BIO-INSPIRED PHYSIOLOGICAL SIGNAL(S) AND MEDICAL IMAGE(S) NEURAL PROCESSING SYSTEMS BASED ON DEEP LEARNING AND MATHEMATICAL MODELING FOR IMPLEMENTING BIO-ENGINEERING APPLICATIONS IN MEDICAL AND INDUSTRIAL FIELDS

EDITED BY: Francesco Rundo, Giuseppe Luigi Banna, Concetto Spampinato and Sabrina Conoci

PUBLISHED IN: Frontiers in Oncology, Frontiers in Computer Science, Frontiers in Materials, Frontiers in Bioengineering and Biotechnology, Frontiers in Medicine, Frontiers in Neuroinformatics and Frontiers in Neurorobotics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-916-7
DOI 10.3389/978-2-88971-916-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

BIO-INSPIRED PHYSIOLOGICAL SIGNAL(S) AND MEDICAL IMAGE(S) NEURAL PROCESSING SYSTEMS BASED ON DEEP LEARNING AND MATHEMATICAL MODELING FOR IMPLEMENTING BIO-ENGINEERING APPLICATIONS IN MEDICAL AND INDUSTRIAL FIELDS

Topic Editors:
Francesco Rundo, STMicroelectronics, Italy
Giuseppe Luigi Banna, United Lincolnshire Hospitals NHS Trust, United Kingdom
Concetto Spampinato, University of Catania, Italy
Sabrina Conoci, University of Messina, Italy

Citation: Rundo, F., Banna, G. L., Spampinato, C., Conoci, S., eds. (2021). Bio-Inspired Physiological Signal(s) and Medical Image(s) Neural Processing Systems Based on Deep Learning and Mathematical Modeling for Implementing Bio-Engineering Applications in Medical and Industrial Fields. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-916-7

Table of Contents

frontiers in Neuroinformatics

EDITORIAL
published: 29 October 2021
doi: 10.3389/fninf.2021.763699

Editorial: Bio-Inspired Physiological Signal(s) and Medical Image(s) Neural Processing Systems Based on Deep Learning and Mathematical Modeling for Implementing Bio-Engineering Applications in Medical and Industrial Fields

Francesco Rundo[1]*, *Giuseppe Luigi Banna*[2], *Concetto Spampinato*[3] and *Sabrina Conoci*[4]

[1] STMicroelectronics, ADG Central R&D, Catania, Italy, [2] United Lincolnshire Hospitals NHS Trust, Louth, United Kingdom, [3] PerCeiVe Lab, University of Catania, Catania, Italy, [4] Department of Chemical, Biological, Pharmaceutical and Environmental Sciences, University of Messina, Messina, Italy

Keywords: neural computing, bio-inspired systems, bio-engineering, industrial applications, automotive

Editorial on the Research Topic

Bio-Inspired Physiological Signal(s) and Medical Image(s) Neural Processing Systems Based on Deep Learning and Mathematical Modeling for Implementing Bio-Engineering Applications in Medical and Industrial Fields

Edited and reviewed by:
Antonio Fernández-Caballero, University of Castilla-La Mancha, Spain

***Correspondence:**
Francesco Rundo
francesco.rundo@st.com

Received: *25 August 2021*
Accepted: *27 September 2021*
Published: *29 October 2021*

Citation:
Rundo F, Banna GL, Spampinato C and Conoci S (2021) Editorial: Bio-Inspired Physiological Signal(s) and Medical Image(s) Neural Processing Systems Based on Deep Learning and Mathematical Modeling for Implementing Bio-Engineering Applications in Medical and Industrial Fields. Front. Neuroinform. 15:763699. doi: 10.3389/fninf.2021.763699

This Research Topic gathers different contributions highlighting novel types of bio-inspired mathematical models with neuroimaging and physio-signal data, mainly applied in medical and industrial field. The target of this Research Topic was to collect scientific contributions which were able to highlight the undoubted advantages that modern techniques of Deep Learning and Bio-inspired Neural computing can offer in addressing the main issues in the medical and industrial field. The correlated target of the Research Topic was also to highlight the significant contribution of multimodal data analysis (signals and images) in the medical and industrial field. The collected accepted articles contribute significantly to the achievement of the Research Topic target, confirming, as below detailed, the capabilities of the combined approach of bio-inspired mathematical models and multimodal data analysis.

The first accepted article of this Research Topic (Folego et al.) proposed an early detection method of Alzheimer's disease (AD) by means of a 3D Convolutional Deep Network applied to MRI imaging of the analyzed subjects. It is well-known that early diagnosis is paramount to the development and success of interventions, and neuroimaging represents one of the most promising areas for early detection of AD. Through the neuroimaging deep analysis the authors investigated the implementation of such biomarkers suitable to classify brain images into AD, mild cognitive impairment (MCI), and cognitively normal (CN) groups. Their deep network solution embedded domain adaptation to improve the performance in neuro-imaging classification. The performance of the proposed solution shows promising results for CN discrimination (67.3% with TPF metric) against lower performance for MCI and AD early detection. Anyway, the collected results outperformed the compared similar pipelines.

Still in the medical field, the contribution proposed by Zhen et al. is of considerable interest. The authors proposed a deep learning-based pipeline for accurate diagnosis of a Liver cancer

based on the analysis of multimodal data i.e., MRI imaging combined with such clinical data of the analyzed patients. Several scientific contributions confirmed that early-stage diagnosis and treatment of liver cancer can improve survival rates. Dynamic contrast-enhanced MRI provides the most comprehensive information for differential diagnosis of liver tumors. However, MRI-based diagnosis is affected by subjective physicians experience in addition to the difficulties implicit in the method (limited spatial resolution). The authors showed that deep learning solutions they implemented may supply to these limitations. To improve the performance of the developed deep architecture, the authors processed MRI images combined with clinical data. They analyzed a dataset of 1,210 patients with liver tumors ($N = 31{,}608$ images) used as training set while the learned model were validated in an external independent extended cohort of 201 patients ($N = 6{,}816$ images).Using only unenhanced images, the proposed deep classifier performs well in distinguishing malignant from benign liver tumors (AUC, 0.946; 95% CI 0.914–0.979 vs. 0.951; 0.919–0.982, $P = 0.664$). The authors tried to combine unenhanced MRI images with clinical data. This setup improved significantly the performance of classifying malignancies as hepatocellular carcinoma (AUC, 0.985; 95% CI 0.960–1.000), metastatic tumors (0.998; 0.989–1.000), and other primary malignancies (0.963; 0.896–1.000). The very promising results confirmed that multimodal data processing through *ad-hoc* deep classifier can be considered as valuable tool for robust differential diagnosis of liver cancer.

In Henriques-Pons et al. the authors proposed an interesting mathematical analysis to the flow cytometry labeling. Conventional flow cytometry analysis relies on the creation of dot plot sequences, based on two fluorescence parameters at a time, to evidence phenotypically distinct populations. Anyway, results of such classical approach is not always robust. The authors proposed an interesting mathematical analysis named MCTA (Multiparametric Color Tendency Analysis) which considers multiple labelings simultaneously, extending and complementing conventional analysis. The MCTA method executes the background fluorescence exclusion, spillover compensation, and a user-defined gating strategy for subpopulation analysis. The performance showed are very promising and then the MCTA it deserves to be further investigated for applications in the biomolecular field.

Still in medical oncology field, an interesting literature review was proposed by Sa et al. in which the Fluoro-Deoxyglucose Positron Emission Tomography/Computed Tomography (PET/CT) imaging method was analyzed for Pediatric Rhabdomyosarcoma (RMS) staging and prognosis estimation. Thirteen consecutive patients with pathologically confirmed RMS underwent PET/CT scan for evaluation of therapy response. About the baseline PET/CT, most RMSs are located in the pelvic cavity, and upper arms ranked second. About evaluation of the disease spread, lymph node metastases were seen in eight patients, and eight patients showed distant metastasis to the lung, liver, and bone. The median SUVmax, SUVmean, and SUVpeak of primary sites were 7.1, 4.0, and 5.9, respectively. The scientific paper confirmed that PET/CT scan methodology could be a valid tool for RMS assessing as this kind of cancer demonstrates increased glycolytic activity. Again, *ad-hoc* mathematical analysis of imaging data showed very promising results in medical field.

A very interesting approach in the field of neurology was proposed by Li et al. in which the authors implemented a mathematical approach for distinguishing Epileptiform Discharges (ED) from normal electroencephalograms (EEG). It is well-known that EDs are of fundamental importance in understanding the physiology of epilepsy. To aid in the clinical diagnosis, classification, prognosis, and treatment of epilepsy, it is important to implement robust solutions to distinguish epileptiform discharges from normal electroencephalogram (EEG). This is a challenging task. To take on this challenge, they proposed to use a multiscale complexity measure, the scale-dependent Lyapunov exponent (SDLE). The authors analyzed 640 multi-channel EEG segments, each 4 s long. Among these segments, 540 are short epileptiform discharges, and 100 are from healthy controls. They noticed that such SDLE features can be effective in distinguishing epileptiform discharges from normal EEG. They tested different machine learning classifier such as Random Forest Classifier (RF) and Support Vector Machines (SVM), obtaining an accuracy around 99% in distinguishing ED from normal EEG. The confirmed robustness of the approach based on SDLE analysis suggest further investigation of the approach with the target to use that widely in a clinical setting.

Another promising pipeline based on multimodal data analysis for applications in the medical field can be found in the contribution (Bartoletti et al.) in which the authors correlated such clinical data of the patients with Bioelectric Impedance Analysis (BIA) in order to select the patients candidate to prostate biopsy. The authors analyzed a cohort of one-hundred 40 consecutive candidates to prostate biopsy and 40 healthy volunteers. For each recruited subject the following clinical data have been collected: PSA and PSA density determinations (PSAD), Digital Rectal Examination (DRE), and the novel BIA test. The targets of the proposed investigation were to determine accuracy of BIA test in comparison to PSA, PSAD levels and MRI and obtain Prostate Cancer (PCa) prediction by BIA test. The authors performed a lot of experimental results which confirmed what follow: Combined PSA, PSAD, DRE, and trans rectal ultrasound test failed to discern patients with PCa from those with benign disease (62.86% accuracy; sensitivity of 83% and a specificity of 59%). The accuracy in discerning PCa increased up to 75% by BIA test (sensitivity 63.33% and specificity 83.75%) confirming the effectiveness of the proposed approach.

Another field of application of the analyzed Research Topic was the investigation of bio-inspired models and solutions. This includes the study proposed by Jia et al. which analyzed the impact of some Neuron Dendritic Spine Patterns. Some pattern abnormalities of dendritic spine, seems to be correlated to multiple nervous system diseases, such as Parkinson's disease and schizophrenia. The analysis of these spine patterns can help to bring-up a valid model of the pathogenesis of these diseases. The authors investigated the application of the bio-inspired reaction-diffusion model to simulate the formation patterns dynamic of dendritic spine. The authors also investigated

the deep regulation mechanisms of dendritic spine. The authors was able to define a robust mathematic model-based pathogenesis research for neuron diseases correlated to the dendritic spine pattern abnormalities analysis. With this bio-inspired model, it will be possible to better study the pathogenesis of certain neurodegenerative diseases in order to identify an effective treatment.

The field of medical oncology has been significantly explored in the Research Topic. The authors of the contribution reported in Gao et al. have invested the application of modern machine learning methods to the assessment of the tumor grading and such bio-marker (Ki67, GFAP expression level, S100 expression level, etc.) of such type of brain cancer (Gliomas). The grading and pathologic biomarkers play a key role in the diagnosis and treatment of the Glioma. The authors proposed a pipeline aimed to use conventional machine learning algorithms to predict the tumor grades and pathologic biomarkers based on the analysis of magnetic resonance imaging (MRI) data. The authors analyzed a dataset of 367 glioma patients, who had pathological reports and underwent MRI imaging scans. The extracted MRI image features was processed by several type of machine learning based classifier such as Logistic Regression based classifier (LR), Support Vector Machine (SVM), and Random Forests (RF). The RF algorithm outperformed the compared classifiers i.e., Logistic Regression and SVM. The RF classifier on glioma grades achieved a predictive performance (AUC: 0.79, accuracy: 0.81) and a predictive performance of AUC: 0.85, accuracy: 0.80 for the Ki67 expression AUC: 0.72 and accuracy: 0.81 for the GFAP and AUC: 0.60 accuracy: 0.91 for the S100 expression level. The robustness of this results encourages further investigation in this area.

In the field of neurovascular diagnostics, multimodal data analysis showed significant contribution. The authors of the contribution (Zhang et al.) investigated the application of PLA-combined ferroferric graphene oxide aspirin (Fe_3O_4-GO-ASA) multifunctional nanobubbles in the prevention and treatment of thrombotic events in a preclinical study. The experimental results confirmed that PLA-combined Fe_3O_4-GO-ASA nanobubbles treatment has significantly inhibit thrombosis (concentration of 80 mg·mL−1 interacted with the rabbit blood). The proposed approach also showed a relevant ultrasonic imaging effect and a good magnetic targeting for an efficient diagnostic of the vascular disease.

The multi-modal bio-medical data analysis also involved the study of the motility and shape of the tumor cells of a subject to characterize the biology of the tumor and the selection of the most appropriate treatment. The authors of the study reported in D'Orazio et al. analyzed motility and such shape features of cancer cells *in vitro* in order to assess diagnosis and treatment. They combined fluorescence time-lapse microscopy (TLM) and label-free imaging, with cell tracking, quantitative representation of cell trajectories, and a recent machine learning (ML) strategies based on peer prediction algorithm. The implemented *ad-hoc* cooperative learning approaches in order to discriminate with high accuracy non-cancer vs. cancer cells of high vs. low malignancy. They investigated the performance of the proposed solution in the treatment of prostate cancer. Comparison with standard classification methods validated their promising proposed approach.

Still with reference to prostate cancer, a pre-clinical study reported in Merisaari et al. analyzed the impact of the mathematical modeling of MRI-DWI (Diffusion Weighted Imaging) sequences in prostate cancer assessment. Classical approaches assess the tumor growth by weekly examination of the DWI imaging by using a 7T MRI scanner. The authors implemented *ad-hoc* mathematical models for performing additional DWI examination. They observed significant changes in their DWI data mathematical features (stretched exponential and kurtosis) during the tumor growth confirming that the proposed approach shows promising performance as a robust tool for the prostate cancer follow-up.

In the field of medical neurology, an interesting multi-modal approach was presented by the authors of the contribution reported in Chen et al. They investigated the performance of such machine learnings approaches for seizure prediction based on the analysis of EEG data. They implemented a multi-dimensional enhanced seizure prediction framework, which embedded a graph state encoder, and a space-time predictor. The input data was the multi-channels EEG. Their proposed model analyzed the space-time relationship of the input data with the seizure event in epileptic subject. The authors validated their approach on a public dataset retrieving a sensitivity of 98.61%, confirming the effectiveness of the proposed solution.

Unfortunately, to date, there are particularly aggressive tumors that unfortunately denote an high rate of mortality. Among these it is worth mentioning the pancreatic cancer which in some forms is particularly lethal. The authors of the contribution reported in Han et al. investigated the reliability of radiomics pipeline applied to contrast-enhanced CT scan imaging data for discriminating pancreatic cystadenomas from pancreatic neuroendocrine tumors (PNETs). They implemented effective machine-learning pipelines to perform this discrimination. They retrospectively analyzed 120 patients, including 66 pancreatic cystadenomas patients and 54 PNETs patients. They identified 48 had-crafted radiomic features from contrast-enhanced CT images to be classified by classical machine learning methods such as linear discriminant analysis (LDA), Random Forest, Adaboost, support vector machine (SVM), k-nearest neighbor (KNN), logistic regression (LR), and so on. The proposed deep classifier shows promising ability of differentiating pancreatic cystadenomas from PNETs. Specifically, the RF-based classifier, as well as Xgboost+RF, demonstrated the best discriminative ability, with the highest AUC of 0.997 in the testing group.

In the context of the bio-inspired models' study (one of the main target of the Research Topics), the contribution reported in Quan et al. shows considerable interest. The authors proposed *FusionNet* a novel deep fully residual convolutional neural network for image segmentation in connectomics. Neuro-connectomics tries to generate comprehensive brain connectivity maps using high-throughput, nano-scale electron microscopy. Deep learning showed very promising performance in image processing and computer vision, leading to a recent explosion

in popularity. For these reasons the author implemented *ad-hoc* deep architecture named *FusionNet* to perform automatic segmentation of neuronal structures in connectomics data. FusionNet is a fully convolutional deep architecture which embeds semantic segmentation ability combined with residual layers which improve the overall segmentation capability. The authors successfully validated their deep backbone providing robust comparison with other popular electron microscopy and confirming the very promising performance of the proposed architecture.

The analyzed scientific contributions confirmed that the convolutional architectures show remarkable skills in image processing and computer vision tasks in medical and industrial fields. In the contribution reported in Tang et al. the authors investigated the use of specific convolutional architecture for the diagnosis of Intramucosal Gastric Cancer. A deep convolutional neural network was implemented to learn a retrospectively collected 3,407 endoscopic images from 666 gastric cancer patients. The deep network performance was validated over a test-set composed by 228 images from 62 independent patients. The implemented deep architecture was able to discriminate intramucosal Gastric Cancer from advanced Gastric Cancer with an AUC of 0.942, a sensitivity of 90.5%, and a specificity of 85.3% in the testing dataset. The reported results confirmed that deep learning can be effectively used to improve diagnostic accuracy in medical oncology.

Medical branch related to oncology is certainly pathological anatomy. Also in this context, multimodal analysis together with machine learning methods can significantly improve the histopathological characterization of tumors. More in detail, in the article reported in Hao et al. the authors proposed a low dimensional three-channel features based breast cancer histopathological images classification approach suitable to discriminate benign breast cancer from malignant ones. Several hand-crafted image descriptors including gray level co-occurrence matrix on different directions, average pixel value of each channel, Hu invariant moment (HIM), wavelet features, and son on, are defined as input of Support Vector Machine classifier. Experiments on specific dataset showed that the proposed pipeline achieved an accuracy of 90.2–94.97% at the image level and 89.18–94.24% at the patient level, which outperforms many state-of-the-art methods.

As the analyzed Research Topic covered applications in the industrial field, it is of particular interest to include among the publications of this scientific issues, the contribution reported in Rundo et al. The authors investigated the usage of such neuro-physiological signals (specifically the PhotoPlethysmoGraphy) in the automotive field. With the aim to develop an intelligent driving assistance system, the authors implemented a complex system (both hardware and software) which embeds different deep architectures with innovative attention mechanisms. Specifically, the proposed system detects and tracks the car driver attention level from analysis of correlated PhotoPlethysmoGraphy sampled from a bio-sensor embedded in the car steering. This retrieved attention assessment was correlated to driving scenario risk evaluation made by another deep architecture which embeds modern self-attention mechanisms. A combined intelligent monitor evaluates the congruence between the driver's level of attention as above determined with the driving scenario risk level, generating such alerts if there is an inconsistency between the two assessments. The approach was applied in a specific automotive use-case, namely the tracking of pedestrians in a common driving scenario. In addition, further computer vision systems will assist the driver's attention assessment. The high accuracy of the proposed driver's attention detection system based on physiological analysis (more than 95%) as well as of the parallel driving risk assessment systems confirmed the effectiveness of the proposed method and of the underlying hardware platform. Therefore, also in industrial (automotive) field, the multimodal data analysis combined with bio-inspired deep processing showed high capability in addressing typical issues.

In conclusion of this scientific review, the authors hope that the reader will find in this Research Topic a useful reference for the state of the art in the emerging field of bio-inspired deep models and multimodal data analysis for addressing issues and problems in the medical and industrial field.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

Conflict of Interest: FR was employed by the company STMicroelectronics.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 28 May 2020
doi: 10.3389/fonc.2020.00680

Deep Learning for Accurate Diagnosis of Liver Tumor Based on Magnetic Resonance Imaging and Clinical Data

Shi-hui Zhen[1,2†], Ming Cheng[2†], Yu-bo Tao[2], Yi-fan Wang[1], Sarun Juengpanich[1], Zhi-yu Jiang[1], Yan-kai Jiang[1,2], Yu-yu Yan[2], Wei Lu[3,4], Jie-min Lue[1], Jia-hong Qian[2], Zhong-yu Wu[5], Ji-hong Sun[3], Hai Lin[2]* and Xiu-jun Cai[1]**

[1] Department of General Surgery, School of Medicine, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou, China, [2] State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou, China, [3] Department of Radiology, School of Medicine, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou, China, [4] Department of Medical Imaging, Hwa Mei Hospital, University of Chinese Academy of Sciences, Ningbo, China, [5] Department of Surgical Oncology, School of Medicine, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:
Francesco Rundo,
STMicroelectronics, Italy

Reviewed by:
Mohammed Benjelloun,
University of Mons, Belgium
Francesca Trenta,
University of Catania, Italy

***Correspondence:**
Ji-hong Sun
sunjihong@zju.edu.cn
Hai Lin
lin@cad.zju.edu.cn
Xiu-jun Cai
srrsh_cxj@zju.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 18 February 2020
Accepted: 09 April 2020
Published: 28 May 2020

Citation:
Zhen S, Cheng M, Tao Y, Wang Y, Juengpanich S, Jiang Z, Jiang Y, Yan Y, Lu W, Lue J, Qian J, Wu Z, Sun J, Lin H and Cai X (2020) Deep Learning for Accurate Diagnosis of Liver Tumor Based on Magnetic Resonance Imaging and Clinical Data. Front. Oncol. 10:680. doi: 10.3389/fonc.2020.00680

Background: Early-stage diagnosis and treatment can improve survival rates of liver cancer patients. Dynamic contrast-enhanced MRI provides the most comprehensive information for differential diagnosis of liver tumors. However, MRI diagnosis is affected by subjective experience, so deep learning may supply a new diagnostic strategy. We used convolutional neural networks (CNNs) to develop a deep learning system (DLS) to classify liver tumors based on enhanced MR images, unenhanced MR images, and clinical data including text and laboratory test results.

Methods: Using data from 1,210 patients with liver tumors ($N = 31,608$ images), we trained CNNs to get seven-way classifiers, binary classifiers, and three-way malignancy-classifiers (Model A-Model G). Models were validated in an external independent extended cohort of 201 patients ($N = 6,816$ images). The area under receiver operating characteristic (ROC) curve (AUC) were compared across different models. We also compared the sensitivity and specificity of models with the performance of three experienced radiologists.

Results: Deep learning achieves a performance on par with three experienced radiologists on classifying liver tumors in seven categories. Using only unenhanced images, CNN performs well in distinguishing malignant from benign liver tumors (AUC, 0.946; 95% CI 0.914–0.979 vs. 0.951; 0.919–0.982, $P = 0.664$). New CNN combining unenhanced images with clinical data greatly improved the performance of classifying malignancies as hepatocellular carcinoma (AUC, 0.985; 95% CI 0.960–1.000), metastatic tumors (0.998; 0.989–1.000), and other primary malignancies (0.963; 0.896–1.000), and the agreement with pathology was 91.9%.These models mined diagnostic information in unenhanced images and clinical data by deep-neural-network, which were different to previous methods that utilized enhanced images. The sensitivity and specificity of almost every category in these models reached the same high level compared to three experienced radiologists.

Conclusion: Trained with data in various acquisition conditions, DLS that integrated these models could be used as an accurate and time-saving assisted-diagnostic strategy for liver tumors in clinical settings, even in the absence of contrast agents. DLS therefore has the potential to avoid contrast-related side effects and reduce economic costs associated with current standard MRI inspection practices for liver tumor patients.

Keywords: liver cancer, liver mass, deep learning, diagnosis, artificial intelligence, MRI

INTRODUCTION

Liver cancer is the second leading cause of cancer-related deaths worldwide (1) and the incidence rate has been growing on a global scale (2), which is in contrast to the stable incidence or declining trends for most cancers (3). Hepatocellular carcinoma (HCC) accounts for 90% of primary liver cancers, and could result in a major global health problem. Early-stage HCC detection and diagnosis can allow the patients to receive the treatment earlier and achieve better survival rates (1). All the continental and national liver disease societies have recommended that surveillance should be carried out for high-risk patients with cirrhosis (4). Ultrasonography as the preferred test for surveillance is unsatisfactory because of the limitations of ultrasound operator dependency and its low sensitivity to small liver cancers (5). Dynamic contrast-enhanced imaging is recommended as the first-line diagnostic tool for HCC when the screening test result is abnormal (6). Undoubtedly, compared with computed tomography (CT), enhanced MRI is the better choice because of its various tissue contrast mechanisms and it being only way to assess all major and auxiliary imaging features (7). However, enhanced MRI could not be used widely like ultrasonography in screening and surveillance owing to its high cost and the risk of contrast- related side effects (8–12).

Actually, even with enhanced MRI, it still remains challenging to diagnose, owing to liver tumor diversity and complex imaging features. In addition to HCC, primary malignant tumors in the liver include intrahepatic cholangiocarcinoma (ICC), mixed hepatocellular-cholangiocarcinoma (HCC-CC), and other rare tumors (13, 14). The liver is also the target of metastasis for many malignant tumors, such as colorectal, pancreatic, neuroendocrine, breast cancer, etc. Moreover, there are several types of benign tumors in the liver, such as cyst, hemangioma, focal nodular hyperplasia (FNH), adenomas, high-risk cirrhotic nodules [regenerative nodules (RN), and dysplastic nodules (DN)] (14). The evaluations of images are generally subjective and are possibly affected by radiologists' experience (15, 16).

Unlike traditional image-dependent "semantic" features evaluation from human experts, deep learning can automatically learn feature representations from sample images with convolutional neural networks (CNNs) (17, 18). It has shown to match or even surpass human performance in the application of specific tasks and may even discover additional differential features not yet identified in current radiological practice (19). CNNs have been achieving good performances in medical imaging for several tumor types (20–22), but for liver tumors only a few exploratory studies have been reported (16, 23–25). These studies trained models based on enhanced images from contrast-enhanced CT/MRI in small-scale (<500 patients) datasets, however, they only considered specific liver tumor categories and did not simulate clinical practice conditions, which restricted their utility in the diagnostic decision-making phases of the workflow. In addition, the potential diagnostic value of clinical information and unenhanced sequences, including T1 pre-contrast, T2, and diffusion sequences, were not evaluated in deep-learning models. Here, we report the results of a large study of liver tumors, which covered all types of hepatic local lesions except for inflammatory masses. There were two aims of this study: First, we developed CNNs that implemented assisted diagnosis for liver tumors by classifying them in seven categories. Second, we developed a CNN that utilized unenhanced sequences to distinguish malignant from benign tumors, then, our modified CNN that combines unenhanced images with clinical data achieved end-to-end output for precise classification of malignant tumors. These models were integrated into the DLS. In an external independent cohort, their performance was compared with experienced radiologists that had read all sequences and clinical information.

MATERIALS AND METHODS

Patients

This study was approved by the independent institutional review boards of Sir Run Run Shaw Hospital and performed according to the Helsinki declaration. In this study, a sample size was not prespecified. Instead, we included the largest possible number of patients with liver tumors to ensure that deep learning models were trained as fully as possible. The inclusion criteria were as follows: (1) participants had liver tumors, and (2) participants accepted enhanced MRI inspection. The exclusion criteria were as follows: (1) those who had accepted treatment related to the lesion before MRI inspection, including surgery, transcatheter arterial chemoembolization (TACE), radiofrequency ablation, chemotherapy, radiotherapy, targeted drug therapy, etc.; (2) those with inflammatory lesions; (3) those with a clinically diagnosed malignancy (without pathology confirmed); (4) any missing important medical records or laboratory results of the malignancy individuals; and (5) unqualified image quality. Only patients who had malignancies confirmed by biopsy or post-surgical histopathology were enrolled. The diagnosis of some benign lesions was supported by histopathology, but the labels of other benign lesions without surgery provided by the imaging diagnosis report were considered as our gold standard. For these liver masses diagnosed with a combination of clinical

information and imaging criteria, the follow-up time was 6–10 months. Those labels were the result of a consensus as explained by the radiology department: The report was firstly given by the doctor who read the images. Then, the report was reviewed by the senior doctor. In case of a disagreement, the final decision was confirmed by a department conference. The test set consists of liver focal lesions enhanced by MRI in the same hospital between July 2018 to December 2018 according to the include and exclude criteria above (**Figure S1**). In addition, general demographics for all patients, pathology reports (if any), and malignant patients' related medical history, liver-related symptoms, and laboratory test results were documented.

Taxonomy

Three groups of models were trained. The first task was to classify the liver focal lesions of the training set into seven categories: 0. cyst, 1. hemangioma, 2. focal liver lesion (FNH), 3. other benign nodules (cirrhotic nodules, RN, DN, rare benign tumors), 4. HCC, 5. metastatic malignant tumors from other sites (colorectal, breast, lung, pancreas, etc.), 6. primary hepatic malignancy other than HCC (intrahepatic cholangiocarcinoma, adenocarcinoma, etc.). The second task was to divide all the tumors into two categories: a. benign (include 0, 1, 2, 3 above), b. malignant (include 4, 5, and 6). In the third task, malignant tumors were divided into three categories the same as category 4, 5, and 6 in the seven-category method above. The test set was also classified as described above and labeled. Those classification models included all liver mass-like lesions (except for inflammatory lesions). In category 3, 5, and 6, several different specific tumor types were all mixed in the training and test set. Three experienced radiologists independently read and labeled the MRI of the validation set following this standard and they could refer to additional information such as medical history, laboratory test results, etc. according to daily clinical work habits.

MRI Acquisition Protocol

Abdominal MRIs were performed in a supine position with 1.5-T, 3-T, and 750 W MR scanners, including GE MR Singna HDX 3.0T, GE MR Singna HD Excite scanners 1.5T, Simen MR Skyra 3.0T, Simen magnetom avanto Dot1.5T, and GE discovery MR 750 scanners. T2-weighted sequence, diffusion-weighted sequence (b-values: 800 s/mm^2) from standard institutional liver MR imaging protocols were performed with acquisition times of 2–2.5 and 2–2.5 min. Contrast-enhanced T1 sequences were used with acquisition times of 12–18 s. Two different contrast agents were used, i.e., Gadopentetate dimeglumine and Gadoxetic Acid Disodium (Primovist) at doses of 0.2 and 0.1 mmol/kg, respectively. Except for pre-contrast T1, T2, and diffusion images, post- contrast images were analyzed, including late arterial phase (∼15 s post-injection), portal venous phase (∼60 s post-injection), and equilibrium phase (∼3 min post-injection). Imaging parameters varied across different scanners and time frames (**Table S4**).

Image Processing

Eligible MRI images were downloaded from the Picture Archiving and Communication Systems (PACS) and stored as Digital Imaging and Communications in Medicine (DICOM) files. Six sequences were selected and the region of interest (ROI) on T2 sequence was annotated. All the sequences were resampled to a resolution of 0.7 × 0.7 × 7 mm. Then the annotations of the other five sequences were generated according to the origin and spacing information of sequences, which were checked manually. Software was developed to correct cases that were not matched. SimpleITK was used to read DICOM images. After preprocessing, resampling, and configuration, the DICOM files were converted to images in preparation for the training. Some common augmentation methods were performed on the images such as rotation, flipping, scaling, shifting, and shearing.

Deep Learning Model Development

Deep CNNs have achieved good results in medical image classification in recent years. The most straightforward way to improve the performance of deep CNNs is to increase their depth and width. Szegedy et al. (26) proposed a deep CNN architecture codenamed Inception that increased the width of each stage. Multiple versions of Inception-Net have been widely used in classification tasks. Residual connections introduced by He et al. (27) make it easy to train very deep networks. Inception- ResNet, which combines both ideas, outperforms the previous networks. We utilized a Google Inception-ResNet V2 CNN architecture (28) that was pre- trained on approximately 1.28 million images (1,000 object categories) from the 2014 ImageNet Large Scale Visual Recognition Challenge (29), then we removed the final classification layer from the network and retrained it with our dataset, fine-tuning the parameters across all layers (30). Our CNN was trained using backpropagation. The loss function was defined as categorical cross entropy between predicted probability and the true class labels in multi-class classification problems. We used stochastic gradient descent (SGD) optimization, with the same global learning rate of 0.1, a decay factor of 0.5 every 20 epochs, a momentum of 0.9, and batch size of 16 for training the weights. The TensorFlow (31) deep learning framework was used to train, validate, and test our network. We resized each image to 299 × 299 pixels in order to make it compatible with the original dimensions of the Inception-ResNet V2 network architecture. During training, images were augmented. Each image was rotated randomly between −40° and 40° and flipped vertically and horizontally with a probability of 0.5. Five-fold cross-validation was used in training CNN, and the parameters of the model with the highest average accuracy on the cross-validation sets were used to train CNN on the whole training set so as to get the final model.

For our dataset, either six (T1, T2 diffusion, late arterial, portal venous, and equilibrium) or three (T1, T2 and diffusion) sequences and clinical data were applied as inputs of our model. Each group of images from six or three sequences can be input to the network through different channels. For the three-way classification model, we modified the network to receive the image and clinical data as inputs simultaneously. The convolution layers were used to extract features from images, and then these features together with encoded clinical data were input to the fully connected layer for classifying liver tumors. Our deep learning model can accept clinical data as input. Clinical data was encoded using one-hot encoding. For example, gender can

be encoded as [0,1] for male and [1,0] for female. The output of network is a one-dimensional vector about the predicted value for each group of images, which is expressed as [P1, P2,..., Pi], i represents i-category classification, Pi represents the predicted value for the i-th category. To calculate the individual-level predicted value, the predicted vector for each image-group of one patient was summed up, then the category with the largest value was used as the final diagnosis of this patient.

Statistical Analysis

Upon finishing the training phase, the performance was evaluated using the validation set, which is composed of images from patients of an external independent dataset not used during the training. Then the predicted probability of each patient was obtained by aggregating the predicted probability values of each group of images. For classification purposes, the ROC curve was used to show the diagnostic ability of the deep learning model in discriminating specific class from others. The ROC curve and the corresponding area under ROC curve (AUC) for each class were calculated in each model using the python library sklearn (32). Differences between various AUCs were compared by using a Delong test. Average sensitivity and specificity of radiologists were displayed in ROC charts, then the sensitivity and specificity of radiologists' consensus were used to compare with the models (see in Tables). Ninety-five percentage CIs for sensitivity, specificity, and accuracy were calculated with the Clopped-Pearson method (33). *P*-values for sensitivity and specificity comparisons were computed using McNemar's test (binomial distribution for exact probabilistic method). In addition, the agreements between the predicted results with pathological/formal report were compared using the Cohen's Kappa statistic (34, 35) and *P*-values were estimated by a two-sample two-tailed z-test score. All tests were two-tailed, and $p \leq 0.05$ was considered to be indicative of statistical significance.

Confusion matrixes demonstrated the mis-classification similarity between the CNNs and human experts. Element (i, j) of each confusion matrix represented the empirical probability to predict class j given that the ground truth was class i.

To analyze the internal features learned by the CNNs of validation sets, the Barnes- Hit implementation of the t-SNE technique (t-distributed Stochastic Neighbor Embedding) (36) was used to reduce the dimensionality and facilitate the visualization of the classes. The values associated with the last fully hidden connected layer were used as an input, and theta was set to 0.5, perplexity to 50, and 10,000 iterations. Colored point clouds represented the different disease categories, representing how the algorithm clusters the diseases.

Saliency Map

To gain further intuition into how the network made its decisions, saliency maps (37), that can visualize the pixels that a network is fixating on for its prediction, are increasingly being used (20, 38). Back-propagation is an application of the chain rule of calculus to compute loss gradients for all weights in the network. The loss gradient can also be back-propagated to the input data layer. By taking the L1 norm of this input layer loss gradient across the RGB channels, the resulting heat map intuitively represents the importance of each pixel for diagnosis. We generated saliency maps about seven categories and typical individual examples to visualize areas on the images that were deemed important for the classification results. All saliency maps were produced using Keras 2.2.0.

RESULTS

Training and Validation Cohort

Between January 1, 2014 to June 30, 2018, MRI images were obtained for the training set from the hepatic focal lesions database in Sir Run Run Shaw Hospital affiliated to Medicine School, Zhejiang University. According to the inclusion and exclusion criteria (**Figure S1**), the complete training set consisted of 31608 MRI images from 1,210 individuals, including 5,268 groups, and each group included six images from different scan sequences (T2, diffusion, Pre-contrast T1, late arterial, portal venous, equilibrium phase). Between July 1, 2018 to December 31, 2018, 6,816 images from 201 individuals for the validation set were obtained from Sir Run Run Shaw Hospital according to strict enrollment criteria to minimize selective bias (**Figure S1**), which ensured that the validation set could reflect the disease composition and distribution waiting to be diagnosed in real-world clinical scenarios. Only malignancies that had been confirmed by biopsy or post-surgical histopathology were enrolled. The diagnosis of some benign lesions was supported by histopathology, but the labels of other benign lesions without surgery provided by formal imaging diagnosis reports were considered as our gold standard. The study classified hepatic local lesions into seven categories: 0. cyst, 1. hemangioma, 2. focal liver lesion (FNH), 3. other benign nodules [cirrhotic nodules, regenerative nodules(RN), dysplastic nodules(DN), rare benign tumors], 4. HCC, 5. metastatic malignant tumors from other sites (colorectal, breast, lung, pancreas, etc.), and 6. primary hepatic malignancy other than HCC (adenocarcinoma, Intrahepatic cholangiocarcinoma, etc.). 0, 1, 2, and 3 above belonged to benign, 4, 5, and 6 belonged to malignant (see Methods, Taxonomy in details). Baseline characteristics of the training set and validation set are shown in **Table S1**. The disease composition and distribution of the validation set were not exactly the same as the training set (**Table S2**). According to TRIPOD statement (39), this validation set can be regarded as an external independent set. The training dataset in the current study is the largest published liver-enhanced MRI dataset with the most types of liver tumors (non- inflammatory lesion).

Deep-Learning Frameworks for Liver Tumor Classification

Our CNN computational strategy is demonstrated in **Figure 1**. In the training set, liver tumors were marked in T2 sequences by trained senior abdominal radiologists based on formal diagnostic reports. Then, six images from six sequences (T2, diffusion, Pre-contrast T1, late arterial, portal venous, and equilibrium phase) were obtained for each cross section of the lesion by processing the images. The medical history and laboratory test results of individuals with malignant tumors were searched from the medical record system and encoded by the auto encoder to obtain

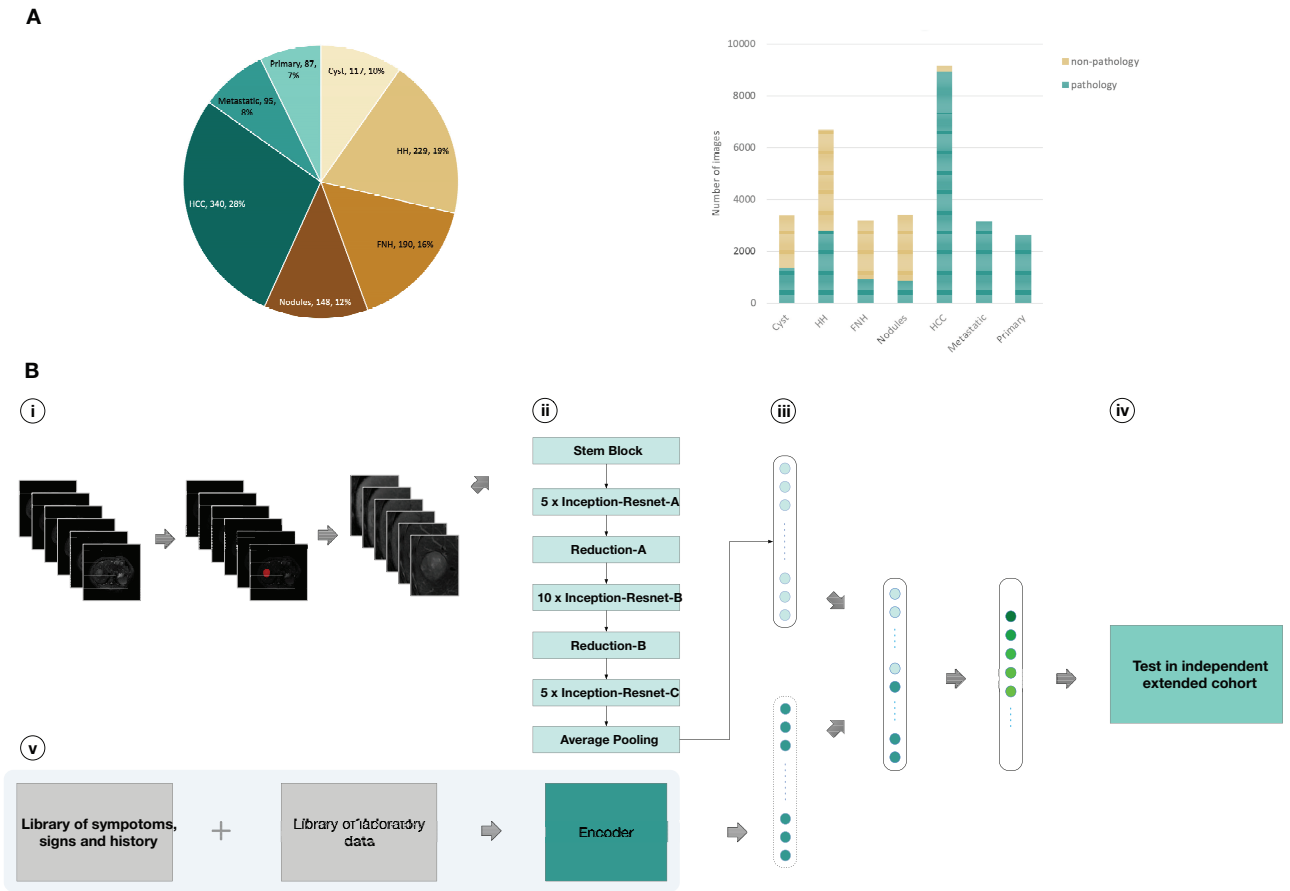


FIGURE 1 | Data and strategy. **(A)**, Number of patients and images per category. **(B)**, Strategy for development and validation. (**B**, i), Magnetic resonance images of patients in training set were first downloaded from the PACS; Liver tumors was outlined in related images of T2-weighted sequences as Regions of interest (ROI) by ITK-SNAP software; pre-processed images and obtained six different scan sequences pictures for each cross section of the lesion (T2, diffusion,T1 pre-contrast, late arterial phase, portal venous phase, equilibrium phase); (**B**, ii), six sequences of each cross section input to CNN as a whole-image from six channels, encoded clinical data was input to CNN; (**B**, iii), the Inception-ResNetV2 architecture was used and fully trained using the training set or partly retrained using the new-training set with clinical data; (**B**, iv), classifications were performed on images from an independent validation set, and the values were finally aggregated per patient to extract the T-SNE and the statistics; (**B**, v), Clinical data was encoded using one-hot encoding as preparation for three-way malignancy classifier. HH, Hepatic hemangioma; Nodules, other benign nodules; Metastatic, Metastatic malignancy from other sites; Primary, Primary malignancy except HCC.

clinical data, including age, gender, cirrhosis-related history, other cancers, tumor marker(AFP, CEA,CA-125,CA19-9, PSA, Ferritin), and liver function (albumin, total bilirubin, prolonged prothrombin time, hepatic encephalopathy, ascites). The coding table is in **Table S3**. The images and clinical information can be used as direct input to the neural networks according to different tasks (**Figure 1**). Based on the computational strategy outlined in **Figure 1**, three group models were trained. First, CNNs were developed to classify images into seven categories. Model A and B used six sequences and three un enhanced sequences (T2, diffusion, Pre-contrast T1) to train CNN, respectively. Second, six sequences (Model C) and three unenhanced sequences (Model D) were utilized to train CNNs for binary classification of benign and malignant. Third, malignant cases with complete clinical data in the training set were selected as a new training set and the integral computational pipeline (**Figure 1**) was applied to train new models in order to classify malignant tumors in to three categories. Model F and Model G, respectively, utilized six and three sequences alongside clinical data as the direct input, while Model E utilized only six imaging sequences as input. The 5-fold cross-validation results during training were reported in **Supplementary Materials-Data File S1**. Then three experienced radiologists were asked to independently classify MRI images in the validation set through the Picture Archiving and Communication Systems (PACS). They could refer to other information about patients such as their medical history, laboratory tests, and so on. This design could better reflect the true level of doctors in daily clinical situation than previous works (16, 23–25) that asked doctors to make judgements only based on images. When the results of the three radiologists were inconsistent, then they discussed together in order to get a diagnoses referred to as reader consensus.

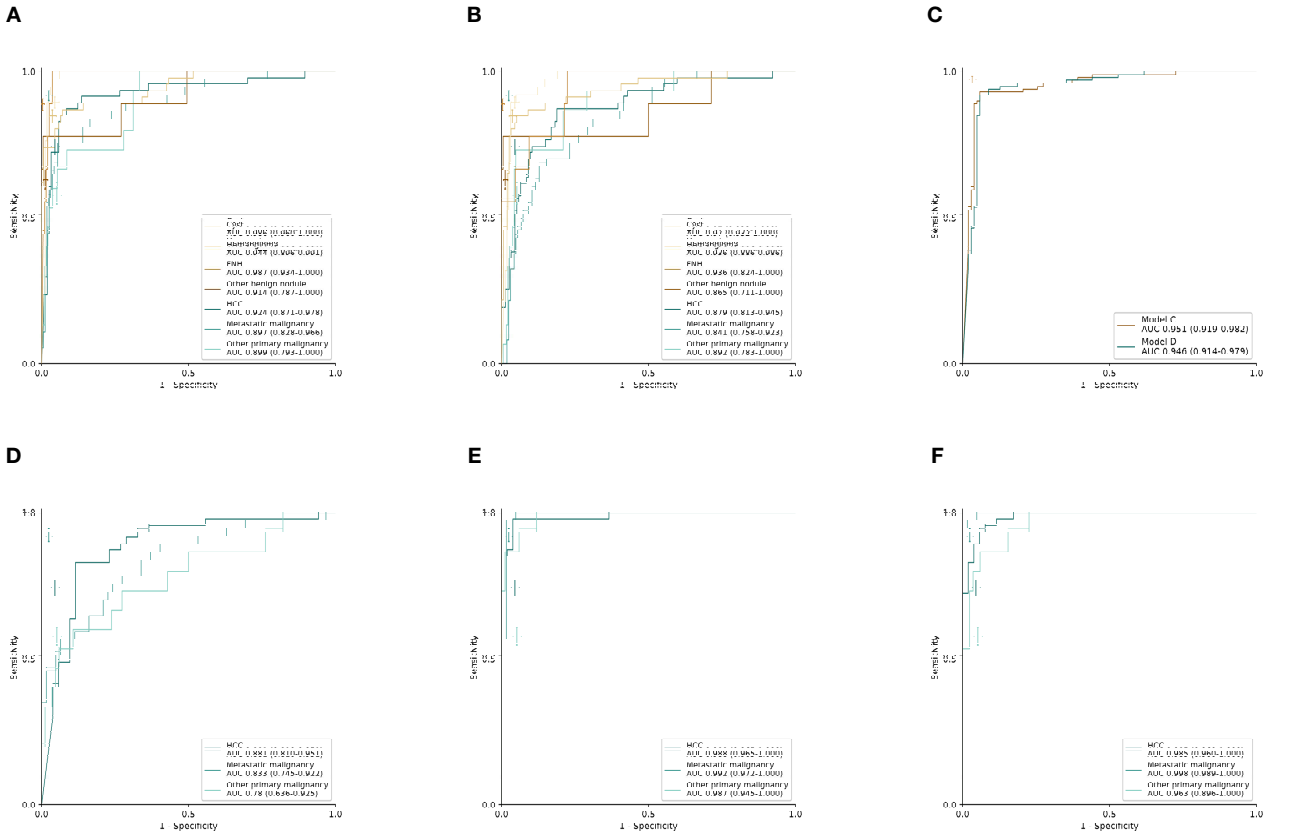


FIGURE 2 | Performance of CNN models and radiologists in external validation set. **(A–C)** Receiver operating characteristic (ROC) curves in the validation set (n = 201 patients). **(A)** Model A: seven-way classifier with six sequences. **(B)** Model B: seven-way classifier with three unenhanced sequences. **(C)** Model C,D: binary classifier for benign and malignancy with six sequences and three unenhanced sequences. **(D–F)** ROC curves in the new validation set of malignant tumors (n = 99 patients). **(D–F)**, **(D)** Model E: three-way classifier with six sequences. **(E)** Model F: three-way classifier with six sequences and clinical data. **(F)** Model G: three-way classifier with three sequences and clinical data. The crosses indicate the performance of average radiologists for each category, the length of the cross represents the confidence Interval (CI).

Deep-Learning Models in Seven-Way Classification Diagnosis

Using the computational pipeline of **Figure 1**, Inception-ResNet V2 was first trained to classify liver tumors in to seven categories according to clinical practical criterion and tested in the validation cohort. Model A with six sequences achieved a high performance in seven-way classification in the validation set (**Figure 2A**), with AUC values for each category ranging from 0.897 (95% CI 0.828–0.966, metastatic malignancy) to 0.987 (95% CI 0.934–1.000, FNH), which was better than Model B with three unenhanced sequences (**Figure 2B**), although no significant difference existed (**Table 1**). Compared with the average level of readers, Model A achieved a competitive level for most categories, but for metastatic malignancy, the average reader seemed to perform better than Model A (**Figure 2A**). The performance between the model and reader consensus was further compared (**Table 1**). The model sensitivity for seven categories ranged from 53.3% (95% CI 26.6–78.7%, other primary malignancy) to 100% (95% CI 66.4–100%, FNH). The sensitivity of reader consensus for seven categories ranged from 55.6% (95% CI 21.2–86.3%, benign nodules) to 94.7% (95% CI 82.3–99.4%, FNH). There was no significant difference ($p > 0.05$) in the sensitivity for each category between Model A and reader consensus, except for metastatic malignancy ($P = 0.003$). In this category, reader consensus had a better performance than Model A because doctors could refer to clinical information including related medical history and laboratory test results, which are of great value for the diagnosis of metastatic tumors. For seven categories, the specificity of model A and reader consensus ranged from 91.6% (95% CI 86.0–95.4%, HCC) to 99.5% (95% CI 97.1–100%, benign nodules) and 94.8% (95% CI 90.0–97.7%, HCC) to 100% (95% CI 98.1–100%, FNH), respectively. Among all categories, only the specificity of FNH demonstrated a significant difference ($P = 0.008$) between Model A and reader consensus. Model A showed a higher sensitivity of 77.8% (95% CI 40.0–97.2%) and specificity of 99.5% (95% CI 97.1–100%) than reader consensus for benign nodules. For benign nodules, the number of patients with an accurate diagnosis by models was more than every experienced radiologist (**Figure S2** for confusion matrix). Seven cases of cirrhosis nodules were

TABLE 1 | Diagnostic performance of seven-way classifiers and radiologists in the validation set.

	Value (95%CI)			McNemar's *P*-value[a]	Delong *P*-value
	Model A	**Model B**	**Reader consensus**	**Model A vs. Reader**	**Model A vs. Model B**
Cyst					
AUC	0.986 (0.960, 1.000)	0.970 (0.932, 1.000)			0.147
Sensitivity, %	89.5 (75.2, 97.1)	89.5 (75.2, 97.1)	94.7 (82.3, 99.4)	0.688	
Specificity, %	96.9 (93.0, 99.0)	96.9 (93.0, 99.0)	98.2 (94.7, 99.6)	0.727	
Hemangioma					
AUC	0.944 (0.897, 0.991)	0.936 (0.886, 0.986)			0.66
Sensitivity, %	82.6 (68.6, 92.2)	87.0 (73.7, 95.1)	89.1 (76.4, 96.4)	0.508	
Specificity, %	95.5 (90.9, 98.2)	93.6 (88.5, 96.9)	98.1 (94.5, 99.6)	0.344	
FNH					
AUC	0.987 (0.934, 1.000)	0.936 (0.824, 1)			0.102
Sensitivity, %	100 (66.4, 100)	66.7 (29.9, 92.5)	88.9 (51.8, 99.7)	1.000	
Specificity, %	95.8 (92.0, 98.2)	94.8 (90.63, 97.5)	100 (98.1, 100)	0.008	
Benign nodules					
AUC	0.914 (0.787, 1.000)	0.865 (0.711, 1.000)			0.139
Sensitivity, %	77.8 (40.0, 97.2)	66.7 (29.9, 92.5)	55.6 (21.2, 86.3)	0.500	
Specificity, %	99.5 (97.1, 100)	99.5 (97.1, 100)	98.4 (95.5, 99.7)	0.625	
HCC					
AUC	0.925 (0.871, 0.978)	0.879 (0.813, 0.9452)			0.137
Sensitivity, %	87.2 (74.3, 95.2)	74.5 (59.7, 86.1)	87.2 (74.3, 95.2)	1.000	
Specificity, %	91.6 (86.0, 95.4)	86.4 (79.9, 91.4)	94.8 (90.0, 97.7)	0.267	
Metastatic malignancy					
AUC	0.897 (0.828, 0.966)	0.841 (0.758, 0.923)			0.039
Sensitivity, %	59.6 (42.1, 75.3)	40.5 (24.8, 57.9)	89.2 (74.6, 97.0)	0.003	
Specificity, %	97.6 (93.9, 99.3)	97.0 (93.0, 99.0)	97.6 (93.9, 99.3)	1.000	
Primary malignancy except HCC					
AUC	0.899 (0.793, 1.000)	0.892 (0.783, 1.000)			0.844
Sensitivity, %	53.3 (26.6, 78.7)	46.7 (21.3, 73.4)	60.0 (32.3, 83.7)	0.688	
Specificity, %	97.9 (94.6, 99.4)	96.8 (93.1, 98.8)	96.2 (92.4, 98.5)	0.727	

[a] *P-value was calculated between Model A (seven-classification) vs. Reader consensus using the McNemar's test.*

all predicted correctly, indicating that the network performs well in differentiating high-risk cirrhosis nodules from HCC. Two wrongly predicted cases of benign nodules were confirmed as bile duct adenoma (predicted HCC) and angiomyolipoma (predicted HCC), respectively (**Figure S3**). HCC false negative cases possessed some common features, such as small tumor sizes or being highly differentiated (**Figure S4**). Those wrongly predicted cases were associated with a lack of enough similar cases in the training set, which may mean the network was not fully trained.

The internal features learned by the CNN using t-SNE (t-distributed Stochastic Neighbor Embedding) (36) were examined (**Figure 3**). Each point represents a group of liver tumor images from six different sequences projected from the 2048-dimensional output of the CNN's last hidden layer into two dimensions. For benign tumors, we observe clusters of points of the same clinical category clustered together (**Figure 3A**), whereas the point distributions of three categories of malignant tumors were not very clearly separated, which is consistent with the observation in the statistical indicators. **Figure 4** shows saliency maps that identify the pixels on which the Inception-ResNet V2 model was fixating its attention on for prediction. As is seen, the network fixates most of its attention on the liver lesions themselves and ignores the background, which is in line with the clinical implication that the lesion and nearby region are more informative of diagnosis. However, the patterns are not specific enough to extract traditional radiologic features, and overall, the saliency map suggests that the deep learning model considered the most important regions when making the prediction, as presented in **Figure 4**.

Deep-Learning Models in Malignancy Diagnosis

Model D was trained on the more challenging task of distinguishing benign and malignant tumors using only using three unenhanced sequences, which exhibited comparable performance to Model C with six sequences. Validated in the independent set, the AUC value was 0.946 (95% CI 0.914–0.979) for Model D and 0.951(95% CI 0.919–0.982) for Model

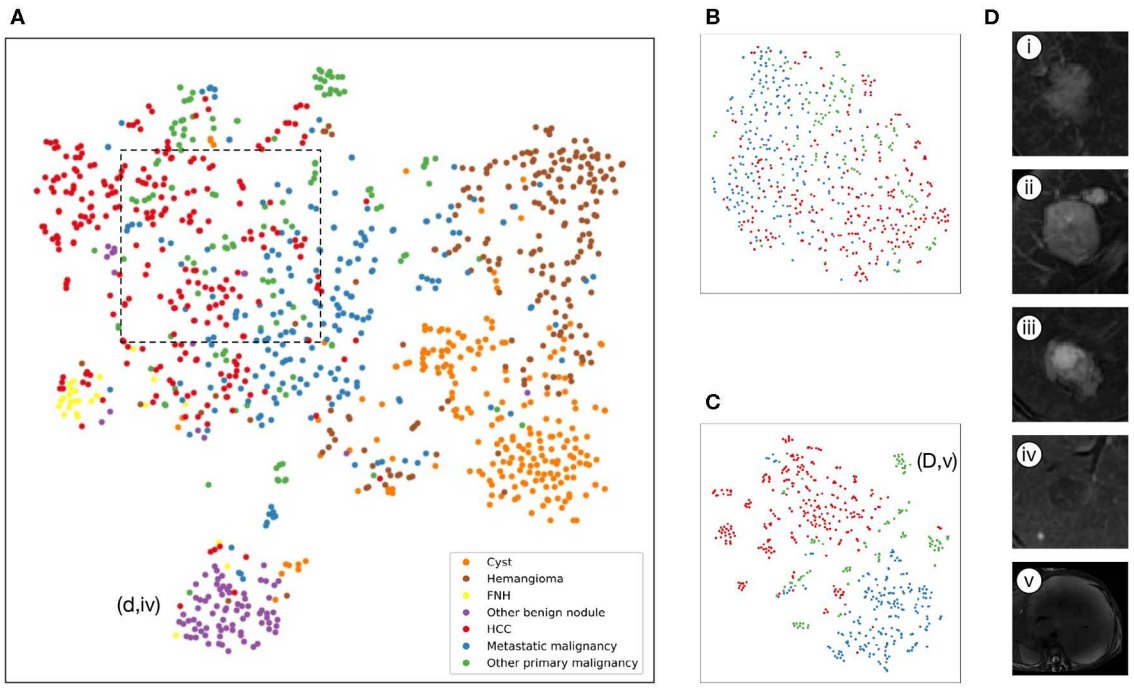


FIGURE 3 | Illustration of classifiers learned by deep-learning projected to 2 dimensions for visualization via the t-SNE algorithm using values of the last fully connected layer in the CNNs of the validation set. **(A–C)**, Scatterplots where each point represents an image of lesions and the color represents the true category, show how the algorithm clusters. **(A)**, Model A: seven-way classifier with six sequences images, shows that seven clusters of the same clinical classes, and we can see benign tumor clusters are better than that of three malignant tumors. The purple point clouds(benign nodules) are effectively divided from red point clouds (HCC). **(B)**, Model E: three-way classifier with six sequences images for malignant tumors. **(C)**, Model G: three-way classifier with three sequences images and clinical data for malignant tumors, shows that three different color point clouds are more effectively clustered than Model E. **(D)** Insets of T2 images show some categories. (i), Hepatocellular carcinoma (ii), Metastatic malignant tumors from pancreas (iii), Intrahepatic cholangiocarcinoma (iv) DN that are difficult to identify with HCC. (v) malignant fibrous histiocytoma represented by outlier point clouds of c(d,v).

C, but two ROC curves exhibited no significant difference in Delong's test ($P = 0.664$), which demonstrated that Model C and Model D have similar performances (**Figure 2C**). The sensitivity of Model C and Model D was 91.9% (95% CI 84.7–96.5%) and 90.9% (95% CI 83.4–95.8%), respectively, which were slightly lower than 99.0% (95% CI 94.5–100%) of reader consensus, but did not reach statistical significance ($P = 0.375$ and 0.219, respectively, estimated by the McNemar's test using binomial distribution). Specificity of Model C and Model D also had no significant difference compared with consensus ($P = 0.549$ and 0.754, respectively) (**Table 2**). In Model D, 93.6% (44/47) of HCC was correctly predicted as a malignancy, while the whole seven individuals of RN/DN nodules were all correctly predicted as benign. It showed that the network without enhanced images can effectively differentiate malignant from benign tumors, even for HCC and high-risk cirrhosis nodules that are difficult to distinguish in traditional HCC diagnostic imaging frameworks, such as LI-RADS, in the absence of contrast agents. As for atypical HCC that were wrongly predicted (very small lesions, benign tumors with carcinogenesis, highly differentiated, etc.), more cases for training are needed.

Deep-Learning Models in Three-Way Classification Diagnosis for Malignancy

An approach that joined clinical data to CNN resulted in a much higher performance in Model F and Model G (**Figures 2E,F**, **Table 3**). The AUCs of Model G improved to 0.985 (95% CI 0.960–1.000, HCC), 0.998 (95% CI 0.989–1.000, metastatic malignancy), and 0.963 (95% CI 0.896–1.000, other primary malignancy), which were significantly better than those of 0.881 (95% CI 0.810–0.951), 0.833 (95% CI 0.745–0.922) and 0.780 (95% CI 0.636–0.925) in Model E (**Figure 2D**) with six sequences (P-values of 0.002, 0.0002, and 0.008, respectively). However, no statistical significance was observed in each category between Model F and Model G (P-values of 0.002, 0.0002, and 0.008, respectively), which demonstrated that after adding clinical data, classifiers with six sequences and three sequences had similar performances. Sensitivity and specificity of each category in Model G had no statistical significance compared with those of reader consensus (**Table 3**). Among three categories, the latter two included many specific tumors from different sites and histopathological sources (**Table S2**). The CNN network with the new approach is highly inclusive with the tumor type

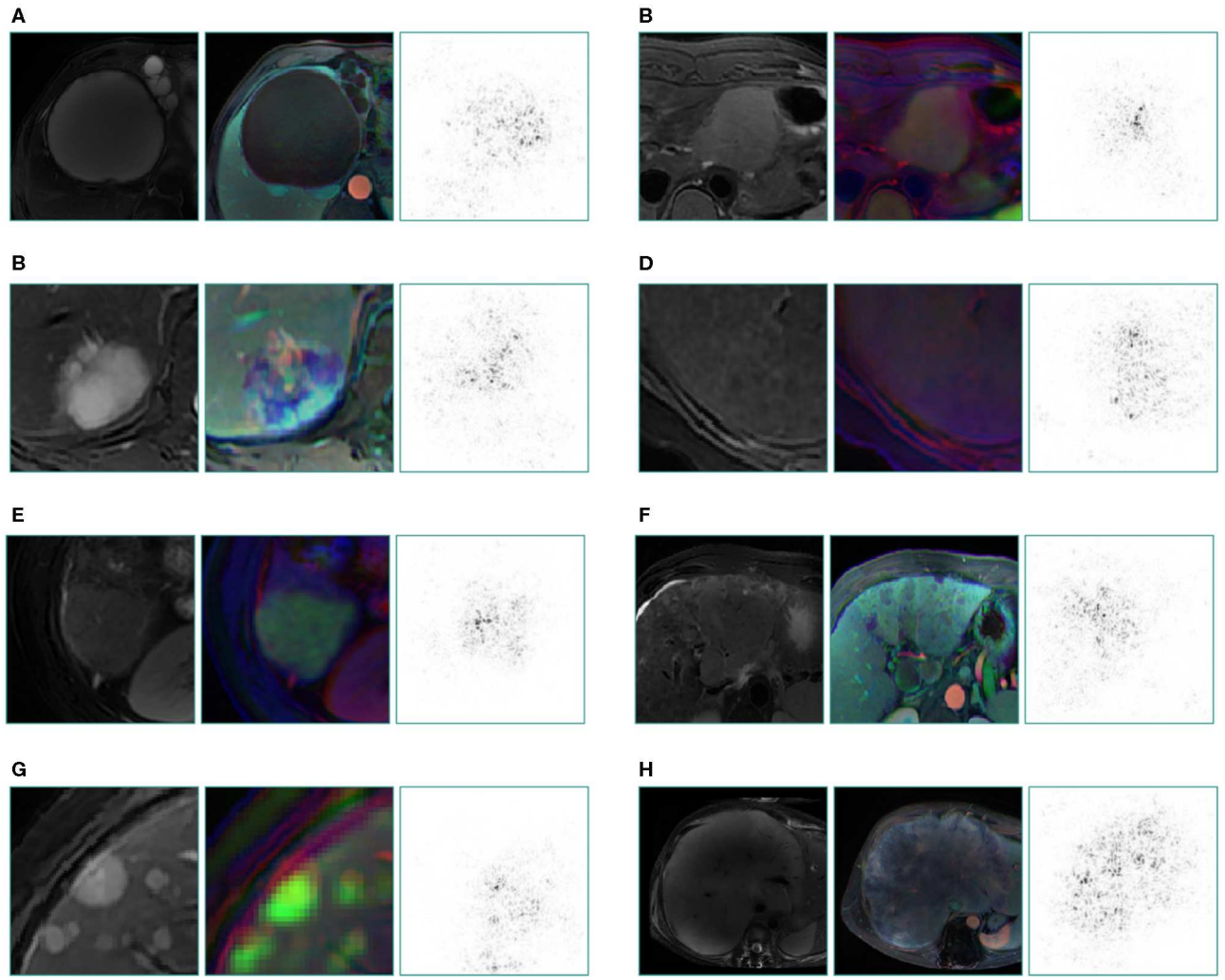


FIGURE 4 | Saliency map for example images from seven categories of the validation set and a special case which not appeared in training set. These maps reveal the pixels that most influence a CNN's prediction. Saliency maps show the pixel gradients with relative to the CNN loss function. Darker pixels represent pixels with greater influence. Clear correlation between lesions and saliency maps are revealed. We selected T2 image as the original control, the middle is reconstructed image of three sequences (the left column is from three plain scan sequences, the right column is from three enhanced sequences), the right is a corresponding saliency map. **(A)** cyst, **(B)** FNH, **(C)** hemangioma, **(D)** benign nodule, **(E)** HCC, **(F)** primary adenocarcinoma, **(G)** metastatic malignancy originating from pancreas, **(H)** malignant fibrous histiocytoma, which still gains a good display although this rare type did not appear in the training set.

complexity. From t-SNE visualization of the last hidden layer representations of Model G and Model E (**Figures 3B,C**), clusters of points which belonged to the same clinical category were better gathered together in Model G than Model E, therefore, this also demonstrated that the modified end-to-end CNN with unenhanced images and clinical data as collaborative inputs achieves better classification performances than the network using enhanced images.

Consistency Evaluation Between Models/Radiologists and Gold Standard

The agreement was then measured comparing deep-learning models, and the radiologists' consensus with the pathological/formal reports using Cohen's Kappa statistic (**Table 4**). It was observed that the agreement of all deep-learning models and radiologists with pathological/formal reports reached statistical significance ($P < 0.01$, estimated by the two-sample two-tailed z-test score), indicating the consistency between them. According to the interpretation Cohen suggested about the Kappa results (34), Model A had substantial consistency (kappa > 0.6), while Model C, Model D, Model E, and Model F had almost perfect consistency compared with the gold standard(kappa > 0.8). Regarding time spent, it can take a radiologist several minutes to analyze a patient's imaging depending on the difficulty of each individual, but deep-learning models just need a few seconds.

TABLE 2 | Diagnostic performance of binary classifiers and radiologists in the validation set.

	Value (95%CI)			McNemar's *P*-value		Delong *P*-value
	Model C	**Model D**	**Reader consensus**	**Model C vs. Reader**	**Model D vs. Reader**	**Model C vs. Model D**
AUC	0.951 (0.919, 0.982)	0.9416 (0.914, 0.979)				0.664
Sensitivity, %	91.9 (84.7, 96.5)	90.9 (83.4, 95.8)	89.1 (76.4, 96.4)	0.375	0.219	
Specificity, %	94.1 (87.6, 97.8)	94.1 (87.6, 97.8)	90.4 (79.0, 96.8)	0.549	0.754	

TABLE 3 | Diagnostic performance of three-way classifiers and radiologists in the malignancy validation set.

	Value (95%CI)				McNemar's *P*-Value[a]	Delong *P*-Value	
	Model E	**Model F**	**Model G**	**Reader consensus**	**Model G vs. Reader**	**Model E vs. Model G**	**Model F vs. Model G**
HCC							
AUC	0.879 (0.808, 0.949)	0.972 (0.938, 1.000)	0.951 (0.906, 0.997)			0.002	0.792
Sensitivity, %	93.6 (82.5, 98.7)	95.7 (85.5, 99.5)	95.7 (85.5, 99.5)	89.1 (76.4, 96.4)	0.289		
Specificity, %	67.3 (52.9, 79.7)	96.2 (86.8, 99.5)	90.4 (79.0, 96.8)	90.4 (79.0, 96.8)	0.063		
Metastatic malignancy							
AUC	0.814 (0.722, 0.907)	0.980 (0.947, 1.000)	0.985 (0.958, 1.000)			0.0002	0.403
Sensitivity, %	59.5 (42.1, 75.3)	100 (90.5, 100)	94.6 (81.8, 99.3)	89.2 (74.6, 97.0)	0.688		
Specificity, %	93.6 (84.3, 98.2)	96.8 (88.8, 99.6)	100 (94.2, 100)	95.1 (86.3, 99.0)	1.000		
Primary malignancy except HCC							
AUC	0.761 (0.613, 0.909)	0.989 (0.951, 1.000)	0.905 (0.801, 1.000)			0.008	0.081
Sensitivity, %	53.3 (26.6, 78.7)	86.7 (59.5, 98.3)	73.3 (44.,9 92.2)	60.0 (32.3, 83.7)	0.688		
Specificity, %	95.2 (88.3, 98.7)	100 (95.7, 100)	96.4 (89.9, 99.3)	91.6 (83.4, 96.5)	0.250		

[a] *P value was calculated between Model G (three-sequence images + clinical data) vs. Reader consensus using the McNemar's test.*

TABLE 4 | Consistency analysis of models and radiologists compared with pathological or formal report.

	Accuracy	**Kappa**	**Z score**	***P*-Value**
Model A	79.1 (72.8, 84.5)	0.744	22.9	<0.01
Model B	71.1 (64.4, 77.3)	0.649	20.0	<0.01
Model C	93.5 (89,2, 96.5)	0.861	12.2	<0.01
Model D	93.0 (88.6, 96.1)	0.851	12.1	<0.01
Model E	72.7 (62.9, 81.2)	0.541	7.31	<0.01
Model F	93.9 (87.3, 97.7)	0.901	11.8	<0.01
Model G	91.9 (84.7, 96.5)	0.867	11.4	<0.01
Reader 1	88.1 (82.8, 92.2)	0.853	25.7	<0.01
Reader 2	77.1 (70.7, 82.7)	0.723	22.8	<0.01
Reader 3	84.6 (78.8, 89.3)	0.810	24.4	<0.01
Reader consensus	86.1 (80.5, 90.5)	0.829	25.1	<0.01

DISCUSSION

The findings of the current study show the feasibility and potential superiority of the integrated DLS in liver tumor MRI diagnosis in clinical situations. Deep learning achieves a performance on par with experienced radiologists in classifying liver tumors to seven categories. Utilizing unenhanced sequences, DLS can distinguish malignant from benign tumors, and then by combining medical texts and laboratory test results lead to a precise diagnosis for malignant tumors, even better than three experienced radiologists who have considered enhanced sequences. To the best of our knowledge, this is the largest study in the field of deep-learning-guided liver tumor diagnosis based on MR images worldwide, which has the most variable types of focal liver lesions (only inflammatory lesions excluded).

Evaluations of MR images by radiologists are generally subjective and are possibly influenced by their experience to an extent (40), even in LI-RADS which is majorly applicable to patients at high risk for HCC. Deep learning models have advantages in overcoming these problems. CNNs learning feature representations with an automated procedure and the interpretation maintains consistency and, therefore, diagnostic reproducibility. Thus, for developing countries such as China or other undeveloped countries, where there is an unbalanced distribution of medical resources between urban and rural areas, the deep-learning models could help in bridging the diagnosis gap of MRI between national hospitals and primary care hospitals, which can also be served as a quick and reliable opinion for junior radiologists in the diagnosis of hepatic lesions.

The confusion matrix (**Figure S2**) shows that both the seven-way classifier and doctors have poor performances in malignant classification, especially with HCC and other primary malignancies that are difficult to distinguish from each other, which may be because that intrahepatic cholangiocarcinoma have some similar imaging manifestations with HCC. It is worth noting that some rare types, such as malignant fibrous histiocytoma which belongs to other primary malignancies, have not been fully studied in the training set due to a lack of sufficient cases. The same situation exists with metastatic tumors with different origins such as pancreatic neuroendocrine cancer, lung squamous cell carcinoma, and so on. The inclusion of these individuals to the study reduced the diagnosis efficacy of the model, however, considering the clinical application scenarios, we did not exclude these individuals. After adding clinical data, the performance of three-way CNNs are greatly accelerated with an accuracy over 90%, which is better than radiologists, even in the model utilizing unenhanced sequences. Moreover, the binary CNN with unenhanced images exhibits almost the same performance with that using enhanced sequences and the prediction accuracy for seven RN/DN and HCC achieved 100% and over 90%, respectively. These results indicated that deep learning can mine more information in unenhanced sequences and clinical data to make judgments than human experts. Therefore, CNN has the potential ability to use non-contrast MRI to make diagnoses of liver tumors, even for high-risk cirrhotic nodules. This advantage could protect patients from potential gadolinium unsafety, especially for those allergic to contrast agents or those who cannot tolerate it with liver or kidney failure. Considering the high expenditure of gadolinium and hepatobiliary contrast agents, non-contrast MRI could be served as a potential cost-effective screening and surveillance tool (12) for high-risk patients under the assistance of DLS. More patients with cirrhotic nodules and small HCCs will be included in future multi-center prospective collaborative research and the DLS will be further validated.

In addition, two false negative cases in puncture biopsies were selected to be validated separately, which were not included in the training or test set. Combined with the patient's medical history and blood test results, they were still considered as malignant and were verified in subsequent treatment (**Figure S5**). But the seven-way CNNs and binary CNNs suggested they were malignant. Image-based features represent the phenotype of the entire tumor in three dimensions and not just the portion that was punctured for pathological testing (11, 17), and thereby can assess the condition of the entire lesion and yield more comprehensive and accurate results (41), which is another advantage of CNN-based image recognition.

Many previous studies have gained impressive achievements in medical image classification using deep neural networks, such as in the classification of skin cancer (20), lung cancer screening (21), diabetic retinopathy examination (22), liver fibrosis assessment (42–44), etc. However, owing to the diversity and complexity of liver masses, there are only a few studies (16) that applied deep learning for multi-classification of liver tumors using CT (23) and MRI (16, 24, 25). Hamm et al. (16) selected six common specific hepatic lesion types ($n = 494$ patients) and utilized enhanced sequences to train the model. The model was not validated in an independent set. The study of Yamashita et al. (25) was also based on enhanced sequences in a small dataset. Whereas, our study evaluated the value of non-enhanced sequences (T2, diffusion, T1 pre-contrast) and achieved good performances in a binary model, which has the potential to reduce high costs and the risk of contrast-related side effects. Moreover, different from all previous studies, we modified the input layer to receive a variety of data input and used the concatenate operation to combine image features with clinical data features, and then classified liver tumors through a fully connected layer. This end-to-end deep learning model with clinical data and images can fully utilize comprehensive information to improve diagnostic performance.

The DLS is applicable for patients with all types of liver tumors, except inflammatory lesions. The MR images used in the study were gained from different MRI scanners and acquisition protocols, which contributed to increased data diversity and heterogeneousness in training the algorithm and demonstrated the robustness of models. Therefore, once the DLS is established, radiologists just need to perform a standardized selection of ROI for liver tumors on a T2 sequence in the daily workflow of MRI reading to conduct such analysis, which is extremely convenient and timesaving for clinical applications. We are working on constructing a cloud-based multi-institutional artificial intelligence platform and a user-friendly website to provide freely accessible telemedical assistance for clinicians to accelerate the interpretation of MR images, meanwhile, to collect the related follow-up pathological information and feed it back to DLS so as to continuously improve its performance.

Our study has several limitations, which should be acknowledged. First, the study is a single-center retrospective study, although the validation set has maximally simulated the scenario of a clinical practice, multicenter prospective research is still necessary to evaluate performance in a real-world, clinical setting. Second, more patients with some specific types of focal liver diseases (RN, DN, small HCC, HCC without pathology, inflammation, etc.) need to be included in future training, in order to be applicable across the full distribution and spectrum of lesions encountered in a clinical practice. Third is a potential problem for medical applications about the interpretability of "black box" algorithms (45). In the current study, darker pixel regions in the saliency map and clusters of point clouds in t-SNE revealed that CNN prediction at least follows some aspects of human expert knowledge, which can be seen as an application of interpretable deep-learning on multimodal medical data, in addition, research into explainable AI and evaluation of interpretability is occurring at a rapid pace (46, 47).

In the future, we would ideally include some types of less common liver masses, such as abscesses, adenomas, rare malignancies, etc. Importantly, further high-quality, prospective, multicenter studies will also be performed, especially for high-risk patients with cirrhosis. For these patients, the main differentials for small HCC are benign regenerative/ dysplastic nodules and pre-malignant nodules. These are quite challenging clinically. This is an inherent problem in retrospective liver nodule research as most of these will not be confirmed

histologically, however many of these nodules when followed turn out to have been early malignancy. Therefore, early diagnosis of these premalignant nodules may be the most important value that deep learning networks can offer over human experts, which would have to be explored by a rigorously designed prospective study. On the other hand, we will explore the utilization of more ideal data visualization tools (46–48) to allow further degrees of visual understanding of how algorithms make decisions, through identifying relevant imaging features and showing where these features are found on an image, so as to make the "black box" more transparent.

CONCLUSIONS

In summary, we have developed a deep learning-based system which can supply a reliable and timesaving assisted diagnosis in a clinical setting by classifying liver tumors on MRI to seven categories with high accuracy. Meanwhile, it can use non-enhanced MRI to distinguish malignant tumors from benign tumors, and, after adding clinical data, it can provide accurate classification and diagnosis for malignant tumors, which could avoid contrast-related side effects and reduce costs. The DLS was trained with data in a various acquisition condition, and this classification system covers most types of liver tumors, which is unprecedented. All of these suggested a good potential of DLS for clinical generalization. Further prospective multicenter studies in larger patient populations and high-risk cirrhosis patients are still needed.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Independent institutional review boards of Sir Run Run Shaw Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

SZ, WL, ZJ, JS, and YW created the datasets, interpreted the data, outlined ROI, and defined the clinical labels. MC, YJ, YY, and JQ developed the network architecture and data/modeling infrastructure, training, and testing setup. SZ and MC created the figures, wrote the methods, and performed additional analysis requested in the review process, and performed statistical analysis. SZ wrote the manuscript. XC, JL, and ZW provided clinical expertise and guidance on the study design. HL and YT advised on the modeling techniques. XC and HL supervised the project.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (61672452 to HL) and National Natural Science Foundation of China (81827804 to XC).

ACKNOWLEDGMENTS

We thank team members Pro. Hu and Pro. Sun from the radiology department of Sir Run Run Shaw hospital for their assistance in collecting data for this study, outlining ROI, and labeling images. We are also grateful to team member Pro. Lin from State Key Laboratory of CAD&CG, Zhejiang University for their suggestions to the modeling techniques and scientific visualization.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00680/full#supplementary-material

Figure S1 | Flowchart of the procedures in the selection of training and validation set.*Lesion-related treatment included surgery, transcatheter arterial chemoembolization, radiofrequency ablation, chemotherapy, radiotherapy, targeted drug therapy, etc. Number of patients with typical malignant imaging performance and in advanced stages without surgical indications, usually received therapy directly without biopsy. Reasons for exclusion included a unified method for preprocessing resulted in some images not being of a good enough quality owing to different parameters in multiple sequences of multiple machines from different manufacturers. For patients in the follow-up, 82 individuals were reviewed by ultrasound or radiology inspection in the following 6–10 months from the current examination date, and the diagnosis remains consistency.

Figure S2 | Confusion matrix comparison between CNNs and radiologists. These show the number of patients for each class of gold standard and predicted class. The number of patients who were correctly predicted are found on the diagonal. **(A)** Confusion matrices for Model A and radiologists for the seven-way classification task in the validation set reveal similarities in misclassification between human experts and the CNN; confusion matrices for Model B, which performance is inferior to Model A. **(B)** Confusion matrices for Model E,F,G and radiologists for the three-way classification task in new validation set reveal the CNN with clinical data achieve higher performance than experienced experts.

Figure S3 | Example HCC false positive cases. (i), Epithelioid angiomyolipoma. (ii), Bile duct adenoma. **(A)** T2WI **(B)** DWI **(C)** Precontrast-T1WI **(D)** late arterial phase **(E)** portal venous phase **(F)** equilibrium phase.

Figure S4 | Example HCC false negative cases. (i). Hepatocellular adenoma with carcinogenesis. (ii), small hepatocellular carcinoma. **(A)** T2WI **(B)** DWI **(C)** Precontrast-T1WI **(D)** late arterial phase **(E)** portal venous phase **(F)** equilibrium phase.

Figure S5 | Example false negative case in puncture biopsy report, actually positive in clinical assessment and verified in subsequent treatment. **(A–E)** T2-Weighted image of a lesion in the same section **(A)** Jan 14, 2017, histopathology report after biopsy showed: Nodular cirrhosis. But combined with history and tumor indicators, it was still considered that the risk of malignancy was extremely high and TACE was given. **(B,C)**,The second and third TACE were given in Feb 23, 2017 and May 18, 2017, the lesion shrunk enough to almost disappear. **(D,E)**, The lesion enlarged by the follow-up on Nov 21,2017 and Jan 18, 2018, which suggested it was a malignant tumor. **(F,G)** T2-Weighted image of the lesion of another case. Histopathology report after biopsy showed: Fibrous tissue hyperplasia, a small amount of shed atypical cells were seen. MRI report showed:

Liver cancer with intrahepatic metastasis, left and right intrahepatic portal vein tumor thrombus. With comprehensive consideration, this case is a malignant tumor, but the local biopsy leads to a negative result because of partial necrosis of the lesion.

Table S1 | Baseline characteristics of the training and validation set.

Table S2 | Disease distribution of each category in the training and validation cohort.

Table S3 | Medical text and laboratory data coding table.

Table S4 | Imaging parameters in various sequences of magnetic resonance imaging.

REFERENCES

1. European Association for the Study of the Liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol.* (2018) 69:182–236. doi: 10.1016/j.jhep.2018.03.019
2. Forner A, Vilana R, Ayuso C, Bianchi L, Solé M, Ayuso JR, et al. Diagnosis of hepatic nodules 20 mm or smaller in cirrhosis: prospective validation of the noninvasive diagnostic criteria for hepatocellular carcinoma. *Hepatology.* (2008) 47:97–104. doi: 10.1002/hep.21966
3. DeSantis CE, Siegel RL, Sauer AG, Miller KD, Fedewa SA, Alcaraz KI, et al. Cancer statistics for African Americans, 2016: progress and opportunities in reducing racial disparities. *CA Cancer J Clin.* (2016) 66:290–308. doi: 10.3322/caac.21340
4. Sherman M. Surveillance for hepatocellular carcinoma. *Best Pract Res Clin Gastroenterol.* (2014) 28:783–93. doi: 10.1016/j.bpg.2014.08.008
5. Singal A, Volk ML, Waljee A, Salgia R, Higgins P, Rogers MA, et al. Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. *Aliment Pharmacol Ther.* (2009) 30:37–47. doi: 10.1111/j.1365-2036.2009.04014.x
6. Omata M, Cheng AL, Kokudo N, Kudo M, Lee JM, Jia J, et al. Asia-Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. *Hepatol Int.* (2017) 11:317–70. doi: 10.1007/s12072-017-9799-9
7. An C, Rakhmonova G, Choi JY, Kim MJ. Liver imaging reporting and data system (LI-RADS) version 2014: understanding and application of the diagnostic algorithm. *Clin Mol Hepatol.* (2016) 22:296–307. doi: 10.3350/cmh.2016.0028
8. Khawaja AZ, Cassidy DB, Al Shakarchi J, McGrogan DG, Inston NG, Jones RG. Revisiting the risks of MRI with Gadolinium based contrast agents-review of literature and guidelines. *Insights Imaging.* (2015) 6:553–8. doi: 10.1007/s13244-015-0420-2
9. Rogosnitzky M, Branch S. Gadolinium-based contrast agent toxicity: a review of known and proposed mechanisms. *Biometals.* (2016) 29:365–76. doi: 10.1007/s10534-016-9931-7
10. Gulani V, Calamante F, Shellock FG, Kanal E, Reeder SB, International Society for Magnetic Resonance in Medicine. Gadolinium deposition in the brain: summary of evidence and recommendations. *Lancet Neurol.* (2017) 16:564–70. doi: 10.1016/S1474-4422(17)30158-8
11. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet.* (2018) 391:1301–14. doi: 10.1016/S0140-6736(18)30010-2
12. Kim HA, Kim KA, Choi JI, Lee JM, Lee CH, Kang TW, et al. Comparison of biannual ultrasonography and annual non-contrast liver magnetic resonance imaging as surveillance tools for hepatocellular carcinoma in patients with liver cirrhosis (MAGNUS-HCC): a study protocol. *BMC Cancer.* (2017) 17:877 doi: 10.1186/s12885-017-3819-y
13. Sia D, Villanueva A, Friedman SL, Llovet JM. Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterology.* (2017) 152:745–61. doi: 10.1053/j.gastro.2016.11.048
14. Venkatesh SK, Chandan V, Roberts LR. Liver masses: a clinical, radiologic, pathologic perspective. *Clin Gastroenterol.* (2014) 12:1414–29. doi: 10.1016/j.cgh.2013.09.017
15. Mitchell DG, Bruix J, Sherman M, Sirlin CB. LI-RADS (Liver Imaging Reporting and Data System): summary, discussion, and consensus of the LI-RADS Management Working Group and future directions. *Hepatology.* (2015) 61:1056–65. doi: 10.1002/hep.27304
16. Hamm CA, Wang CJ, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. *Eur Radiol.* (2019) 29:3338–47. doi: 10.1007/s00330-019-06205-9
17. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin.* (2019) 69:127–57. doi: 10.3322/caac.21552
18. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539
19. Greenspan H, van Ginneken B, Summers RM. Deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging.* (2016) 35:1153–9. doi: 10.1109/TMI.2016.2553401
20. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* (2017) 542:115–8. doi: 10.1038/nature21056
21. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med.* (2019) 25:954–61. doi: 10.1038/s41591-019-0447-x
22. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA.* (2017) 318:2211–23. doi: 10.1001/jama.2017.18152
23. Yasaka K, Akai H, Abe O, Kiryu S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology.* (2018) 286:899–908. doi: 10.1148/radiol.2017170706
24. Trivizakis E, Manikis GC, Nikiforaki K, Drevelegas K, Constantinides M, Drevelegas A, et al. Extending 2-D convolutional neural networks to 3-D for advancing deep learning cancer classification with application to MRI liver tumor differentiation. *IEEE J Biomed Health Inform.* (2019) 23:923–30. doi: 10.1109/JBHI.2018.2886276
25. Yamashita R, Mittendorf A, Zhu Z, Fowler KJ, Santillan CS, Sirlin CB, et al. Deep convolutional neural network applied to the liver imaging reporting and data system (LI-RADS) version 2014 category classification: a pilot study. *Abdom Radiol.* (2019) 45:24–35. doi: 10.1007/s00261-019-02306-7
26. Szegedy C, Liu W, Jia YQ, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: *IEEE Conference on Computer Vision and Pattern Recognition (Cvpr)*. Boston, MA (2015). p. 1–9.
27. He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (Cvpr)*. Las Vegas, NV (2016). p. 770–8.
28. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: *Thirty-First AAAI Conference on Artificial Intelligence*. San Francisco, CA (2017). p. 4278–84.
29. Russakovsky O, Deng J, Su H, J.Krause, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis.* (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y
30. Shao L, Zhu F, Li X. Transfer learning for visual categorization: a survey. *IEEE Trans Neural Netw Learn Syst.* (2015) 26:1019–34. doi: 10.1109/TNNLS.2014.2330900
31. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. *arXiv:1603.04467 [cs.DC]*. (2016).
32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* (2011) 12:2825–30.
33. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika.* (1934) 26:404–13. doi: 10.1093/biomet/26.4.404

34. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med.* (2012) 22:276–82. doi: 10.11613/BM.2012.031
35. Reed JF III. Homogeneity of kappa statistics in multiple samples. *Comput Methods Programs Biomed.* (2000) 63:43–6. doi: 10.1016/s0169-2607(00)00074-2
36. van der Maaten L. Accelerating t-SNE using tree-based algorithms. *J Mach Learn Res.* (2014) 15:3221–45.
37. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualizing image classification models and saliency maps. *arXiv:1312.6034v2 [cs.CV].* (2014).
38. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, et al. A deep learning model to predict a diagnosis of Alzheimer disease by using (18)F-FDG PET of the brain. *Radiology.* (2019) 290:456–64. doi: 10.1148/radiol.2018180958
39. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med.* (2015) 13:1 doi: 10.1186/s12916-014-0241-z
40. Blachar A, Federle MP, Ferris JV, Lacomis JM, Waltz JS, Armfield DR, et al. Radiologists' performance in the diagnosis of liver tumors with central scars by using specific CT criteria. *Radiology.* (2002) 223:532–9. doi: 10.1148/radiol.2232010801
41. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
42. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut.* (2019) 68:729–41. doi: 10.1136/gutjnl-2018-316204
43. Choi KJ, Jang JK, Lee SS, Sung YS, Shim WH, Kim HS, et al. Development and validation of a deep learning system for staging liver fibrosis by using contrast agent-enhanced CT Images in the liver. *Radiology.* (2018) 289:688–97. doi: 10.1148/radiol.2018180763
44. Xue LY, Jiang ZY, Fu TT, Wang QM, Zhu YL, Dai M, et al. Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis. *Eur Radiol.* (2020) 25:342–52. doi: 10.1109/TVCG.2018.2864812
45. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* (2019) 17:195. doi: 10.1186/s12916-019-1426-2
46. Ming Y, Qu H, Bertini E. RuleMatrix: visualizing and understanding classifiers with rules. *IEEE Trans Vis Comput Graph.* (2018) 25:342–52. doi: 10.1109/TVCG.2018.2864812
47. Wang CJ, Hamm CA, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features. *Eur Radiol.* (2019) 29:3348–57. doi: 10.1007/s00330-019-06214-8
48. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *IEEE I Conf Comp Vis.* (2017) 618–26. doi: 10.1109/ICCV.2017.74

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 14 July 2020
doi: 10.3389/fmed.2020.00281

Utility of [^{18}F] Fluoro-Deoxyglucose Positron Emission Tomography/Computed Tomography for Staging and Therapy Response Evaluation in Pediatric Rhabdomyosarcoma: A Case Series and Literature Review

*Ri Sa[1], Danyan Liu[2], Hongguang Zhao[1], Sen Hou[1], Qiuyu Lin[1] and Feng Guan[1]**

[1] Department of Nuclear Medicine, The First Hospital of Jilin University, Changchun, China, [2] Department of Radiology, The First Hospital of Jilin University, Changchun, China

OPEN ACCESS

Edited by:
Francesco Rundo,
STMicroelectronics, Italy

Reviewed by:
Domenico Albano,
University of Brescia, Italy
Gaurav Malviya,
University of Glasgow,
United Kingdom

***Correspondence:**
Feng Guan
guanfeng1972@163.com

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: 13 March 2020
Accepted: 21 May 2020
Published: 14 July 2020

Citation:
Sa R, Liu D, Zhao H, Hou S, Lin Q and Guan F (2020) Utility of [^{18}F] Fluoro-Deoxyglucose Positron Emission Tomography/Computed Tomography for Staging and Therapy Response Evaluation in Pediatric Rhabdomyosarcoma: A Case Series and Literature Review. Front. Med. 7:281. doi: 10.3389/fmed.2020.00281

Background: The role of [^{18}F] fluoro-deoxyglucose [[^{18}F] FDG] positron emission tomography (PET)/computed tomography (CT) in pediatric rhabdomyosarcoma (RMS) is not well-established. This manuscript explores the role of staging and therapy response evaluation of PET/CT in a series of patients with RMS.

Methods: Thirteen consecutive patients with pathologically proven RMS underwent baseline PET/CT scan and a second PET/CT for evaluation of therapy response. Maximum standardized uptake value (SUV_{max}), mean standardized uptake value (SUV_{mean}), highest standardized uptake peak value (SUV_{peak}), metabolic tumor volume (MTV), and total lesion glycolysis (TLG) were obtained from baseline PET/CT and were used as potential predictors for evaluation of metabolic treatment response.

Results: On baseline PET/CT, most RMSs are located in the pelvic cavity, and upper arms ranked second. The primary lesions were large and showed invasion to the surrounding tissues. Lymph node metastases were seen in eight patients, and eight patients showed distant metastasis to the lung, liver, and bone. The median SUV_{max}, SUV_{mean}, and SUV_{peak} of primary sites were 7.1, 4.0, and 5.9, respectively. The median MTV and TLG were 196.6 cm^3 and 780.2, respectively. After therapy, six patients received complete metabolic response (CMR) and non-CMR occurred in seven patients on the second PET/CT. SUV_{max}, SUV_{peak}, MTV, and TLG in patients with CMR were significantly lower than those in patients with non-CMR.

Conclusions: Primary sites and metastatic lesions of RMS demonstrate increased glycolytic activity, which may allow them to be imaged using [^{18}F] FDG PET/CT. Metabolic parameters derived from the baseline PET/CT have potential value for predicting CMR to therapy in pediatric RMS.

Keywords: PET/CT, rhabdomyosarcoma, standardized uptake value (SUV), metabolic tumor volume (MTV), total lesion glycolysis (TLG)

INTRODUCTION

Rhabdomyosarcoma (RMS) is the most common soft tissue sarcoma in children and adolescents, but the disease is still rare, with about 400 newly diagnosed cases each year in Europe and a similar incidence in the USA (1). RMS is a high-grade, malignant neoplasm in which the current therapy strategies for RMS are multimodal approach, comprising surgical resection, chemotherapy, and/or radiation therapy in order to control the primary tumor to the utmost (2). The chance of cure with widely metastatic and recurrent disease of RMS is very low, and patients experience months of intensive, multifaceted therapies that can bring life-threatening acute toxicities and, in some cases, life-changing late effects (3).

The potential factors which are associated with ultimate outcome of RMS are controversial (4, 5). Nevertheless, it is well-accepted that the favorable status at the end of therapy is crucial to promising outcome of RMS. By contrast, residual mass in patients with RMS at the end of therapy had a significant impact on failure-free survival (6). Thus, identifying predictors that can select therapy-sensitive tumors is considered as an important step to manage treatment strategies to avoid poor outcome. [^{18}F] fluoro-deoxyglucose [[^{18}F] FDG] positron emission tomography (PET)/computed tomography (CT) is one of the most advanced multimodal techniques available. [^{18}F] FDG PET/CT can accurately detect the extent and metabolic activity of tumor lesions in patients with malignant tumors, aiding staging and evaluation of therapy response in many malignant tumors. However, to date, only several retrospective studies qualitatively interpreted the images to demonstrate the potential role of [^{18}F] FDG PET/CT in RMS but failed to include volumetric [^{18}F] FDG uptake measurements such as metabolic volume and lesion glycolysis. Therefore, the aim of this study was to demonstrate the utility of metabolic parameters derived from baseline PET/CT in staging and evaluation of therapy response of RMS.

PATIENTS AND METHODS

Patients

We enrolled the patients with the following inclusion criteria: histopathologically confirmed RMS from January 2008 to November 2018; under age of 18 years old. Those with other malignant tumors beyond RMS were excluded.

[^{18}F] Fluoro-deoxyglucose PET/CT

Baseline PET/CT was performed before therapies, and the second PET/CT for evaluating therapy response was performed within 4 weeks after antitumor therapy. PET/CT scans including semiquantitative and metabolic volumes measurement were reviewed by two nuclear physicians who were unaware of the results of any other imaging tests and the clinical data by consensus.

Patients were asked to fast for 6 h. [^{18}F] FDG (4.07 MBq/kg) with a radiochemical purity of >95% (Sumitomo Corporation, Japan) was given through intravenous injection, and the patients were asked to rest for approximately 60 min after the injection. Four children (age ≤3 years old) who could not give collaboration during the tests were sedated after injection of [^{18}F] FDG immediately prior to commencement of imaging. The PET/CT scan was performed using a hybrid PET/CT scanner (Biograph 16HR; Siemens, Germany). Here, all [^{18}F] FDG PET/CTs were performed with a standard acquisition and reconstruction protocol. Quality assurance and quality control procedures for the PET system were carried out accurately on a daily basis. The scan range was from the skull to the upper part of mid-thigh or feet. A low-dose CT protocol (100 mAs, 140 kV, tube rotation time of 0.5 per rotation, pitch of 6, slice thickness of 5 mm, and shallow breathing) was first applied and followed by the PET scan (3-min emission scan per table position).

In each examination, the representative lesion was defined as a single lesion with the highest FDG uptake, even though patients had multiple active lesions. A region of interest (ROI) was drawn around the representative lesion with boundaries drawn to include the lesion in transaxial, coronal, and sagittal views. Maximum standardized uptake value (SUV_{max}) was calculated from a single voxel exhibiting the SUV_{max} within a representative lesion. Mean standardized uptake value (SUV_{mean}) was the mean value of voxels within the ROI. Highest standardized uptake peak value (SUV_{peak}) was the SUV_{mean} of a 1-cm^3 three-dimensional ROI showing the highest value in the representative lesion. They were obtained in the same area as the pretreatment lesion in case of no abnormal uptake after treatment. Metabolic tumor volume (MTV) and total lesion glycolysis (TLG) were calculated as volumetric parameters in the representative lesion as well as in whole-body lesions. MTV was defined as the volume showing abnormal FDG uptake greater than any parts of the liver in this study. TLG was calculated as the product of MTV multiplied by the SUV_{mean}. The metabolic parameters were measured on the Siemens MIWP workstation (Syngo MIWP; Siemens Medical Solutions, Erlangen, Germany).

Therapy Response Assessment

The SUV_{max}, SUV_{mean}, SUV_{peak}, total MTV, and total TLG were determined as metabolic PET parameters for each patient. The second PET/CT scans were used to classify therapy response into four categories according to the European Organization for Research and Treatment of Cancer (EORTC) criteria: complete metabolic response (CMR, no FDG uptake within the tumor volume), partial metabolic response (PMR, SUV_{max} reduction >25% after treatment), stable metabolic disease (SMD, SUV_{max} increase or decrease <25%), progressive metabolic disease (PMD, SUV_{max} increase >25% or increase in the extension of tumor uptake >20% in the longest dimension or the appearance of new FDG uptake) (7). Patients with PMR, SMD, and PMD were then grouped as non-complete metabolic responders (non-CMR).

Statistical Analysis

Statistical analysis was performed using SPSS software (version 22; IBM). The continuous variables with non-normal distribution are presented as median (range). The categorical variables are reported as number (percentage). The factors that may have affected CMR in RMS were analyzed. Fisher exact tests and χ^2 tests were used to test the significance of categorical data such

as gender (male or female), tumor size (<5.0 cm or ≥5.0 cm), tumor margin (T1, confined to anatomic site of origin, or T2, extension and/or fixative to surrounding tissue), lymph node metastases (yes or no), distant metastases (yes or no), and therapy (chemotherapy only or surgery/RT+ chemotherapy). The non-parametric Mann–Whitney *U*-test was used to compare quantitative data when it was not normally distributed including age, SUV_{max}, SUV_{mean}, SUV_{peak}, MTV, and TLG. All *p*-values were two-sided, and $p < 0.05$ was considered statistically significant in all analyses.

Literature Search

We performed a PubMed/Medline search by using MeSH terms focusing on articles on RMS-related and RMS-specific infections and on use of nuclear imaging with [^{18}F] FDG PET/CT scans. Basic information was collected, including author, journal, year published, and number of patients. Specific data were collected, including the study purpose, recorded parameters, and main findings of PET or PET/CT.

RESULTS

Thirteen consecutive patients with newly diagnosed RMS in our department were enrolled. Clinical characteristics, therapy regimen, treatment response, and outcome of all patients with RMS are shown in **Table 1**. The findings of baseline [^{18}F] FDG PET/CT are shown in **Table 2**.

The age of the 13 patients ranged from 2 months (case 3) to 13 years (case 9), and six of them were female. The diagnosis was established by ultrasonography (US)-guided fine-needle aspiration in all patients before baseline PET/CT. Only one patient (case 9) was histologically confirmed as alveolar RMS (ARMS).

On baseline PET/CT, RMS was most commonly located in the pelvic cavity, while upper arms ranked second. The primary lesions in the majority of patients (except cases 4, 6, and 7) were large, in which the diameter were larger than 5 cm. Primary sites in three patients (cases 2, 4, and 6) were confined to the anatomic site of origin, while primary sites in other patients demonstrated invasion to the surrounding tissues. Lymph node metastases were seen in eight patients (cases 1 and 7–13) while cases 1, 3, 5, and 7–11 showed distant metastasis to the lung, liver, and bone. The median SUV_{max}, SUV_{mean}, and SUV_{peak} of target lesions were 7.1 (3.9–15.0), 4.0 (2.4–7.3), and 5.9 (3.4–9.1), and the median MTV and TLG were 196.6 (13.7–824.8) cm^3 and 780.2 (42.6.1–4,023.3), respectively.

Five patients received surgery before the chemotherapy, while six patients received radiation therapy (RT) before the chemotherapy. After the therapy, all patients received the second PET/CT for evaluation of metabolic treatment response, in which six patients achieved CMR, and non-CMR occurred in seven patients (PMR in three patients, SMD in two patients, and PMD in two patients) on the second PET/CT. All patients with CMR and two patients with PMR on the PET/CT turned to observation, and other patients prolonged or changed the therapy. The median follow-up time was 9.0 months, and two patients (cases 3 and 11) died. The results of [^{18}F] FDG PET/CT of cases 1 and 9 are shown in **Figures 1**, **2**, which were confirmed as CMR and PMR at the second PET/CT, which were performed for evaluation of therapy response, while the two patients received the third PET/CT during the follow-up while case 1 obtained CMR and case 9 suffered from PMD.

Comparisons of the clinical characteristics and metabolic parameters derived from baseline PET in RMS patients regarding therapy response (CMR and non-CMR) are shown in **Table 3**. SUV_{max}, SUV_{peak}, MTV, and TLG in patients with CMR were significantly lower than those in patients with non-CMR.

DISCUSSION

To the best of our knowledge, this is the first study to distinguish CMR from non-CMR in therapy of RMS by testing various metabolic parameters derived from the baseline PET, including SUV_{max}, SUV_{mean}, SUV_{peak}, MTV, and TLG, which would assist the management of treatment and avoid poor outcome. In the current literature, we found several studies have only focused on SUV_{max} to reveal the role of [^{18}F] FDG PET/CT in the field of diagnosis, initial staging, and predicting prognosis of RMS. The findings of these studies are summarized in **Table 4**.

With regard to the staging and restaging, [^{18}F] FDG PET/CT was consistently somewhat better than conventional imaging at identifying unknown primary sites, nodal involvement, and distant metastases, especially bone metastases in RMS (8–12). The primary sites of the RMS are distributed in the whole body, from head to limbs, with relatively large size at diagnosis and poor borders with surrounding tissue, which is not amenable to up-front complete surgical resection of the primary tumor and results in gross residual disease at the initiation of chemotherapy. In our study, the primary lesions only in three patients were limited to the primary sites, and other patients suffered from lymph nodes and/or distant metastases. Lymph node metastases were seen in eight patients in our study, mainly distributing to the neck, mediastinum, hilum, abdomen, and pelvis. In the previous study, approximately 15% of patients with RMS have distant metastases at diagnosis (13), and the molecular hallmark of which is paired box 3 (PAX3) or PAX7 gene fusion with forkhead box protein O1 (FOXO1) (4). According to our study, bone was the most common involved sites in metastatic RMS, and the lung ranked second. One patient (case 1) showed liver metastases in our study, and breast (14, 15) and small bowel metastasis (16) were seen in previous studies.

Complete response at the end of therapy means favorable outcome for malignant tumors. By contrast, patients with non-complete response at the end of therapy predicts poor outcome. Evaluation of therapy response in terms of FDG activity is superior to that based on the radiological performance, which may ignore the activity of tumor. Residual masses at the end of therapy may be reactive scar tissue, mature RMS that occurs as a result of differentiation of RMS cells during

TABLE 1 | Clinical characteristics, treatment response, and outcome in the 13 patients with rhabdomyosarcoma.

No.	Age	Gender (F/M)	Pathological type	Clinical symptom	Therapy	Therapy response	Follow up time	Outcome after the second PET/CT
1	11 y	F	ERMS	Chest pain	RT, chemotherapy	CMR	3.4 y	Observation
2	1 y	M	ERMS	Arm pain	Surgery, chemotherapy	CMR	1.5 y	Observation
3	2 mo	F	ERMS	Facial mass	RT, chemotherapy	SMD	13 mo	Therapy changed, but the patient died
4	7 y	M	ERMS	Abdomen pain	Surgery, chemotherapy	CMR	9 mo	Observation
5	3 y	M	ERMS	Abdomen pain	Chemotherapy	CMR	17mo	Observation
6	3 y	M	ERMS	Mass in hips	Surgery, chemotherapy	CMR	21 mo	Observation
7	12 y	F	ERMS	Mass in toe	RT, chemotherapy	PMR	13 mo	Observation
8	8 y	F	ERMS	Abdomen pain	RT, chemotherapy	PMR	6 mo	Therapy prolonged
9	13 y	M	ARMS	Arm pain	Chemotherapy	PMR	7.2 mo	Observation
10	4 y	M	ERMS	Abdomen pain	RT, chemotherapy	PMD	5.5 mo	Therapy changed
11	4 y	F	ERMS	Abdomen pain	RT, chemotherapy	PMD	6.5 mo	Therapy changed, but the patient died
12	5 y	F	ERMS	Axillary mass	Surgery, chemotherapy	CMR	7.4 mo	Observation
13	11 y	M	ERMS	Mass in hips	Surgery, chemotherapy	SMD	6 mo	Therapy prolonged

y, year; mo, month; F, female; M, male; ARMS, alveolar rhabdomyosarcoma; ERMS, embryonal rhabdomyosarcoma; RT, radiation therapy; CMR, complete metabolic response; PMD, progressive metabolic disease; SMD, stable metabolic disease; PMR, partial metabolic response.

TABLE 2 | Findings of baseline PET/CT in the 13 patients with rhabdomyosarcoma.

No.	Primary site	Size	SUV_{max}	SUV_{mean}	SUV_{peak}	MTV (cm^3)	TLG	T	Lymph node metastases	Distant metastases
1	Mediastinum	b	4.9	3.4	4.1	77.2	229.4	T2	Neck, mediastinum, hilum	Liver, sternum
2	Upper arm	b	5.4	3.8	3.4	37.9	112.3	T1	-	-
3	Maxillofacial	b	7.1	4.3	6.2	25.0	114.1	T2	-	Ilium
4	Intraperitoneal	a	7.2	4.6	5.6	196.6	780.2	T1	-	-
5	Pelvic	b	7.1	4.0	5.9	40.8	160.8	T2	-	Lung
6	Hips	a	4.3	2.6	3.7	13.7	42.6	T1	-	-
7	Toes	a	6.4	3.7	5.9	824.8	2969.3	T2	Neck, mediastinum, hilum, abdomen, pelvic	Rib, ilium
8	Pelvic	b	9.0	4.9	6.4	556.4	3115.8	T2	Neck, mediastinum, abdomen, pelvic	Lung
9	Upper arm	b	9.2	4.3	6.9	264.6	1202.5	T2	Neck, mediastinum, axillary, elbow, hilum	Humerus
10	Pelvic	b	15.0	7.3	9.1	496.7	4023.3	T2	Mediastinum, abdomen, pelvic	Rib, spine, ilium
11	Pelvic	b	12.9	6.9	7.6	319.3	2362.8	T2	Abdomen, pelvic	Lung
12	Axillary	b	3.9	2.4	3.4	46.1	121.4	T2	Axillary	-
13	Hips	b	7.0	3.4	5.5	402.9	1814.1	T2	Pelvic, inguen	-

Size: a, < 5 cm in diameter; b, ≥ 5 cm in diameter; T, tumor; T1, confined to anatomic site of origin; T2, extension and or/fixative to surrounding tissue; -, no metastatic lesions. SUV_{max}, maximum standardized uptake value; SUV_{mean}, mean standardized uptake value; SUV_{peak}, standardized uptake peak value; MTV, metabolic tumor volume; TLG, total lesion glycolysis.

therapy, or residual viable tumor (6). Either scar tissue or mature RMS is non-malignant without and malignant behavior, which would not need additional therapy immediately. The data referring to the role of [^{18}F] FDG PET/CT evaluation of therapy response in RMS are limited (17, 18); worse, the roles of volume-based and intensity-based PET parameters

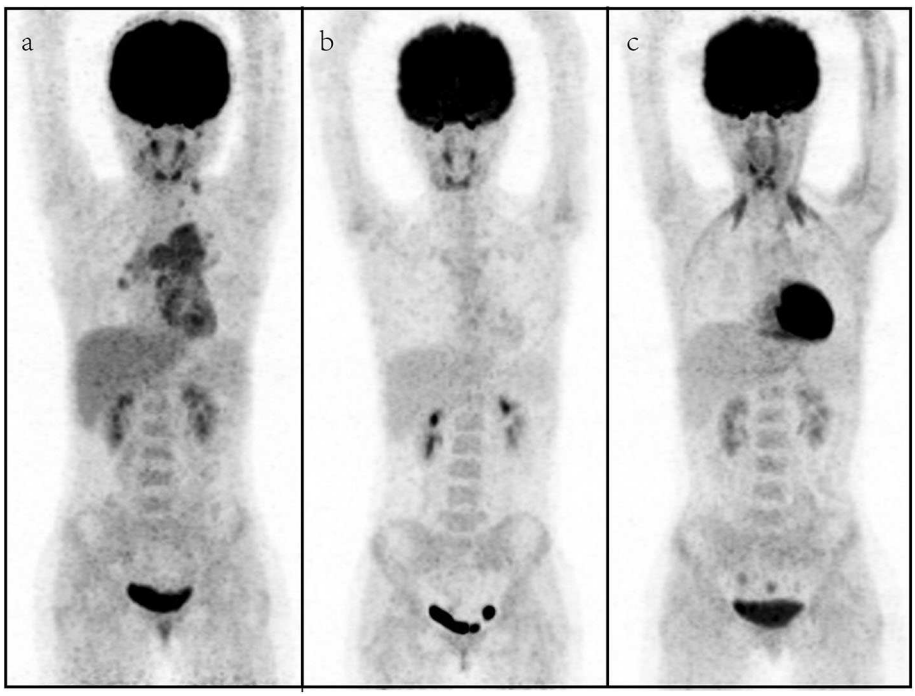


FIGURE 1 | A representative case of an 11-year-old female with rhabdomyosarcoma. Baseline maximum intensity projection (MIP, **a**), showing hypermetabolic disease in laterocervical, mediastinal, and hilar lymph nodes, liver, and sternum [maximum standardized uptake volume (SUV_{max}): 4.9, SUV_{mean}: 3.4, SUV_{peak}: 4.1, metabolic tumor volume (MTV): 77.2 cm^3, total lesion glycolysis (TLG): 229.4]. PET/CT **(b)** after therapy showing a complete metabolic response with no fluoro-deoxyglucose (FDG) uptake in the previous sites. PET/CT **(c)** for follow-up revealed no FDG uptake lesions.

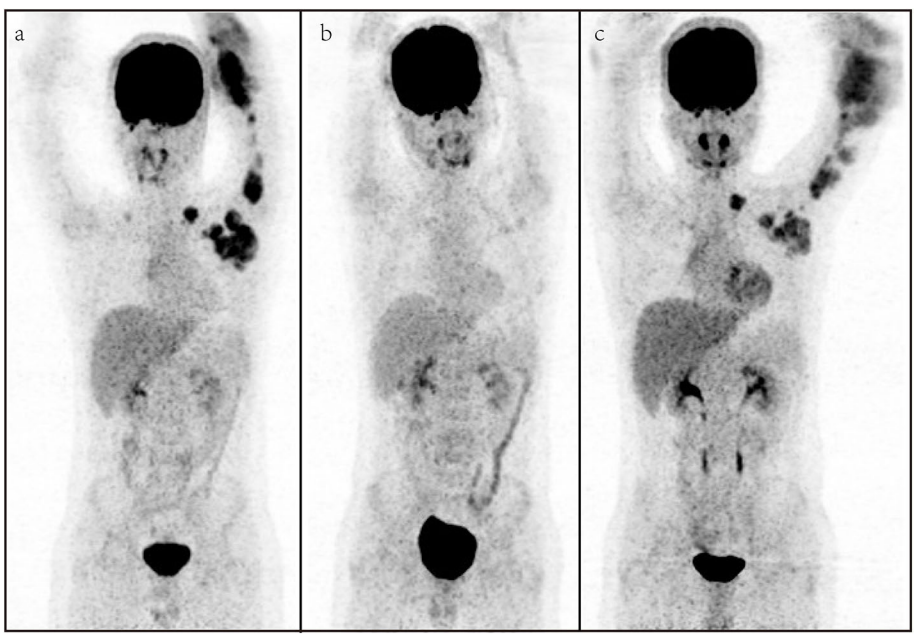


FIGURE 2 | A representative case of a 13-year-old male with rhabdomyosarcoma. Baseline maximum intensity projection (MIP, **a**), showing hypermetabolic disease in laterocervical, mediastinal, axillary, elbow, and hilar lymph nodes and humerus [maximum standardized uptake volume (SUV_{max}): 9.2, SUV_{mean}: 4.3, SUV_{peak}: 6.9, metabolic tumor volume (MTV): 264.6 cm^3, total lesion glycolysis (TLG): 1,202.5]. PET/CT **(b)** after therapy showing a partial metabolic response with fluoro-deoxyglucose (FDG) uptake lymph nodes in the axillary. PET/CT **(c)** for follow-up showing metabolic progression of disease with appearance of new lesions.

TABLE 3 | Comparison of clinical characteristics and metabolic parameters derived from baseline PET in rhabdomyosarcoma patients regarding to the therapy response ($N = 13$).

Patients	Therapy response		Z/x^2 value	p
	CMR ($n = 6$)	Non-CMR ($n = 7$)		
Age-median (range)	4 (1–11 y)	8 (2.0 mo−13 y)	−1.076	0.282
Gender-*n* (%)			0.737	0.592
F	2 (33.3%)	4 (57.1%)		
M	4 (66.7%)	3 (42.9%)		
Size-*n* (%)			0.660	0.559
<5.0 cm	2 (33.3%)	1 (14.3%)		
≥5.0 cm	4 (66.7%)	6 (85.7%)		
SUV_{max}-median (range)	5.2 (3.9–7.2)	9.0 (6.4–15.0)	−2.217	0.027
SUV_{mean}-median (range)	3.6 (2.4–4.6)	4.3 (3.4–7.3)	−1.791	0.073
SUV_{peak}-median (range)	3.9 (3.4–5.9)	6.4 (5.5–9.1)	−2.650	0.008
MTV-median (range, cm^3)	43.5 (13.7–196.6)	402.9 (25.0–824.8)	−2.286	0.022
TLG-median (range)	141.1 (42.6–780.2)	2362.8 (114.1–4023.3)	−2.429	0.015
T –*n* (%)			4.550	0.070
T1	3 (50.0%)	0 (0%)		
T2	3 (50.0%)	7 (100%)		
Lymph node metastases-*n* (%)			3.745	0.103
Yes	2 (33.3%)	6 (85.7%)		
No	4 (66.7%)	1 (14.3%)		
Distant metastases-*n* (%)			3.745	0.103
Yes	2 (33.3%)	6 (85.7%)		
No	4 (66.7%)	1 (14.3%)		
Therapy			0.014	0.906
Chemotherapy only	1 (16.7%)	1 (14.3%)		
Surgery/RT+ chemotherapy	5 (83.3%)	6 (85.7%)		

T1, confined to anatomic site of origin; T2, extension and or/fixative to surrounding tissue. F, female; M, male; T, tumor; T1, confined to anatomic site of origin; T2, extension and or/fixative to surrounding tissue; SUV_{max}, maximum standardized uptake; SUV_{mean}, mean standardized uptake value; SUV_{peak}, standardized uptake peak value; MTV, metabolic tumor volume; TLG, total lesion glycolysis; RT, radiation; CMR, complete metabolic response.

including MTV and TLG were not available for the evaluation of therapy response.

In our study, we observed an inspiring phenomenon that some patients even diagnosed as RMS with multiple lymph nodes and distant metastases would experience a complete response by the end of therapy. By analyses, the clinical characteristics including age, gender, TNM stage, therapy approaches, and SUV_{mean} level were not associated with the therapy response at the end of therapy. SUV_{max}, SUV_{peak}, MTV, and TLG on baseline PET were significant predictors for differentiating CMR from non-CMR. Lower SUV_{max}, SUV_{peak}, MTV, and TLG derived from the baseline PET were more likely to receive CMR after therapy (**Table 3**). SUV_{max} of primary sites of RMS was quite different in previous studies, in which it ranged from 0 to 32 (9, 11, 17–22, 24). High SUV_{max} was more prevalent among patients with less favorable clinical and pathological features including unfavorable primary site, alveolar pathology, presence of regional or distant metastasis, and high-risk group, relating to worse prognosis (23). However, there exists criticism in using SUV to evaluate therapy response, since it is calculated only from one voxel of the ROI. SUV_{max} only represents the highest metabolic activity within tumor lesions and easily affected by noise. SUV_{mean} is calculated from the average SUV value of the entire tumor, and SUV_{peak} is the mean SUV of a 1-cm^3 three-dimensional ROI showing the highest value in the representative lesion. These factors hardly reflect the internal metabolic characteristics of tumor. MTV refers to the volume of tumor tissue, and TLG represents the metabolic activity and metabolic volume of the tumor tissue. Thus, compared to SUV, MTV and TLG have advantages in reflecting the tumor metabolic burden, which could provide a more accurate evaluation of treatment response.

Metabolic parameters derived from baseline PET, such as SUV, MTV, and TLG, are known to be applicable for evaluating therapy response in many malignant tumors. In advanced non-small-cell lung carcinoma, patients with higher values of MTV and TLG had higher probability of disease progression compared to those patients presenting with lower values, while SUV_{max} did not show a correlation with progressive disease (PD) status. MTV also resulted in being significantly different among partial response (PR), stable disease (SD), and PD groups, while SUV_{max} was confirmed to not be associated with response to therapy (25). Besides, some papers have correlated metabolic parameters derived from the baseline PET to metabolic therapy response. In locally advanced cervical cancer, patients with higher values of MTV and TLG had a higher probability of non-CMR compared to those patients presenting with lower values, while SUV_{max} was confirmed to not be associated with response to therapy (26). Similar findings were available in the mantle cell lymphoma (27). However, Voglimacci et al. (28) found that higher cervical SUV_{max} and TLG were significantly associated with poor response to chemoradiotherapy. Nakajo et al. (29) reported that the positive and negative predictive values for non-responders in esophageal cancer treated by chemoradiotherapy were 77 and 69% in MTV and 76 and 100% in TLG, respectively. On balance, SUV and other PET/CT metabolic parameters require further prospective investigation to help tailoring of treatment.

Our study has some limitations: first, the major limitation of our study is the small sample size, including only 13 patients, which limits the statistical power of our data for a definite conclusion and hardly provides cutoff value of differentiating CMR from non-CMR. Therefore, in the future, if enough cases of RMS are available, it would be given. Second, patients in our study did not receive the same treatment approaches, which

TABLE 4 | Overview of published studies on pediatric rhabdomyosarcoma and [^{18}F] FDG PET/CT.

Study	Journal	Year	Image	Patients number	Purpose	Method	SUV_{max} of primary sites	Findings
Ben Arush MW et al. (10); Ricard et al. (11); Dong et al. (19)	J Pediatr Hematol Oncol Clin Nucl Med; Q J Nucl Med Mol Imaging	2006; 2011; 2012	PET or PET/CT	3 cases; 13 cases; 28 cases	Staging and restaging	SUV_{max}	-; 3.7 (range,2.0–6.9); 6.0 (range, 1.4–22.6)	PET/CT can be useful in staging and restaging pediatric RMS, especially for assessing lymph nodes and bone involvement, and for detecting unknown primary sites of RMS, with potential therapeutic strategy alteration.
Klem et al. (20)	J Pediatr Hematol Oncol	2007	PET/CT	31 cases	Staging	Comparison with CIMs; SUV_{max}	6.4 (range, 2.4–12.7)	CIMs was equivocal with PET for the detection of regional or distant spread. Comparing to the final clinical determination of disease, PET was 77% sensitive and 95% specific.
Federico et al. (9); Eugene et al. (17); Tateishi et al. (12); Völker et al. (21)	Pediatr Blood Cancer; Nucl Med Commun; Ann Nucl Med;J Clin Oncol	2013; 2012; 2009;2007	PET/CT; PET	17 cases; 13 cases; 35 cases; 12 cases	Staging	Comparison with CIMs; SUV_{max}	7.2 (range, 2.5–19.2); 6.2 ± 3.8 (range, 2.7–15.4); –; 7.0 ± 3.4	PET/CT performed better than CIM in identifying nodal, bone, bone marrow, and soft tissue disease in children with RMS. CIM remains essential for detection of pulmonary nodules.
Dharmarajan KV et al. (18)	Int J Radiat Oncol Biol Phys	2012	PET	97 cases	Predicting local control	SUV_{max}	7.0 (range, 0–31)	Negative postradiation PET predicted improved LRFS. Negative baseline and preradiation PET findings suggested statistically insignificant trends toward improved LRFS.
Casey DL et al. (22)	Int J Radiat Oncol Biol Phys	2014	PET	107 cases	Predicting outcome	SUV_{max}	8.1 (range, 0–22)	The baseline SUV_{max} (<9.5 vs. ≥9.5) was predictive of PFS and OS, but not LC. The SUV_{max} after induction chemotherapy (<1.5 vs. ≥1.5) was similarly predictive of PFS and was associated with LC and OS. A positive PET after local therapy was predictive of worse PFS, LC, and OS.
El-Kholy et al. (23); Baum et al. (24)	Nucl Med Commun; J Nucl Med	2019; 2011	PET/CT	98 cases; 41 cases	Predicting outcome	Visual analysis; SUV_{max}; SUV_{max}/SUV_{liver}	-	High SUV_{max} was more prevalent among patients with less favorable clinical and pathological features including unfavorable primary site, alveolar pathology, presence of regional or distant metastasis, and high-risk group. Higher SUV_{max} was significantly related to the presence of regional or distant metastasis with worse prognosis.

-, not provided.

SUV_{max}, maximum standardized uptake; RMS, rhabdomyosarcoma; ARMS, alveolar rhabdomyosarcoma; CIMs, conventional imaging modalities; OS, overall survival; PFS, progression-free survival; LC, local control; LRFS, local relapse-free survival.

would influence the final findings of this study. Third, the present study would not be able to provide the relationship of initial FDG uptake and survival because the follow-up periods were too short to reliably assess survival.

In summary, [^{18}F] FDG PET/CT seems to be useful in staging RMS patients, and metabolic parameters extracted from baseline PET/CT have potential value in distinguishing CMR from non-CMR, which is worth to select therapy-sensitive patients. Additional prospective validation studies in a multicenter cohort are required for further understanding of the role of metabolic parameters extracted from baseline PET/CT in RMS.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of The First Hospital of Jilin University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

RS and FG designed the study. RS, DL, and SH participated in patient data acquisition. RS, HZ, and QL analyzed scans and data. RS and FG wrote and revised the draft. All authors read and approved the final manuscript.

REFERENCES

1. Meyer WH. Maintenance chemotherapy in rhabdomyosarcoma: the new standard of care. *Lancet Oncol*. (2019) 20:1476–7. doi: 10.1016/S1470-2045(19)30618-7
2. Chen C, Dorado Garcia H, Scheer M, Henssen AG. Current and future treatment strategies for rhabdomyosarcoma. *Front Oncol*. (2019) 9:1458. doi: 10.3389/fonc.2019.01458
3. Skapek SX, Ferrari A, Gupta AA, Lupo PJ, Butler E, Shipley J, et al. Rhabdomyosarcoma. *Nat Rev Dis Primers*. (2019) 5:1. doi: 10.1038/s41572-018-0051-2
4. Malempati S, Weigel BJ, Chi YY, Tian J, Anderson JR, Parham DM, et al. The addition of cixutumumab or temozolomide to intensive multiagent chemotherapy is feasible but does not improve outcome for patients with metastatic rhabdomyosarcoma: a report from the children's oncology group. *Cancer*. (2019) 125:290–7. doi: 10.1002/cncr.31770
5. Oberlin O, Rey A, Lyden E, Bisogno G, Stevens MC, Meyer WH, et al. Prognostic factors in metastatic rhabdomyosarcomas: results of a pooled analysis from United States and European cooperative groups. *J Clin Oncol*. (2008) 26:2384–9. doi: 10.1200/JCO.2007.14.7207
6. Lautz TB, Chi YY, Tian J, Gupta AA, Wolden SL, Routh JC, et al. Relationship between tumor response at therapy completion and prognosis in patients with group III rhabdomyosarcoma (RMS): a report from the children's oncology group (COG). *Int J Cancer*. (2020). doi: 10.1002/ijc.32896. [Epub ahead of print].
7. Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European organization for research and treatment of cancer (EORTC) PET study group. *Eur J Cancer*. (1999) 35:1773–82. doi: 10.1016/S0959-8049(99)00229-4
8. Norman G, Fayter D, Lewis-Light K, Chisholm J, McHugh K, Levine D, et al. An emerging evidence base for PET-CT in the management of childhood rhabdomyosarcoma: systematic review. *BMJ Open*. (2015) 5:e006030. doi: 10.1136/bmjopen-2014-006030
9. Federico SM, Spunt SL, Krasin MJ, Billup CA, Wu J, Shulkin B, et al. Comparison of PET-CT and conventional imaging in staging pediatric rhabdomyosarcoma. *Pediatric Blood Cancer*. (2013) 60:1128–34. doi: 10.1002/pbc.24430
10. Ben Arush MW, Bar Shalom R, Postovsky S, Militianu D, Haimi M, Zaidman I, et al. Assessing the use of FDG-PET in the detection of regional and metastatic nodes in alveolar rhabdomyosarcoma of extremities. *J Pediatr Hematol/Oncol*. (2006) 28:440–45. doi: 10.1097/01.mph.0000212949.12856.02
11. Ricard F, Cimarelli S, Deshayes E, Mognetti T, Thiesse P, Giammarile F. Additional benefit of F-18 FDG PET/CT in the staging and follow-up of pediatric rhabdomyosarcoma. *Clin Nuclear Med*. (2011) 36:672–7. doi: 10.1097/RLU.0b013e318217ae2e
12. Tateishi U, Hosono A, Makimoto A, Nakamoto Y, Kaneta T, Fukuda H, et al. Comparative study of FDG PET/CT and conventional imaging in the staging of rhabdomyosarcoma. *Ann Nuclear Med*. (2009) 23:155–61. doi: 10.1007/s12149-008-0219-z
13. Hawkins DS, Spunt SL, Skapek SX. Children's oncology group's 2013 blueprint for research: Soft tissue sarcomas. *Pediatr Blood Cancer*. (2013) 60:1001–8. doi: 10.1002/pbc.24435
14. Bai X, Wang X. Solitary breast metastasis from rhabdomyosarcoma shown on FDG PET/CT. *Clin Nuclear Med*. (2018) 43:556–8. doi: 10.1097/RLU.0000000000002125
15. Mehta K, Sakib SMN, Phatak T, Guerrieri C, Partyka L. Rhabdomyosarcoma of the hand metastatic to the breast in an adolescent girl. *Breast J*. (2019) 25:737–8. doi: 10.1111/tbj.13311
16. Sun KK, Shen XJ. Small bowel metastasis from pulmonary rhabdomyosarcoma causing intussusception: a case report. *BMC Gastroenterol*. (2019) 19:71. doi: 10.1186/s12876-019-0990-4
17. Eugene T, Corradini N, Carlier T, Dupas B, Leux C, Bodet-Milin C. ^{18}F-FDG-PET/CT in initial staging and assessment of early response to chemotherapy of pediatric rhabdomyosarcomas. *Nuclear Med Commun*. (2012) 33:1089–95. doi: 10.1097/MNM.0b013e328356741f
18. Dharmarajan KV, Wexler LH, Gavane S, Fox JJ, Schoder H, Tom AK, et al. Positron emission tomography (PET) evaluation after initial chemotherapy and radiation therapy predicts local control in rhabdomyosarcoma. *Int J Radiat Oncol Biol Phys*. (2012) 84:996–1002. doi: 10.1016/j.ijrobp.2012.01.077
19. Dong Y, Zhang X, Wang S, Chen S, Ma C. 18F-FDG PET/CT is useful in initial staging, restaging for pediatric rhabdomyosarcoma. *Q J Nuclear Med Mol Imaging*. (2017) 61:438–46. doi: 10.23736/S1824-4785.17.02792-3
20. Klem ML, Grewal RK, Wexler LH, Schoder H, Meyers PA, Wolden SL. PET for staging in rhabdomyosarcoma: an evaluation of PET as an adjunct to current staging tools. *J Pediatr Hematol/Oncol*. (2007) 29:9–14. doi: 10.1097/MPH.0b013e3180307693
21. Volker T, Denecke T, Steffen I, Misch D, Schonberger S, Plotkin M, et al. Positron emission tomography for staging of pediatric sarcoma patients: results of a prospective multicenter trial. *J Clin Oncol*. (2007) 25:5435–41. doi: 10.1200/JCO.2007.12.2473
22. Casey DL, Wexler LH, Fox JJ, Dharmarajan KV, Schoder H, Price AN, et al. Predicting outcome in patients with rhabdomyosarcoma: role of [(18)f]fluorodeoxyglucose positron emission tomography. *Int J Radiat Oncol Biol Phys*. (2014) 90:1136–42. doi: 10.1016/j.ijrobp.2014.08.005
23. El-Kholy E, El Nadi E, Hafez H, Ahmed S, Younes A, El-Kenanii N, et al. Added predictive value of 18F-FDG PET/CT for pediatric rhabdomyosarcoma. *Nuclear Med Commun*. (2019) 40:898–904. doi: 10.1097/MNM.0000000000001040
24. Baum SH, Fruhwald M, Rahbar K, Wessling J, Schober O, Weckesser M. Contribution of PET/CT to prediction of outcome in children and young adults with rhabdomyosarcoma. *J Nuclear Med*. (2011) 52:1535–40. doi: 10.2967/jnumed.110.082511
25. Polverari G, Ceci F, Bertaglia V, Reale ML, Rampado O, Gallio E, et al. ^{18}F-FDG pet parameters and radiomics features analysis in advanced nsclc treated with immunotherapy as predictors of therapy response and survival. *Cancers*. (2020) 12:1163. doi: 10.3390/cancers12051163
26. Lima GM, Matti A, Vara G, Dondi G, Naselli N, De Crescenzo EM, et al. Prognostic value of posttreatment (18)F-FDG PET/CT and predictors of metabolic response to therapy in patients with locally advanced cervical cancer treated with concomitant chemoradiation therapy: an analysis of intensity- and volume-based PET parameters. *Eur J Nuclear Med Mol Imaging*. (2018) 45:2139–46. doi: 10.1007/s00259-018-4077-1
27. Albano D, Bosio G, Bianchetti N, Pagani C, Re A, Tucci A, et al. Prognostic role of baseline 18F-FDG PET/CT metabolic parameters in mantle cell lymphoma. *Ann Nuclear Med*. (2019) 33:449–58. doi: 10.1007/s12149-019-01354-9
28. Voglimacci M, Gabiache E, Lusque A, Ferron G, Ducassou A, Querleu D, et al. Chemoradiotherapy for locally advanced cervix cancer without aortic lymph node involvement: can we consider metabolic parameters of pretherapeutic FDG-PET/CT for treatment tailoring? *Eur J Nuclear Med Mol Imaging*. (2019) 46:1551–9. doi: 10.1007/s00259-018-4219-5
29. Nakajo M, Jinguji M, Nakabeppu Y, Nakajo M, Higashi R, Fukukura Y, et al. Texture analysis of (18)F-FDG PET/CT to predict tumour response and prognosis of patients with esophageal cancer treated by chemoradiotherapy. *Eur J Nuclear Med Mol Imaging*. (2017) 44:206–14. doi: 10.1007/s00259-016-3506-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 08 September 2020
doi: 10.3389/fbioe.2020.01006

Distinguishing Epileptiform Discharges From Normal Electroencephalograms Using Scale-Dependent Lyapunov Exponent

Qiong Li[1], *Jianbo Gao*[2,3,4]*, *Qi Huang*[5], *Yuan Wu*[5]* *and Bo Xu*[3]

[1] School of Computer, Electronics and Information, Guangxi University, Nanning, China, [2] Center for Geodata and Analysis, Faculty of Geographical Science, Beijing Normal University, Beijing, China, [3] Institute of Automation, Chinese Academy of Sciences, Beijing, China, [4] International College, Guangxi University, Nanning, China, [5] The First Affiliated Hospital of Guangxi Medical University, Nanning, China

OPEN ACCESS

Edited by:
Francesco Rundo,
STMicroelectronics (Italy), Italy

Reviewed by:
Nenad Filipovic,
University of Kragujevac, Serbia
Michael Ming-Yuan Wei,
Texas Commission on Environmental Quality, United States

***Correspondence:**
Jianbo Gao
jbgao.pmb@gmail.com
Yuan Wu
nwuyuan@stu.gxmu.edu.cn

Specialty section:
This article was submitted to Biomaterials, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: 16 April 2020
Accepted: 31 July 2020
Published: 08 September 2020

Citation:
Li Q, Gao J, Huang Q, Wu Y and Xu B (2020) Distinguishing Epileptiform Discharges From Normal Electroencephalograms Using Scale-Dependent Lyapunov Exponent. Front. Bioeng. Biotechnol. 8:1006. doi: 10.3389/fbioe.2020.01006

Epileptiform discharges are of fundamental importance in understanding the physiology of epilepsy. To aid in the clinical diagnosis, classification, prognosis, and treatment of epilepsy, it is important to develop automated computer programs to distinguish epileptiform discharges from normal electroencephalogram (EEG). This is a challenging task as clinically used scalp EEG often contains a lot of noise and motion artifacts. The challenge is even greater if one wishes to develop explainable rather than black-box based approaches. To take on this challenge, we propose to use a multiscale complexity measure, the scale-dependent Lyapunov exponent (SDLE). We analyzed 640 multi-channel EEG segments, each 4 *s* long. Among these segments, 540 are short epileptiform discharges, and 100 are from healthy controls. We found that features from SDLE were very effective in distinguishing epileptiform discharges from normal EEG. Using Random Forest Classifier (RF) and Support Vector Machines (SVM), the proposed approach with different features from SDLE robustly achieves an accuracy exceeding 99% in distinguishing epileptiform discharges from normal control ones. A single parameter, which is the ratio of the spectral energy of EEG signals and the SDLE and quantifies the regularity or predictability of the EEG signals, is introduced to better understand the high accuracy in the classification. It is found that this regularity is considerably greater for epileptiform discharges than for normal controls. Robustly having high accuracy in distinguishing epileptiform discharges from normal controls irrespective of which classification scheme being used, the proposed approach has the potential to be used widely in a clinical setting.

Keywords: EEG, epileptiform discharges, power spectral density (PSD), scale-dependent Lyapunov exponent (SDLE), random forest classifier, support vector machine (SVM)

1. INTRODUCTION

Epilepsy is a common disorder of the brain (Li et al., 2019). Approximately 8–10% of people would experience an epileptic seizure during their lifetime (Gavvala and Schuele, 2016). In adults, the risk of the recurrence of seizure within the 5 years following a new-onset or a second seizure is 35 and 75%, respectively (Gavvala and Schuele, 2016). These percentages are even higher in children, with 50% of the recurrence within the 5 years following a single unprovoked seizure, and 80% after two unprovoked seizures (Camfield and Camfield, 2015). In the United States in 2011, about 1.6 million seizure patients made emergency department visits; approximately 25% of these visits were for new-onset seizures (Gavvala and Schuele, 2016). The exact incidence of epileptic seizures in low-income and middle-income countries is unknown, however it is speculated to exceed that in high-income countries (Ba-Diop et al., 2014).

Electroencephalography (EEG) provides a continuous measure of cortical function with excellent time resolution, and thus remains the primary diagnostic test of brain function, especially in those with epileptic seizures, even though new functional imaging procedures such as functional MRI (fMRI), single-photon emission computed tomography (SPECT), and positron emission tomography (PET) have been increasingly used for assessing anatomical changes in the brain. EEG is especially valuable in investigating patients with known or suspected seizures or encephalopathy. Seizures are however infrequent events in the majority of patients in an outpatient setting, making recording of ictal EEG time-consuming and labor intensive. So far, the mainstay of diagnosis remains to detect interictal (i.e., between seizures) epileptiform discharges. Therefore, epileptiform discharges are of fundamental importance in understanding the physiology of epilepsy. To aid in the clinical diagnosis, classification, prognosis, and treatment of epilepsy, it is critical to develop automated computer programs to distinguish epileptiform discharges from normal EEG.

Many methods have been developed to study EEG. Simple but important features of EEG include the amplitude values (Toet et al., 2005) and the Power Spectral Density (PSD) (Gao et al., 2007). Using wavelet transform is also a popular approach (Adeli et al., 2003; Subasi, 2007; Faust et al., 2015; Chen et al., 2017). Clinically, however, neurologists still rely heavily to visually examine the long continuous EEG signals. Unfortunately, this approach is time-consuming and prone to error due to human fatigue. This issue has motivated much effort to develop automated algorithms to detect epileptiform discharges or other features from EEG (Sharmila and Geethanjali, 2019). Among the notable works along this line are to use entropy (Nicolaou and Georgiou, 2012; Arunkumar et al., 2016, 2017) and complexity measures (Gao et al., 2011, 2012b; Martis et al., 2015; Medvedeva et al., 2016; Pratiher et al., 2016; Sikdar et al., 2018). The majority of the works published are however based on electrocorticogram (ECoG), which is invasively obtained by directly attaching electrodes to the cerebral cortex (Wang et al., 2019). Clinically, the more widely available form of EEG is the non-invasive scalp EEG. Compared with ECoG, scalp EEG signals are much poorer in terms of signal-to-noise ratios (Haufe et al., 2018). Scalp EEG recordings also contain various kinds of artifacts (Islam et al., 2016; Brienza et al., 2019), including eye movements (e.g., blinks), muscle activities (e.g., swallowing, head movements), and the heartbeat (Kappel et al., 2017). These noise and artifacts exacerbates greatly the difficulty in automatically detecting epileptiform discharges from normal controls. Although machine learning based approaches (Mirowski et al., 2008; Shen et al., 2009; Antoniades et al., 2016; Kuswanto et al., 2017; Ullah et al., 2018; van Putten et al., 2018; Subasi et al., 2019) can partly solve some of these problems, overall, the problem remains largely open, especially with regard to the development of explainable non-black-box based approaches.

In this paper, we propose to use scale-dependent Lyapunov exponent (SDLE) to develop a readily explainable approach to automatically detect epileptiform discharges from normal controls. SDLE is a multiscale complexity measure developed to unambiguously distinguish chaos from noise, and more fundamentally to automatically characterize the defining parameters/properties of complex data (Gao et al., 2006, 2007). SDLE stems from two important concepts, the time-dependent exponent curves (Gao and Zheng, 1993, 1994a,b; Gao, 1997) and the finite size Lyapunov exponent (Torcini et al., 1995; Aurell et al., 1996, 1997). SDLE was first introduced in Gao et al. (2006, 2007), and has been further developed in Gao et al. (2009, 2012a) and applied to characterize ECoG (Gao et al., 2011), HRV (Hu et al., 2009, 2010), financial time series (Gao et al., 2013), Earth's geodynamo (Ryan and Sarson, 2008), precipitation dynamics (Fan et al., 2013), sea clutter (Hu and Gao, 2013), THz imagery (Blasch et al., 2012), and randomness (Li et al., 2016). We will show that the proposed approach is very accurate in distinguishing epileptiform discharges from normal controls.

The remainder of the paper is organized as follows. In section 2, we briefly describe the EEG data and analysis methods. In section 3, we present analysis results. In section 4, we summarize our findings.

2. MATERIALS AND METHODS

2.1. Data

The scalp EEG data analyzed here were clinically obtained at the First Affiliated Hospital to Guangxi Medical University. The studies involving human participants were reviewed and approved by the ethics committee of the First Affiliated Hospital to Guangxi Medical University. The participants provided their written informed consent to participate in this study. Fifty-nine epilepsy patients underwent a 3-h video-EEG monitoring with 19-channel EEG recording with electrodes placed on the scalp under the international 10–20 system at 256 Hz sampling rate. The electrode impedances were kept below $10K\Omega$. The 19 scalp electroencephalographic electrodes were arranged according to the names $Fp1$, $Fp2$, $F7$, $F3$, Fz, $F4$, $F8$, $T3$, $C3$, Cz, $C4$, $T4$, $T5$, $P3$, Pz, $P4$, $T6$, $O1$, and $O2$. Since the information yielded by an EEG

channel is essentially the difference of electrical activity between two electrodes in the time-domain (Pardey et al., 1996; Lopez et al., 2016), the amplitude, frequency, and synchronization of the brain waves and background will change (Seeck et al., 2017; Vanherpe and Schrooten, 2017), depending on which montage is chosen (e.g., earlobe reference, averaged reference, or bipolar; Christodoulakis et al., 2013; Geier and Lehnertz, 2017; Rana et al., 2017; Acharya and Acharya, 2019; Rios et al., 2019). In this work, we choose the widely used earlobe reference.

All epileptiform discharges were annotated by an experienced clinical neurophysiologist based on the average montage with an analog bandwidth of 0.1∼70 *Hz* and a notch filter of 50 *Hz*. EEG signals were segmented into 4 *s* epochs, with each epoch assigned a random number. The collected epochs were transformed into European Data Format (EDF) for further analysis. In total, there were 532 EEG recordings of epileptiform discharges and 100 healthy controls, each 4 *s* long, from all the participants. Among the 532 short epileptic discharges, there were 69 spike waves, 82 sharp waves, 174 spike and slow wave complexes, 72 sharp and slow wave complexes, 64 polyspike complexes, 77 polyspike, and slow wave complexes and 2 spike rhythmic discharges. Note the numbers for these seven epileptiform discharges sum up to 540, which is slightly larger than 532. The reason is a few discharges were considered to simultaneously belong to more than 1 of the 7 different epileptiform discharges. For convenience of referencing, the definitions for these 7 epileptiform discharges are given below, together with the number of cases analyzed for each type indicated in the parentheses immediately following each terminology. Examples of their waveforms are shown in **Figure 1**.

- Spike wave (69): the most basic paroxysmal EEG activity, with a duration of 20∼70 *ms*; amplitude varies but typically >50 uV (Kane et al., 2017).
- Sharp wave (82): a transient wave similar to the spike and clearly distinguishable from background activity; its time limit is 70∼200 *ms* (5∼14 *Hz*), amplitude is between 100 and 200 uV, and the phase is usually negative.
- Spike and slow wave complex (174): pattern consisting of a spike followed by a slow wave (classically the slow wave being of higher amplitude than the spike); may be single or multiple (Kane et al., 2017).
- Sharp and slow wave complex (72): pattern consisting of a sharp followed by a slow wave (classically the slow wave being of higher amplitude than the sharp); may be single or multiple (Kane et al., 2017).
- Polyspike complex (64): a sequence of two or more spikes.
- Polyspike and slow wave complex (77): pattern with two or more spikes associated with one or more slow waves.
- Spike rhythm (2): a rare pattern of widespread 10∼25 *Hz* spike rhythm outbreak, with an amplitude of 100∼200 uV and the highest voltage in the frontal area, lasting more than 1 *s*.

Recall that a few epileptiform discharge waveforms were considered to simultaneously belong to more than 1 of these 7 different epileptiform discharges. Because of this, we will not pursue the issue of further characterizing the differences among the 7 epileptiform discharges here.

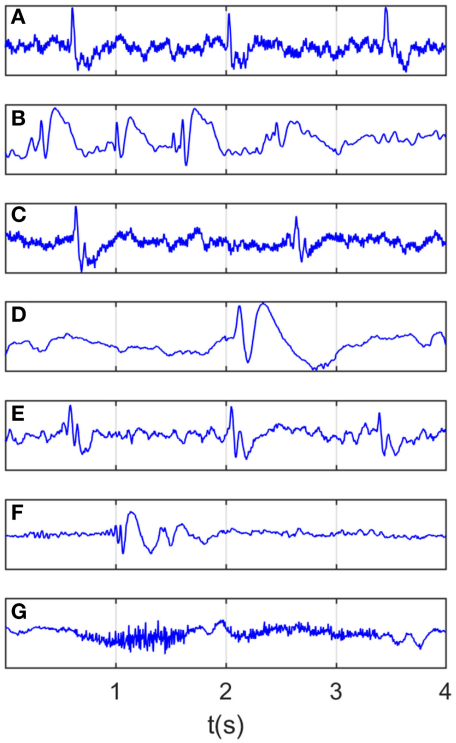


FIGURE 1 | Typical waveforms of the 7 major epileptiform EEG, where **(A-G)**, denotes spike wave, spike and slow wave complex, sharp wave, sharp and slow wave complex, polyspike complex, polyspike and slow wave complex, spike rhythm discharges, respectively.

2.2. Computation of Power Spectral Density (PSD)

PSD of EEG can be readily obtained by taking Fourier transform of the EEG signal, computing the square of the amplitude of the transform to obtain the power, and finally plotting the power against the frequency. In clinical applications, brain waves are often categorized into five bands: delta (0.5∼ 3*Hz*), theta (4∼7 *Hz*), alpha (8∼13 *Hz*), beta (14∼30 *Hz*), and gamma (>30 *Hz*), respectively. To obtain the energy of these waves, one only needs to integrate the PSD curve over the respective wave band. In this work, we integrate the PSD curve for frequencies between 0.5 and 25 Hz for the 10 electrodes with the strongest signals, and then take the average.

2.3. Computation of the SDLE

As with the estimation of PSD, for each subject, we picked up 10 strongest EEG signals from 19 electrodes, computed SDLE from each one of the 10 EEG signals, and took the average.

To compute SDLE, we first need to reconstruct a phase space from the EEG signals. Denote the signal as $x(i), i = 1, \cdots, n$, we construct vectors

$$V_i = [x(i), x(i+L), ..., x(i+(m-1)L)], \tag{1}$$

where m is called the embedding dimension and L the delay time. In practice, m and L have to be chosen properly. This is the issue of optimal embedding. For example, to reconstruct the phase plane of a harmonic oscillator from a sinusoidal signal, the optimal delay time is 1/4 of the period (Gao et al., 2007). Extensive works have been done to optimally determine m and L. Two of the most systematic and most extensively tested approaches are a statistical method called the false nearest neighbor method (Kennel et al., 1992) and a dynamical method based on time-dependent exponent curves developed by Gao and Zheng (1993, 1994a,b). The basic idea of the latter is to choose L in such a way that the motion in the reconstructed phase space is as uniform as possible (in the case of a harmonic oscillator, the reconstructed phase plane is an ellipse, which becomes a circle when L is 1/4 of the period; motion on the circle is the most uniform when compared with motions on ellipses). This is achieved by requiring divergence characterized by time-dependent exponent curves be a minimum when L is varied, and the divergence does not become much larger when m is further increased. This is the method that is employed here. For the EEG signals analyzed in this work, which was sampled with a sampling frequency of 256 Hz, we found $L = 1$ is optimal. With larger sampling frequency, L also has to be larger. For example, when the sampling frequency is 1,024 Hz, L then needs to be 4. As our EEG signal is not that long (4 *s*, or 1,024 points), we also found that $m = 2$ worked very well. After the phase space is reconstructed, we consider an ensemble of trajectories. We denote the initial separation between two nearby trajectories by ϵ_0, and their *average separation* at time t and $t + \Delta t$ by ϵ_t and $\epsilon_{t+\Delta t}$, respectively. The trajectory separation is schematically shown in **Figure 2**. Note $\epsilon_{t+\Delta t}$ is not necessarily larger than ϵ_t. We then examine the relation between ϵ_t and $\epsilon_{t+\Delta t}$, where Δt is small. When $\Delta t \to 0$, we have,

$$\epsilon_{t+\Delta t} = \epsilon_t e^{\lambda(\epsilon_t)\Delta t}, \tag{2}$$

where $\lambda(\epsilon_t)$ is the SDLE given by

$$\lambda(\epsilon_t) = \frac{\ln \epsilon_{t+\Delta t} - \ln \epsilon_t}{\Delta t}. \tag{3}$$

With the above definition, we can readily compute SDLE using the vectors defined by Equation (1). Specifically, we check whether pairs of vectors (V_i, V_j) satisfy the following Inequality:

$$\epsilon_i \le \|V_i - V_j\| \le \epsilon_i + \Delta\epsilon_i, \quad i = 1, 2, 3, \cdots, \tag{4}$$

where ϵ_i and $\Delta\epsilon_i$ are prescribed small distances. Geometrically, a pair of ϵ_i and $\Delta\epsilon_i$ defines a shell, with the former being the diameter of the shell and the latter the thickness of the shell (which reduces to a ball with radius $\Delta\epsilon_k$ when $\epsilon_k = 0$; in a 2-D plane employed here, a ball is a circle described by $(x_i - a)^2 + (x_{i+1} - b)^2 = r^2$, where (a, b) is the center of the circle, and r is the radius). We then monitor the evolution of all such vector pairs (V_i, V_j) within a shell and take the ensemble average over the indices i, j. Since we are most interested in exponential or power-law functions, we assume that taking logarithm and averaging can be exchanged, then Equation (3) can be written as

$$\begin{aligned}\lambda(\epsilon_t) &= \frac{\ln\left\langle \|V_{i+t+\Delta t} - V_{j+t+\Delta t}\| \right\rangle - \ln\left\langle \|V_{i+t} - V_{j+t}\| \right\rangle}{\Delta t} \\ &\approx \frac{\left\langle \ln \|V_{i+t+\Delta t} - V_{j+t+\Delta t}\| - \ln \|V_{i+t} - V_{j+t}\| \right\rangle}{\Delta t}\end{aligned} \tag{5}$$

where t and Δt are integers in units of the sampling time, the angle brackets denote the average over indices i, j within a shell. Note $\left\langle \|V_{i+t+\Delta t} - V_{j+t+\Delta t}\| \right\rangle$ and $\left\langle \|V_{i+t} - V_{j+t}\| \right\rangle$ amount to $\epsilon_{t+\Delta t}$ and ϵ_t, respectively. For EEG signals, the most relevant scaling law for SDLE is

$$\lambda(\epsilon) \sim -\gamma \ln \epsilon, \tag{6}$$

where γ determines the speed of loss of information.

To make the computation of SDLE readily repeated by other researchers, and more importantly, to enable different researchers to readily compare their results, we recommend to use the size of the first shell by $1/\sqrt{10}$ of the standard deviation of the EEG signal, and successive shells shrink by a factor of $1/\sqrt{2}$. Altogether, we used four shells, and then took the average of the four SDLE curves.

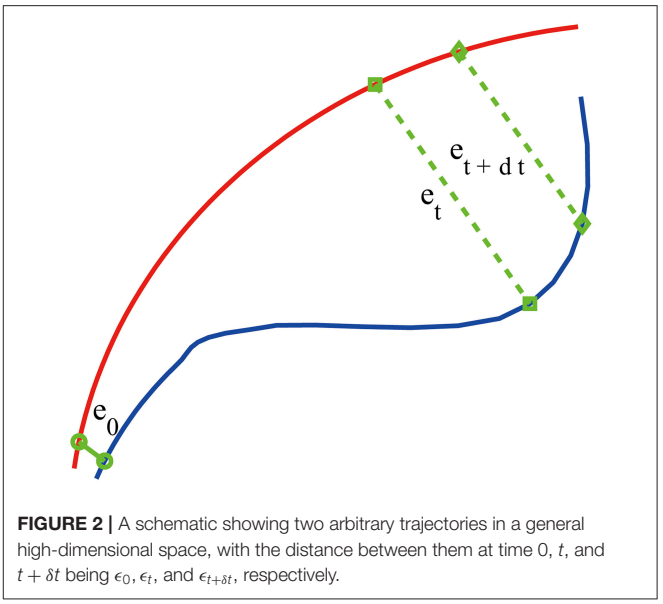


FIGURE 2 | A schematic showing two arbitrary trajectories in a general high-dimensional space, with the distance between them at time 0, t, and $t + \delta t$ being ϵ_0, ϵ_t, and $\epsilon_{t+\delta t}$, respectively.

2.4. Random Forest Classifier (RF)

Random forest (RF) is a learning technique for classification based on ensembles (Cutler et al., 2012). It is not affected by overtraining, does not require normalization of the input data, and has high accuracy. It uses many separate classification trees. Each tree is obtained through a separate bootstrap sample from the data set and classifies the data. A majority vote among the trees provides the final result.

The objective of the RF classifier used here is to classify which of the two classes an EEG signal belongs to: normal or epileptiform discharges. The inputs to the RF classifier are the PSD and a feature extracted from the SDLE curve. Following usual practice, we have randomly taken one-third of the total data as testing data and two-thirds of the data for training the model in this paper.

2.5. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a popular machine learning method for pattern classification (Cristianini and Shawe-Taylor, 2000). It has been widely used in biomedical applications. It aims to find a hyperplane in an N-dimensional space (N, the number of features) that maximizes the distance between two classes of points. Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to the two different classes. The dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane is a two-dimensional plane. When the number of features exceeds 3, it becomes difficult to imagine the shape of the hyperplane, nevertheless, it can be readily computed.

2.6. Evaluation of Performance

The consistency between the diagnosis by the neurologists and machine classification needs to be quantified. This can be accomplished by computing the receiver operating characteristic (ROC) curve and many statistics derived from the ROC curve. A good understanding of these metrics can be based on the confusion matrix, which is a table with two rows and two columns that reports the number of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN). From them we can define three major metrics:

$$sensitivity = \frac{TP}{TP + FN} \tag{7}$$

$$specificity = \frac{TN}{TN + FP} \tag{8}$$

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

Note that the sensitivity is also called true positive rate (TPR) and $1 - specificity$ is also called false positive rate (FPR).

The ROC is a plot of TPR vs. FPR using different threshold values as a sweeping variable. The ROC is a good way to characterize imbalanced data sets, as it does not suffer from class imbalance. The area below the ROC is called area under curve (AUC). Its value takes from 0 to 1. A value of AUC being 0.5 means the classification model has no predictive ability at all. On the other hand, when the value of AUC reaches 1, the prediction ability is 100%. This is equivalent to the ROC being a unit step function.

3. RESULT

We mentioned that for each subject, to compute the SDLE curves, we chose from the 19 electrodes 10 strongest EEG signals, computed the SDLE curves from each EEG signal, then took the average. For each EEG signal, we reconstructed a phase space with $m = 2, L = 1$, then computed 4 $\ln \epsilon_t$ vs. t curves corresponding to 4 shells, with the diameter of the largest shell being $1/\sqrt{10}$ of the standard deviation of the EEG signal, and successive shells shrinking by a factor of $1/\sqrt{2}$. Eight typical $\ln \varepsilon_t$ vs. t curves for epileptiform discharges and normal EEG corresponding to these four shells were shown in **Figure 3**. For simplicity, we call these error growth curves. Note the classic algorithm of computing the Lyapunov exponent amounts to assuming $\epsilon_t \sim \epsilon_0 e^{\lambda_1 t}$, where λ_1 is the largest positive Lyapunov exponent, and estimating λ_1 by $(\ln \epsilon_t - \ln \epsilon_0)/t$ (Wolf et al., 1985). This clearly is inappropriate here since $\ln \epsilon_t$ does not increase with t linearly. In other words, small variations in EEG signals did not really grow exponentially. This difficulty is readily overcome with SDLE, since the latter is the local slopes of such error growth curves, which are always well-defined. The SDLE curves corresponding to the error growth curves of **Figure 3** were shown in **Figure 4**. There are 4 SDLE curves here, corresponding to 4 shells chosen. The left-most curve corresponds to the smallest shell, while the right-most curve corresponds to the largest shell (they often are indistinguishable on larger scales). The most salient feature of these SDLE curves is the scaling behavior described by Equation (6).

It would be desirable to combine the 4 SDLE curves into a single curve. The most rigorous way to estimate the SDLE at a specific scale $\epsilon*$ is to first interpolate each SDLE curve to that scale so that it has a value there, then average the 4 SDLE curves at $\epsilon*$ using the number of pairs of vectors in each shell as the weights. For simplicity, one could also first align the 4 SDLE curves with the left-most curve, and then simply take the arithmetic average (in cases where the 4 curves are indistinguishable, then this alignment operation is unnecessary). To make the proposed method easier to reproduce, we adopted this simplified approach here. For the purpose of distinguishing epileptiform discharges from normal controls, we focused on three SDLEs $\lambda(\epsilon_1)$, $\lambda(\epsilon_2)$, and $\lambda(\epsilon_3)$ at three specific scales ϵ_1, ϵ_2, and ϵ_3, and their average, which was denoted as $\overline{\lambda}(\epsilon)$. The three scales ϵ_1, ϵ_2, and ϵ_3 were specifically indicated in **Figures 3A**, **4A**. These scales correspond to the smallest, intermediate, and boundary scales where the scaling law of Equation (6) holds.

To appreciate how well SDLEs can be used to distinguish epileptiform discharges from normal controls, we formed scatter plots with PSD and SDLEs, where PSD was obtained using Fourier transform, as we explained earlier. The scatter plots with PSD and $\lambda(\epsilon_1)$, PSD and $\lambda(\epsilon_2)$, and PSD and $\overline{\lambda}(\epsilon)$ were shown in **Figures 5–7**, respectively. We observe that in all these three cases, the separation between all seven types of epileptiform discharges and the normal control was excellent. Therefore, we can expect that the classification accuracy will be very high. Below, we specifically evaluate the performance of these three algorithms, which use PSD and $\lambda(\epsilon_1)$, PSD and $\lambda(\epsilon_2)$, and PSD and $\overline{\lambda}(\epsilon)$, respectively.

To compute the classification accuracy, we employed RF and SVM. We randomly took two-thirds of the data as the training data and the remaining one-third of the total data as the testing data. The class distribution of the samples in the training and

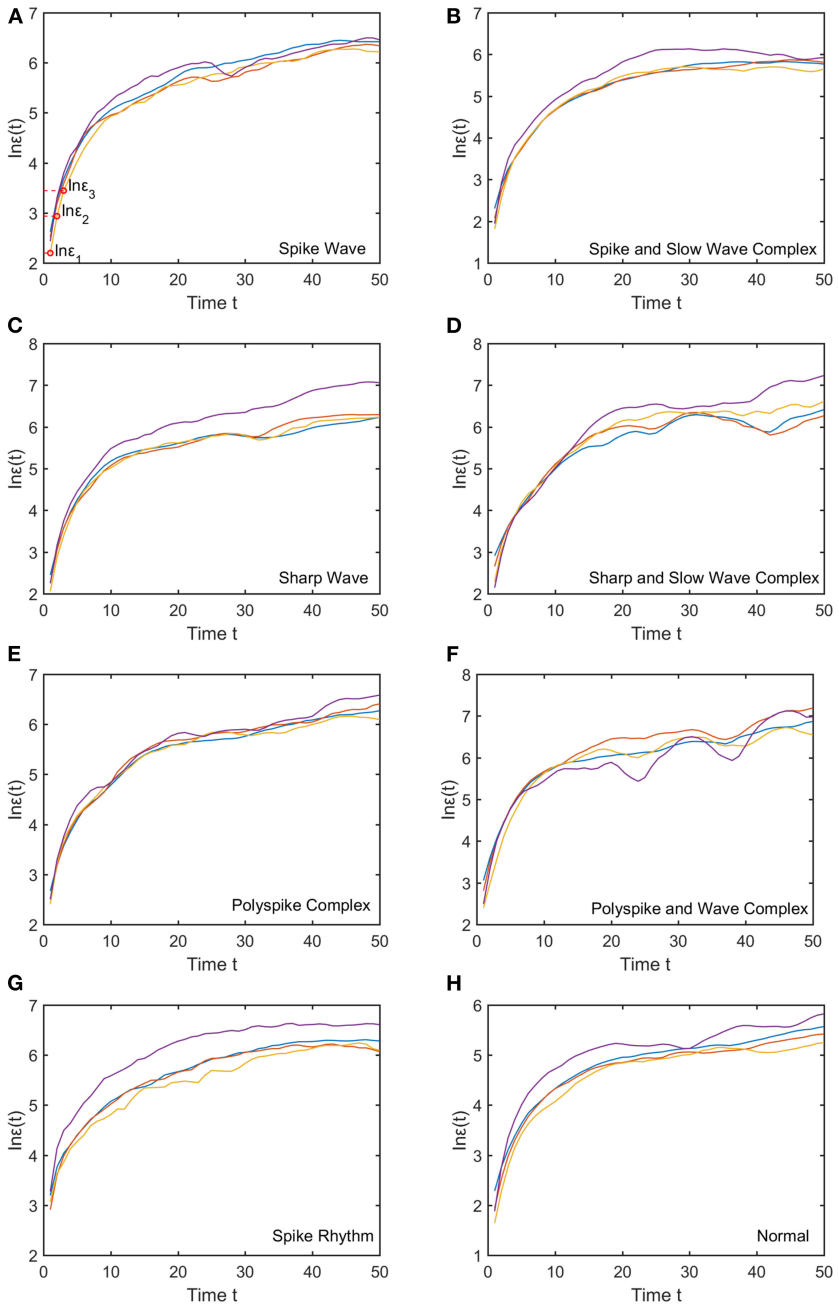


FIGURE 3 | Typical $\ln \varepsilon_t$ vs. t curves for epileptiform discharges and normal EEG, where the four curves correspond to four different shells, with the diameter of the largest shell being $1/\sqrt{10}$ of the standard deviation of the EEG signal, and successive shells shrinking by a factor of $1/\sqrt{2}$. **(A–H)** illustrates the different between the seven types of epileptiform discharges (spike wave, spike and slow wave complex, sharp wave, sharp and slow wave complex, polyspike complex, polyspike and slow wave complex, spike rhythm discharges) and normal EEG.

testing data set is summarized in **Table 1**. The test performance of the classifier can be determined by computing the metrics defined in section 2.6. The confusion matrix in **Table 2** for Algorithm 1, which used PSD and $\lambda(\epsilon_1)$, showed that 1 out of 34 normal subjects was classified incorrectly by the two classification algorithms RF and SVM as the epileptiform discharge, and 1 out of 180 epileptiform discharges was classified incorrectly as the normal subject by RF and SVM. Algorithm 2, which used PSD and $\lambda(\epsilon_2)$, was even better, which only misclassified 1 out of 180 epileptiform discharges as a normal subject by the RF, but without any other errors (the classification accuracy remained the same as that for Algorithm 1 when SVM is used). Algorithm 3, which used PSD and $\bar{\lambda}(\epsilon)$, was also excellent, which only misclassified 1 out of 34 normal subjects as an epileptiform discharge, but without any other errors for both RF and SVM. These were also summarized in **Table 2**. With these confusion matrices, we

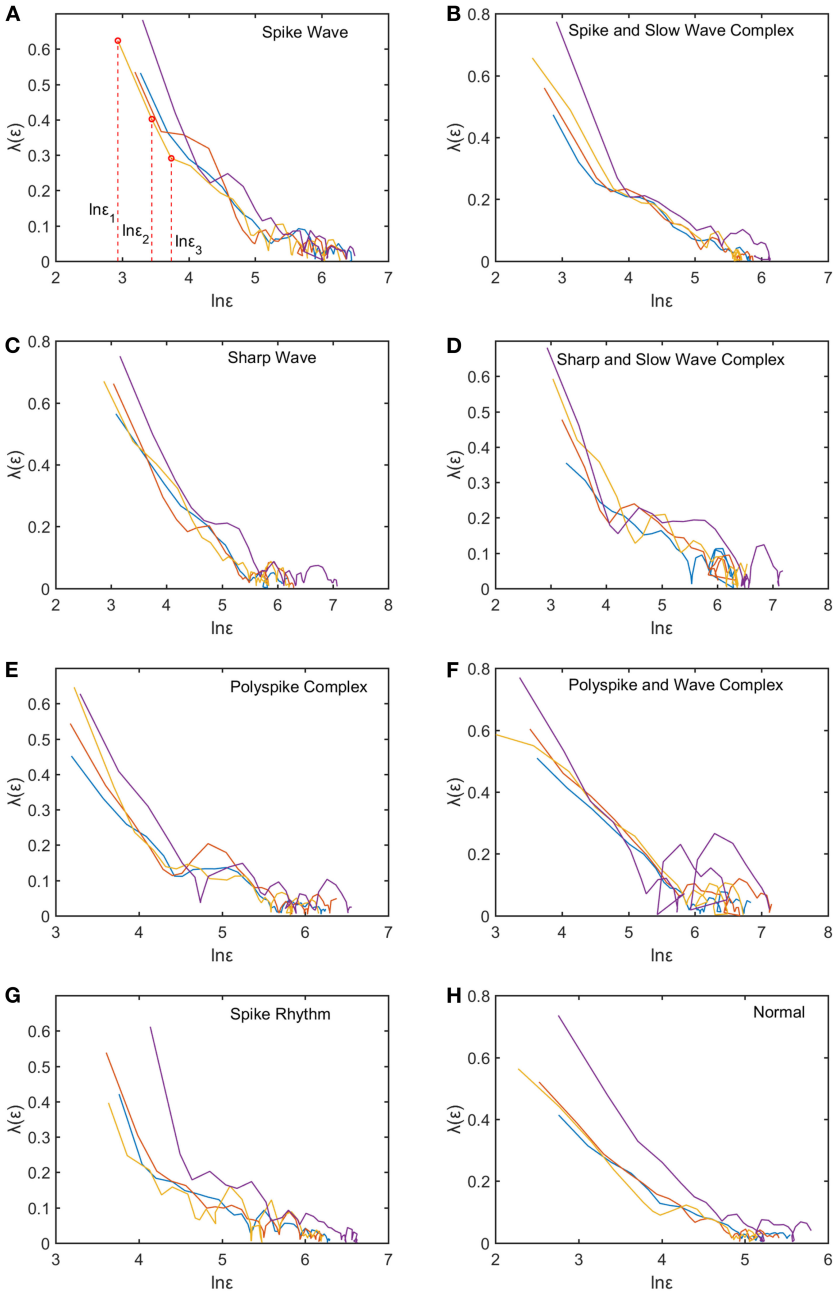


FIGURE 4 | Typical $\lambda(\epsilon)$ vs. $\ln\epsilon$ curves for epileptiform discharges and normal EEG. The four curves represented in four different colors correspond to the error growth curves shown in **Figure 3**. **(A–H)** illustrates the different between the seven types of epileptiform discharges (spike wave, spike and slow wave complex, sharp wave, sharp and slow wave complex, polyspike complex, polyspike and slow wave complex, spike rhythm discharges) and normal EEG.

computed Sensitivity, Specificity, and Accuracy of these three algorithms. They were listed in **Table 3**. We find that all the three algorithms are excellent, with their accuracy all exceeding 99%, for both classification schemes RF and SVM.

The amazing performance of these methods can be further corroborated by the unit step function like ROC curves shown in **Figure 8**. To facilitate comparison of our algorithms with that of Anh-Dao et al. (2018), which achieved a high AUC of 0.945, we also listed the AUC for the three algorithms proposed here in **Table 3**. The AUC of the three algorithms proposed here ranges from 0.9727 to 0.9980, and therefore, are all considerably better than that of Anh-Dao et al. (2018).

4. CONCLUSION AND DISCUSSION

In this paper, we have proposed to employ SDLE for distinguishing epileptiform discharges from normal EEGs, with the aim of being able to use them conveniently in a clinical setting. We found that SDLE computed from scalp EEG signals was mainly characterized by a scaling law described by

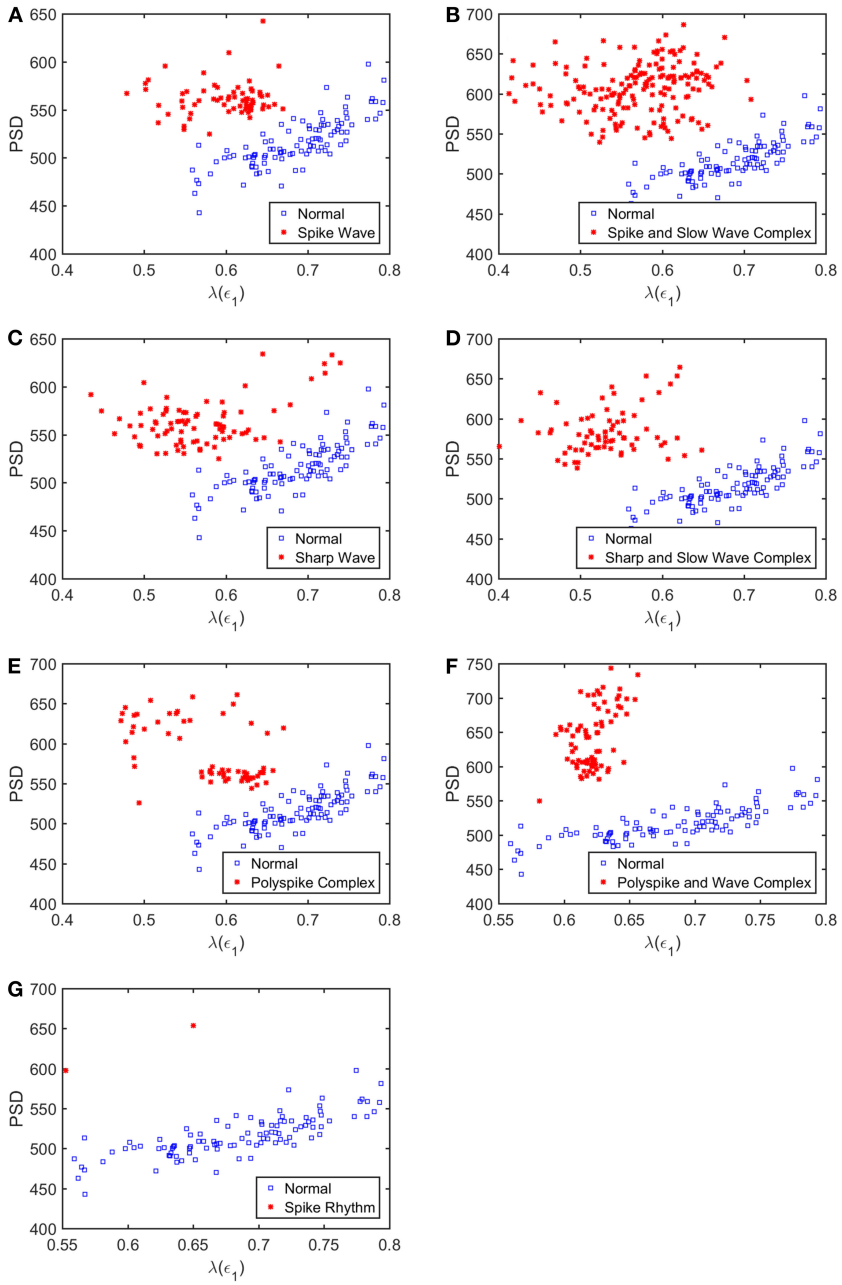


FIGURE 5 | Scatter plots with PSD and λ(e1), where **(A–G)**, illustrates the different between the seven types of epileptiform discharges (spike wave, spike and slow wave complex, sharp wave, sharp and slow wave complex, polyspike complex, polyspike and slow wave complex, spike rhythm discharges) and normal EEG. These plots highly suggest the classification accuracy will be very high.

Equation (6). When the scale parameters were confined to where this scaling law held, SDLE was very effective in distinguishing epileptiform discharges from normal EEG. Using RF and SVM, the proposed approach with different features from SDLE was found to robustly achieve an accuracy exceeding 99% in distinguishing epileptiform discharges from normal control ones.

What is the reason that the choice of concrete classification schemes such as RF or SVM is not critical for the proposed approach to have high accuracy in distinguishing epileptiform discharges from normal control ones? It has to be because of the excellent separations revealed by the scatter plots shown in **Figures 5**–**7**. To better understand the explainability of the proposed approach, we need to understand better the meaning of the SDLE. The definition of SDLE is equivalent to

$$\ln \epsilon_t = \ln \epsilon_0 + \int_0^t \lambda(\epsilon_t) dt. \tag{10}$$

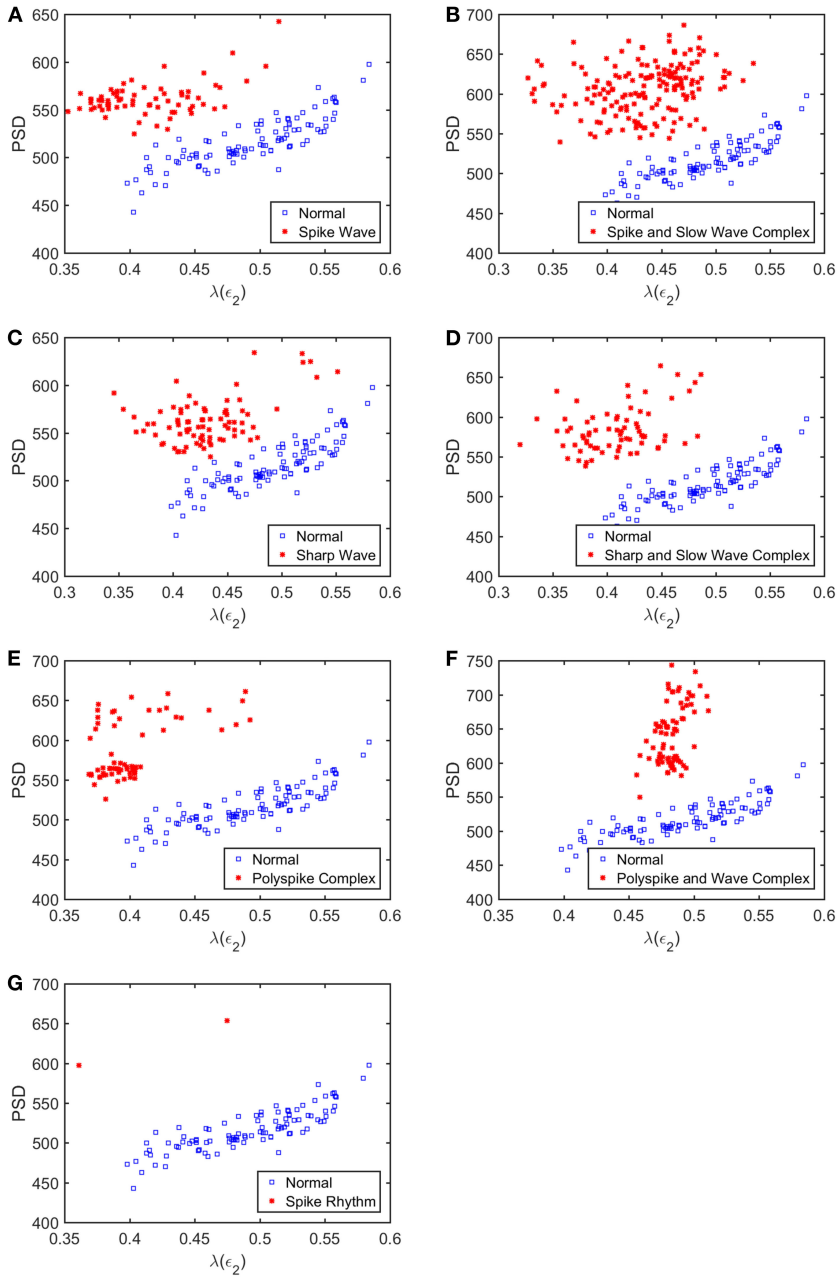


FIGURE 6 | Scatter plots with PSD and λ(e2), where **(A–G)**, illustrates the different between the seven types of epileptiform discharges (spike wave, spike and slow wave complex, sharp wave, sharp and slow wave complex, polyspike complex, polyspike and slow wave complex, spike rhythm discharges) and normal EEG. These plots highly suggest the classification accuracy will be very high.

Letting $\epsilon_{T_{db}} = 2\epsilon_0$, we find the error doubling time T_{db} given by

$$\ln 2 = \int_0^{T_{db}} \lambda(\epsilon_t) dt. \tag{11}$$

As the first approximation, we may consider $1/\lambda(\epsilon)$ to be proportional to the error doubling time (Gao et al., 2009). This understanding motivates us to combine the two parameters PSD and SDLE into a single parameter such as PSD/$\lambda(\epsilon_1)$. Since on average PSD is larger but $\lambda(\epsilon_1)$ (as well as $\lambda(\epsilon_2)$ and $\overline{\lambda}(\epsilon)$, as shown in **Figures 5–7** is smaller for epileptiform discharges than for normal control ones, we can expect that this ratio will be on average larger for epileptiform discharges. In fact, this ratio can be regarded as a measure of the regularity or predictability of EEG signals, since large PSD stems from synchronized firing of neurons, while small SDLE highlights slow divergence and thus considerable regularity and predictability.

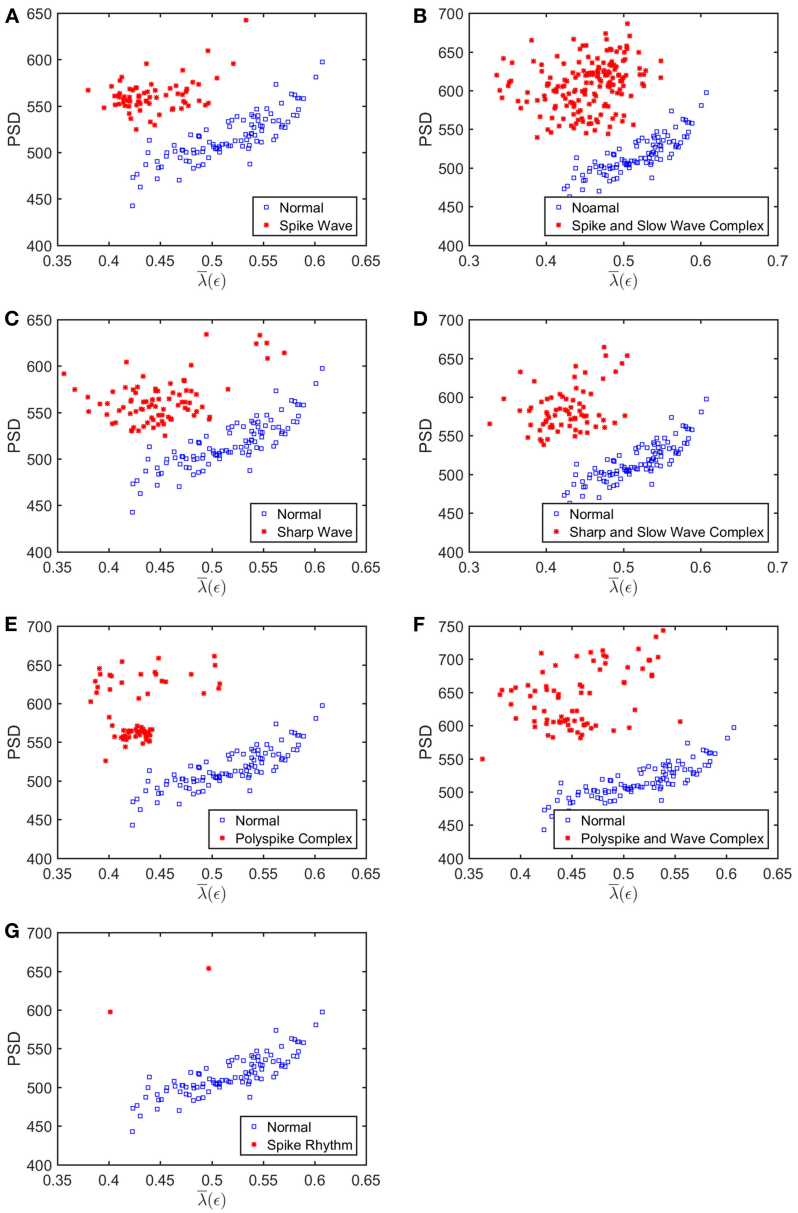


FIGURE 7 | Scatter plots with PSD and $\overline{\lambda}(\epsilon)$, where **(A–G)**, illustrates the different between the seven types of epileptiform discharges (spike wave, spike and slow wave complex, sharp wave, sharp and slow wave complex, polyspike complex, polyspike and slow wave complex, spike rhythm discharges) and normal EEG. These plots highly suggest the classification accuracy will be very high.

TABLE 1 | Class distribution of the samples in the training and testing data sets.

Classes	Training set	Testing set	Total
Normal controls	66	34	100
Epileptiform discharges	360	180	540
Total	426	214	640

Now the question is whether such a single parameter can effectively distinguish normal control ones from epileptiform discharges. For this purpose, we have computed the probability density distribution (PDF) for PSD/$\lambda(\epsilon_1)$ of the epileptiform discharges and the normal control ones. The results are shown in **Figure 9** as the blue and the red curves, respectively. The overlapping of the blue and the red curves defines a right and a left tail for the blue and the red curves; the corresponding probabilities for them are 1.39 and 4.19%, as indicated in the plot. They correspond to the probability that a normal control one may be misclassified as an epileptiform discharge and vice versa. As the classification accuracy with the scheme based on a single parameter will not be higher than that based on two parameters, we can readily understand that the probabilities of

1.39 and 4.19% are the lower bounds that a normal control may be misclassified as epileptiform discharges, and vice verse. This is surely consistent with the probabilities shown in **Table 3** (the case for Algorithm 1). As these misclassification probabilities are very low, we thus can be confident that the proposed approach will always be very accurate in distinguishing epileptiform discharges from normal control ones, no matter what classification schemes are used for classification.

It is interesting to note that if we choose SDLE corresponding to larger scales, such as ϵ_3 indicated in **Figures 3A**, **4A**, an algorithm based on PSD and $\lambda(\epsilon_3)$ would be slightly worse than the three algorithms discussed here, but still slightly better than that of Anh-Dao et al. (2018). This suggests the importance of properly selecting the scale for analysis. On the other end, if we use a three parameter method, for example, using PSD, $\overline{\lambda}(\epsilon)$, and ϵ_∞ (which characterizes the size of an attractor and amounts to the largest scale in **Figure 3**), then the accuracy in distinguishing epileptiform discharges from normal controls can be further improved to 100%. The reason is that ϵ_∞ contains information independent of PSD and SDLE. However, we had not further pursed the issue of improving the accuracy here, since the high accuracy achieved by the easily explainable algorithms presented is already more than satisfying. Overall, our analysis highly

TABLE 2 | Confusion Matrix for the testing data of 180 epileptiform discharges and 34 normal controls: Algorithms 1, 2, 3 use PSD and $\lambda(\epsilon_1)$, PSD and $\lambda(\epsilon_2)$, PSD and $\overline{\lambda}(\epsilon)$, respectively.

Classifier	Algorithms	Result	Epileptiform discharges	Healthy controls
RF	Algorithm 1	Epileptiform discharges	179	1
		Healthy controls	1	33
	Algorithm 2	Epileptiform discharges	179	1
		Healthy controls	0	34
	Algorithm 3	Epileptiform discharges	180	0
		Healthy controls	1	33
SVM	Algorithm 1	Epileptiform discharges	179	1
		Healthy controls	1	33
	Algorithm 2	Epileptiform discharges	179	1
		Healthy controls	1	33
	Algorithm 3	Epileptiform discharges	180	0
		Healthy controls	1	33

TABLE 3 | Classification performance measures.

Classifier	Algorithms	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
RF	Algorithm 1	99.44	97.06	99.07	0.9784
	Algorithm 2	99.44	100.00	99.53	0.9980
	Algorithm 3	100.00	97.06	99.53	0.9953
SVM	Algorithm 1	99.44	97.06	99.07	0.9766
	Algorithm 2	99.44	97.06	99.07	0.9727
	Algorithm 3	100.00	97.06	99.53	0.9953

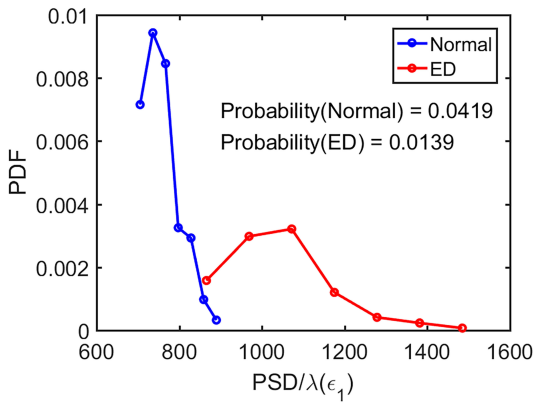


FIGURE 9 | The probability density distribution (PDF) for the ratio PSD/$\lambda(\epsilon_1)$ of the epileptiform discharges (red curve) and normal control ones (blue curve). The overlapping of the blue and the red curves defines a right and left tail for the blue and red curves, respectively; the corresponding probabilities for them are 1.39 and 4.19%, as indicated in the plot.

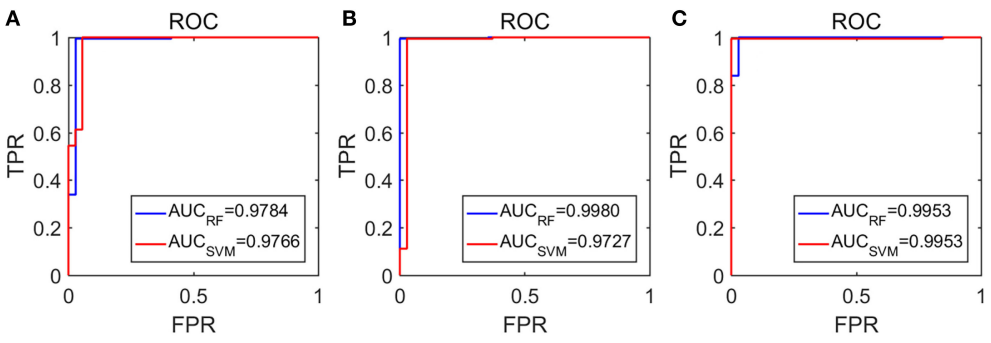


FIGURE 8 | The ROC curves for the testing data: **(A–C)** are for algorithms using PSD and $\lambda(\epsilon_1)$, PSD and $\lambda(\epsilon_2)$, and PSD and $\overline{\lambda}(\epsilon)$, respectively.

suggests that the proposed approach is very promising to be used clinically.

It is worth noting that the epileptiform discharges analyzed here were provided by our collaborators at Guangxi Medical University in two batches: in the first batch, which was about 2/3 of the data analyzed here, the accuracy was similar to that reported here. Then another 1/3 of the data were given to us to further examine whether the accuracy remained as high. It was yes. Nevertheless, the data analyzed here were still quite limited. It would be interesting and important to further validate the proposed approaches with more data in different clinical sets.

Brain activities involve spatial-temporal coordinated dynamics of numerous neurons in different regions of the brain, i.e., involve numerous functional brain networks. To better characterize the synergistic effects among the brain networks, it is important to construct brain networks based on multi-channel EEG signals. Closely related to this network issue is to infer the localization of each type of epileptiform discharges, which is of great clinical importance. These issues have not been pursued in this work, which is obviously a serious limitation of the current study. In the near future, we will examine these issues systematically, especially from the viewpoint of synthesizing network analysis with nonlinear analysis based on complexity science.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The scalp EEG data analyzed here were clinically obtained at the First Affiliated Hospital to Guangxi Medical University. The studies involving human participants were reviewed and approved by the ethics committee of the First Affiliated Hospital to Guangxi Medical University. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

QL performed most of the experimental work. QH and YW provided the data needed for this experiment and engaged in many discussions, together with BX. JG conceived the study, provided overall supervision for the study, directed all phases of the study, and including writing of the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China under Grant Nos. 71661002 and 41671532 and by the Fundamental Research Funds for the Central Universities. It is also supported by the National Key Research and Development Program of China, grant number 2019AAA0103402. One of the authors (JG) also benefited tremendously from participating the long program on culture analytics organized by the Institute for Pure and Applied Mathematics (IPAM) at UCLA, which was supported by the National Science Foundation.

REFERENCES

Acharya, J. N., and Acharya, V. J. (2019). Overview of EEG montages and principles of localization. *J. Clin. Neurophysiol.* 36, 325–329. doi: 10.1097/WNP.0000000000000538

Adeli, H., Zhou, Z., and Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *J. Neurosci. Methods* 123, 69–87. doi: 10.1016/S0165-0270(02)00340-0

Anh-Dao, N. T., Linh-Trung, N., Van Nguyen, L., Tran-Duc, T., and Boashash, B. (2018). A multistage system for automatic detection of epileptic spikes. *Rev J. Electron. Commun*. 8, 1–12. doi: 10.21553/rev-jec.166

Antoniades, A., Spyrou, L., Took, C. C., and Sanei, S. (2016). "Deep learning for epileptic intracranial EEG data," in *Deep learning International Workshop on Machine Learning for Signal Processing (MLSP)* (Vietri sul Mare), 1–6. doi: 10.1109/MLSP.2016.7738824

Arunkumar, N., Ram Kumar, K., and Venkataraman, V. (2016). Automatic detection of epileptic seizures using permutation entropy, Tsallis entropy and Kolmogorov complexity. *J. Med. Imaging Health Inform*. 6, 526–531. doi: 10.1166/jmihi.2016.1710

Arunkumar, N., Ramkumar, K., Venkatraman, V., Abdulhay, E., Fernandes, S. L., Kadry, S., et al. (2017). Classification of focal and non-focal EEG using entropies. *Pattern Recogn. Lett*. 94, 112–117. doi: 10.1016/j.patrec.2017.05.007

Aurell, E., Boffetta, G., Crisanti, A., Paladin, G., and Vulpiani, A. (1996). Growth of noninfinitesimal perturbations in turbulence. *Phys. Rev. Lett*. 77:1262. doi: 10.1103/PhysRevLett.77.1262

Aurell, E., Boffetta, G., Crisanti, A., Paladin, G., and Vulpiani, A. (1997). Predictability in the large: an extension of the concept of Lyapunov exponent. *J. Phys*. A 30:1. doi: 10.1088/0305-4470/30/1/003

Ba-Diop, A., Marin, B., Druet-Cabanac, M., Ngoungou, E., Newton, C., and Preux, P. (2014). "epidemiology, causes, and treatment of epilepsy in Sub-Saharan Africa. *Lancet Neurol*. 13, 1029–1044. doi: 10.1016/S1474-4422(14)70114-0

Blasch, E. P., Gao, J., and Tung, W.-W. (2012). "Chaos-based image assessment for THZ imagery," in *2012 11th International Conference on Information Science, Signal Processing and Their Applications (ISSPA)* (Montreal, QC), 360–365. doi: 10.1109/ISSPA.2012.6310576

Brienza, M., Davassi, C., and Mecarelli, O. (2019). "Artifacts," in *Clinical Electroencephalography*, ed O. Mecarelli (Cham: Springer), 109–130. doi: 10.1007/978-3-030-04573-9_8

Camfield, P., and Camfield, C. (2015). Incidence, prevalence and aetiology of seizures and epilepsy in children. *Epilept. Disord*. 17, 117–123. doi: 10.1684/epd.2015.0736

Chen, D., Wan, S., Xiang, J., and Bao, F. S. (2017). A high-performance seizure detection algorithm based on discrete wavelet transform (dwt) and EEG. *PLoS ONE* 12:e173138. doi: 10.1371/journal.pone.0173138

Christodoulakis, M., Hadjipapas, A., Papathanasiou, E. S., Anastasiadou, M., Papacostas, S. S., and Mitsis, G. D. (2013). "Graph-theoretic analysis of scalp EEG brain networks in epilepsy-the influence of montage and volume conduction,' in *13th IEEE International Conference on Bioinformatics and Bioengineering* (Chania), 1–4. doi: 10.1109/BIBE.2013.6701572

Cristianini, N., and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511801389

Cutler, A., Cutler, D. R., and Stevens, J. R. (2012). "Random forests," in *Ensemble Machine Learning*, eds C. Zhang and Y. Q. Ma (Boston, MA: Springer), 157–175. doi: 10.1007/978-1-4419-9326-7_5

Fan, Q., Wang, Y., and Zhu, L. (2013). Complexity analysis of spatial-temporal precipitation system by PCA and SDLE. *Appl. Math. Model.* 37, 4059–4066. doi: 10.1016/j.apm.2012.09.009

Faust, O., Acharya, U. R., Adeli, H., and Adeli, A. (2015). Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. *Seizure* 26, 56–64. doi: 10.1016/j.seizure.2015.01.012

Gao, J. (1997). Recognizing randomness in a time series. *Phys. D* 106, 49–56. doi: 10.1016/S0167-2789(97)00024-9

Gao, J., Cao, Y., Tung, W.-w., and Hu, J. (2007). *Multiscale Analysis of Complex Time Series: Integration of Chaos and Random Fractal Theory, and Beyond*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9780470191651

Gao, J., Hu, J., Mao, X., and Tung, W.-W. (2012a). Detecting low-dimensional chaos by the "noise titration" technique: possible problems and remedies. *Chaos Solitons Fractals* 45, 213–223. doi: 10.1016/j.chaos.2011.12.004

Gao, J., Hu, J., Tung, W., and Cao, Y. (2006). Distinguishing chaos from noise by scale-dependent Lyapunov exponent. *Phys. Rev. E* 74:066204. doi: 10.1103/PhysRevE.74.066204

Gao, J., Hu, J., and Tung, W.-W. (2011). Complexity measures of brain wave dynamics. *Cogn. Neurodyn.* 5, 171–182. doi: 10.1007/s11571-011-9151-3

Gao, J., Hu, J., and Tung, W.-w. (2012b). Entropy measures for biological signal analyses. *Nonlin. Dyn.* 68, 431–444. doi: 10.1007/s11071-011-0281-2

Gao, J., Hu, J., Tung, W.-W., and Zheng, Y. (2013). Multiscale analysis of economic time series by scale-dependent Lyapunov exponent. *Quant. Fin.* 13, 265–274. doi: 10.1080/14697688.2011.580774

Gao, J., Tung, W., and Hu, J. (2009). Quantifying dynamical predictability: the pseudo-ensemble approach. *Chin. Ann. Math. Ser. B* 30, 569–588. doi: 10.1007/s11401-009-0108-3

Gao, J., and Zheng, Z. (1993). Local exponential divergence plot and optimal embedding of a chaotic time series. *Phys. Lett. A* 181, 153–158. doi: 10.1016/0375-9601(93)90913-K

Gao, J., and Zheng, Z. (1994a). Direct dynamical test for deterministic chaos. *Europhys. Lett.* 25:485. doi: 10.1209/0295-5075/25/7/002

Gao, J., and Zheng, Z. (1994b). Direct dynamical test for deterministic chaos and optimal embedding of a chaotic time series. *Phys. Rev. E* 49:3807. doi: 10.1103/PhysRevE.49.3807

Gavvala, J., and Schuele, S. (2016). New-onset seizure in adults and adolescents: a review. *JAMA* 316, 2657–2668. doi: 10.1001/jama.2016.18625

Geier, C., and Lehnertz, K. (2017). Which brain regions are important for seizure dynamics in epileptic networks? Influence of link identification and EEG recording montage on node centralities. *Int. J. Neural Syst.* 27:1650033. doi: 10.1142/S0129065716500337

Haufe, S., DeGuzman, P., Henin, S., Arcaro, M., Honey, C. J., Hasson, U., et al. (2018). Reliability and correlation of fMRI, ECOG and EEG during natural stimulus processing. *BioRxiv* 2018, 207456. doi: 10.1101/207456

Hu, J., and Gao, J. (2013). Multiscale characterization of sea clutter by scale-dependent Lyapunov exponent. *Math. Probl. Eng.* 2013:584252. doi: 10.1155/2013/584252

Hu, J., Gao, J., and Tung, W.-W. (2009). Characterizing heart rate variability by scale-dependent Lyapunov exponent. *Chaos* 19:028506. doi: 10.1063/1.3152007

Hu, J., Gao, J., Tung, W.-W., and Cao, Y. (2010). Multiscale analysis of heart rate variability: a comparison of different complexity measures. *Ann. Biomed. Eng.* 38, 854–864. doi: 10.1007/s10439-009-9863-2

Islam, M. K., Rastegarnia, A., and Yang, Z. (2016). Methods for artifact detection and removal from scalp EEG: a review. *Neurophysiol. Clin.* 46, 287–305. doi: 10.1016/j.neucli.2016.07.002

Kane, N., Acharya, J., Benickzy, S., Caboclo, L., Finnigan, S., Kaplan, P. W., et al. (2017). A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. revision 2017. *Clin. Neurophysiol. Pract.* 2:170. doi: 10.1016/j.cnp.2017.07.002

Kappel, S. L., Looney, D., Mandic, D. P., and Kidmose, P. (2017). Physiological artifacts in scalp EEG and ear-EEG. *Biomed. Eng. Online* 16:103. doi: 10.1186/s12938-017-0391-2

Kennel, M. B., Brown, R., and Abarbanel, H. D. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* 45:3403. doi: 10.1103/PhysRevA.45.3403

Kuswanto, H., Salamah, M., and Fachruddin, M. I. (2017). Random forest classification and support vector machine for detecting epilepsy using electroencephalograph records. *Am. J. Appl. Sci.* 14, 533–539. doi: 10.3844/ajassp.2017.533.539

Li, F., Liang, Y., Zhang, L., Yi, C., Liao, Y., Jiang, Y., et al. (2019). Transition of brain networks from an interictal to a preictal state preceding a seizure revealed by scalp EEG network analysis. *Cogn. Neurodyn.* 13, 175–181. doi: 10.1007/s11571-018-09517-6

Li, X.-Z., Zhuang, J.-P., Li, S.-S., Gao, J.-B., and Chan, S.-C. (2016). Randomness evaluation for an optically injected chaotic semiconductor laser by attractor reconstruction. *Phys. Rev. E* 94:042214. doi: 10.1103/PhysRevE.94.042214

Lopez, S., Gross, A., Yang, S., Golmohammadi, M., Obeid, I., and Picone, J. (2016). "An analysis of two common reference points for EEGs," in *2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)* (Philadelphia, PA), 1–5. doi: 10.1109/SPMB.2016.7846854

Martis, R. J., Tan, J. H., Chua, C. K., Loon, T. C., Yeo, S. W. J., and Tong, L. (2015). Epileptic EEG classification using nonlinear parameters on different frequency bands. *J. Mech. Med. Biol.* 15:1550040. doi: 10.1142/S0219519415500402

Medvedeva, T. M., Lüttjohann, A., van Luijtelaar, G., and Sysoev, I. V. (2016). "Evaluation of nonlinear properties of epileptic activity using largest Lyapunov exponent," in *Saratov Fall Meeting 2015: Third International Symposium on Optics and Biophotonics and Seventh Finnish-Russian Photonics and Laser Symposium (PALS), Vol. 9917* (Saratov: International Society for Optics and Photonics), 991724.

Mirowski, P. W., LeCun, Y., Madhavan, D., and Kuzniecky, R. (2008). "Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG," in *2008 IEEE Workshop on Machine Learning for Signal Processing* (Cancún), 244-249. doi: 10.1109/MLSP.2008.4685487

Nicolaou, N., and Georgiou, J. (2012). Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Syst. Appl.* 39, 202–209. doi: 10.1016/j.eswa.2011.07.008

Pardey, J., Roberts, S., and Tarassenko, L. (1996). A review of parametric modelling techniques for EEG analysis. *Med. Eng. Phys.* 18, 2–11. doi: 10.1016/1350-4533(95)00024-0

Pratiher, S., Patra, S., and Bhattacharya, P. (2016). "On the marriage of Kolmogorov complexity and multi-fractal parameters for epileptic seizure classification," in *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)* (Noida), 831–836. doi: 10.1109/IC3I.2016.7918797

Rana, A. Q., Ghouse, A. T., and Govindarajan, R. (2017). "Basics of electroencephalography (EEG)," in *Neurophysiology in Clinical Practice*, ed J. Renwick (Cham: Springer), 3–9. doi: 10.1007/978-3-319-39342-1_1

Rios, W. A., Olguín, P. V., Mena, D. A., Cabrera, M. C., Escalona, J., Garcia, A. M., et al. (2019). The influence of EEG references on the analysis of spatio-temporal interrelation patterns. *Front. Neurosci.* 13:941. doi: 10.3389/fnins.2019.00941

Ryan, D., and Sarson, G. (2008). The geodynamo as a low-dimensional deterministic system at the edge of chaos. *Europhys. Lett.* 83:49001. doi: 10.1209/0295-5075/83/49001

Seeck, M., Koessler, L., Bast, T., Leijten, F., Michel, C., Baumgartner, C., et al. (2017). The standardized EEG electrode array of the IFCN. *Clin. Neurophysiol.* 128, 2070–2077. doi: 10.1016/j.clinph.2017.06.254

Sharmila, A., and Geethanjali, P. (2019). A review on the pattern detection methods for epilepsy seizure detection from EEG signals. *Biomed. Eng.* 64, 507–517. doi: 10.1515/bmt-2017-0233

Shen, T.-W., Kuo, X., and Hsin, Y.-L. (2009). "Ant k-means clustering method on epileptic spike detection," in *2009 Fifth International Conference on Natural Computation, Vol. 6* (Tianjin), 334–338. doi: 10.1109/ICNC.2009.639

Sikdar, D., Roy, R., and Mahadevappa, M. (2018). Epilepsy and seizure characterisation by multifractal analysis of EEG subbands. *Biomed. Signal Process. Control* 41, 264–270. doi: 10.1016/j.bspc.2017.12.006

Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Syst. Appl.* 32, 1084–1093. doi: 10.1016/j.eswa.2006.02.005

Subasi, A., Kevric, J., and Canbaz, M. A. (2019). Epileptic seizure detection using hybrid machine learning methods. *Neural Comput. Appl.* 31, 317–325. doi: 10.1007/s00521-017-3003-y

Toet, M. C., Groenendaal, F., Osredkar, D., van Huffelen, A. C., and de Vries, L. S. (2005). Postneonatal epilepsy following amplitude-integrated EEG-detected neonatal seizures. *Pediatr. Neurol.* 32, 241–247. doi: 10.1016/j.pediatrneurol.2004.11.005

Torcini, A., Grassberger, P., and Politi, A. (1995). Error propagation in extended chaotic systems. *J. Phys. A* 28:4533. doi: 10.1088/0305-4470/28/16/011

Ullah, I., Hussain, M., Qazi E. H., and Aboalsamh, H. (2018). An automated system for epilepsy detection using EEG brain signals based on deep learning approach. *Expert Syst. Appl.* 107, 61–71. doi: 10.1016/j.eswa.2018.04.021

van Putten, M. J., de Carvalho, R., and Tjepkema-Cloostermans, M. C. (2018). F85. deep learning for detection of epileptiform discharges from scalp EEG recordings. *Clin. Neurophysiol.* 129, e98–e99. doi: 10.1016/j.clinph.2018.04.248

Vanherpe, P., and Schrooten, M. (2017). Minimal EEG montage with high yield for the detection of status epilepticus in the setting of postanoxic brain damage. *Acta Neurol. Belgica* 117, 145–152. doi: 10.1007/s13760-016-0663-9

Wang, Q., Valdés-Hernández, P. A., Paz-Linares, D., Bosch-Bayard, J., Oosugi, N., Komatsu, M., et al. (2019). EECOG-comp: an open source platform for concurrent EEG/ECOG comparisons-applications to connectivity studies. *Brain Topogr.* 32, 1–19. doi: 10.1007/s10548-019-00708-w

Wolf, A., Swift, J. B., Swinney, H. L., and Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Phys. D* 16, 285–317. doi: 10.1016/0167-2789(85)90011-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 11 September 2020
doi: 10.3389/fonc.2020.01676

Machine Learning-Based Radiomics Predicting Tumor Grades and Expression of Multiple Pathologic Biomarkers in Gliomas

*Min Gao[1], Siying Huang[2], Xuequn Pan[3], Xuan Liao[1], Ru Yang[1] and Jun Liu[1]**

[1] Department of Radiology, Second Xiangya Hospital, Central South University, Changsha, China, [2] School of Computer Science and Engineering, University of New South Wales, Sydney, NSW, Australia, [3] Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda, MD, United States

Background: The grading and pathologic biomarkers of glioma has important guiding significance for the individual treatment. In clinical, it is often necessary to obtain tumor samples through invasive operation for pathological diagnosis. The present study aimed to use conventional machine learning algorithms to predict the tumor grades and pathologic biomarkers on magnetic resonance imaging (MRI) data.

Methods: The present study retrospectively collected a dataset of 367 glioma patients, who had pathological reports and underwent MRI scans between October 2013 and March 2019. The radiomic features were extracted from enhanced MRI images, and three frequently-used machine-learning models of LC, Support Vector Machine (SVM), and Random Forests (RF) were built for four predictive tasks: (1) glioma grades, (2) Ki67 expression level, (3) GFAP expression level, and (4) S100 expression level in gliomas. Each sub dataset was split into training and testing sets at a ratio of 4:1. The training sets were used for training and tuning models. The testing sets were used for evaluating models. According to the area under curve (AUC) and accuracy, the best classifier was chosen for each task.

Results: The RF algorithm was found to be stable and consistently performed better than Logistic Regression and SVM for all the tasks. The RF classifier on glioma grades achieved a predictive performance (AUC: 0.79, accuracy: 0.81). The RF classifier also achieved a predictive performance on the Ki67 expression (AUC: 0.85, accuracy: 0.80). The AUC and accuracy score for the GFAP classifier were 0.72 and 0.81. The AUC and accuracy score for S100 expression levels are 0.60 and 0.91.

Conclusion: The machine-learning based radiomics approach can provide a non-invasive method for the prediction of glioma grades and expression levels of multiple pathologic biomarkers, preoperatively, with favorable predictive accuracy and stability.

Keywords: glioma, biomarkers, machine learning, radiomics, MRI

OPEN ACCESS

Edited by:
Francesco Rundo, STMicroelectronics, Italy

Reviewed by:
Francesca Trenta, University of Catania, Italy
Seyedmehdi Payabvash, Yale University, United States

****Correspondence:***
Jun Liu
junliu123@csu.edu.cn

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *24 May 2020*
Accepted: *29 July 2020*
Published: *11 September 2020*

Citation:
Gao M, Huang S, Pan X, Liao X, Yang R and Liu J (2020) Machine Learning-Based Radiomics Predicting Tumor Grades and Expression of Multiple Pathologic Biomarkers in Gliomas. Front. Oncol. 10:1676. doi: 10.3389/fonc.2020.01676

INTRODUCTION

Gliomas are the most common brain tumors and are often classified as World Health Organization (WHO) grades I-IV, depending on the different tumor cells, and the degree of abnormality (1, 2). As a tumor's grade increases, gliomas process more aggressively (3). Treatment options and responses differ from glioma grades (4). Pathological findings are the premise of rational treatment. Usually, glioma grades are confirmed by pathological examination during surgery or biopsy (5). Then, a following immunohistochemistry (IHC) test determines the molecular biomarkers of tumor tissues at the microscopic level. These pathologic biomarkers, typical proteins, are useful indicators for diagnosis, prognosis, or treatment response (6). However, obtaining such information for gliomas requires invasive approaches. The surgical decision making could be difficult and time-consuming for many patients. Those patients who are not eligible for a surgery or seek non-surgical treatment may have limited treatment options without pathological guidance. Therefore, presurgical glioma grades and the expression of biomarkers are valued and preferred with non-invasive approaches.

At present, the medical imaging can differentiate the tumor phenotype and intra-tumor heterogeneity (7). Conventional magnetic resonance imaging (MRI) is routinely used in the diagnosis and management of glioma patients. T1-weighted contrast-enhanced MRI (T1C) is the current standard for initial brain tumor imaging (8). Radiomics can generate image features with high dimensional data from the intensity histogram, geometry and texture analyses on the entire tumor volume (9). With the emergence of Artificial Intelligence (AI) technologies, advanced informatics tools have become accessible to facilitate machine learning (ML) based radiomics applications using image features as the data source (10). Radiomics is gaining ground in oncology and have the potential to accurately classify or predict tumor characteristics.

Radiomics approaches have been applied for the predictions of glioma grades or differential diagnoses (11, 12). Several studies have reached a prediction accuracy of above 80% using popular ML models. The commonly and frequently used ML algorithms in radiomics include Logistic Regression (LR), Random Forests (RF), Support Vector Machine (SVM), and etc. Each ML method has their own advantages in the classification. For example, LR fits the variables coefficients and predicts a logit transformation of the probability of being one class or the other. SVM separates the classes by finding an optimal hyperplane. RF uses bootstrap aggregating to decision trees and improves classification performance.

When compared to tumor grading, to make predictions at a molecular level is more challenging. Kickingereder et al. reported the association between established MRI features and cancer gene variations (EGFR amplification and CDKN2A loss), but failed to build a sufficient ML model to predict the molecular characteristics (13). In clinic, pathologic biomarkers are more frequently tested for than genetic testing. IDH1 is one important glioma biomarker and IDH1 mutation along with 1p/19q is a part of the molecular diagnosis in the updated 2016 WHO classification (14). Ki67, S100, and GFAP are also the common protein targets for gliomas. IDH1, Ki67, and GFAP were once considered as the golden triad of glioma IHC (15) Ki67 is highly correlated to proliferation that may indicate the tumor grades and prognosis (16–18). S100 has been implicated in the regulation of cellular activities, such as metabolism, motility, and proliferation. Under the pathological conditions of tumor and inflammation, the concentration of the S100 protein increases to the micromole level, which stimulates microglia and astrocytes, and increases the expression of pro-inflammatory cytokines (19–23). GFAP is the most widely used markers of astrocytes (24). Under the condition of injury (trauma or disease), the expression of GFAP in astrocytes rapidly increases (25). GFAP is often used to reveal the astrocytic lineage of glial cells and glial tumor cells, and plays a more significant role in tumor pathology, when compared to the differential diagnosis of astrocytoma. Ki67, S100, or GFAP may not be a reliable diagnostic biomarker for gliomas, because their roles in gliomas are still under investigations, while controversies have been observed in experiments (26). However, there is no doubt that these proteins can provide some insights into the tumor intra-microenvironment.

So far, it is not surprising to know that most radiomics studies favor the prediction of the IDH expression for molecular diagnosis (11, 27), with a few reports on Ki67 (28). In order to expand predictive effects of radiomics, the investigators aimed to assess the prediction feasibility of glioma grades and the pathologic biomarkers of Ki67, S100, and GFAP in gliomas. The investigators believed that the combination of multiple biomarkers can increase the predictive power, and the information obtained can help in understanding the underlying pathologic process in gliomas. The investigators designed the present retrospective study and extracted hundreds of radiomic features from the T1C images of 367 glioma patients. Three machine-learning-based models (LR, SVM, and RF) were built to perform the tasks: (1) classify the glioma grades, and (2) predict the expression levels of Ki67, S100, and GFAP. This study demonstrated that multiple pathologic biomarkers in gliomas can be estimated to the certainty levels of clinical using common ML models on conventional MRI data and pathological records.

MATERIALS AND METHODS

Study Cohort

The investigators retrospectively collected a data set of 420 glioma patients, who had pathological reports and MRI scans performed between October 2013 and March 2019, from the Second Xiangya Hospital of Central South University. The patients who met the following criteria were included: (i) a histopathological diagnosis of primary glioma based on the WHO classification, (ii) the availability of IHC profiles of biomarkers (S100, GFAP, and Ki67), (iii) preoperative MRI data of post-contrast axial T1-weighted (T1C), and (iv) age > 18 years old. Patients were excluded due to the following: (i) secondary gliomas or postoperative recurrence of gliomas, (ii) obvious artifacts in

MRI. Ethics approval was obtained for the present study from the Ethics Committee of the Second Xiangya Hospital, Central South University.

Pathological Evaluation

Patient demographics (age and gender), and histopathologic diagnosis and IHC results were obtained from a surgical pathology report. On these reports, the diagnosis included a specific glioma type by cells (e.g., astrocytoma and oligodendrogliomas) and a given WHO grade (I–IV). The IHC results were presented in the list of glioma biomarkers (e.g., S100, GFAP, or Ki67) and their own expression profile in tumor cells. It is noteworthy that the list was not standard and varied upon the request or availability of the biomarkers at that time. For example, few patients received an IDH1 test before 2017, but after 2016, the WHO classification standard was published, and IDH1 tests became common. So, a patient might have a different set of tested biomarkers, and the number of cases can differ for each biomarker. Their IHC results depended on the scoring system used. The expression levels were usually evaluated by the staining intensity of positive cells, and points were assigned to describe these positive cells by count (e.g., 0 points as negative (−), 1 point as positive (+), 2 points as medium positive, and 3 points as high positive), percentage (e.g., 0 points as none, 1 point less than 5%, 2 points approximately 5–25%, and 3 points above 25%), or the appearance of a clear brown color (e.g., 1 point for light yellow). In the study, the glioma grades were classified as low-grade (WHO I–II, benign) and high-grade (WHO III–IV, malignant), and expression levels of biomarkers were divided into two categories: a low expression scored less than 2 points and a high expression scored 2 points or above.

Imaging Post-processing and Radiomics Features Extraction

Magnetic resonance imaging scans were acquired from different scanners over time. The Picture Archiving and Communication System (PACS) exported the selected DICOM images to a local computer using the RadiAnt DICOM Viewer (Medixant, PL). In order to reduce the influence of different scanning parameters, post-processing and image registration were applied using the Advanced Normalization Tools (ANTS 2.1, PA). Then, the DICOM images were loaded into ITK-SNAP for segmentation and standardization (29). Two neuroradiologists (5 years of experience) drew the region of interest (ROI) around the tumor boundary on the T1C images. The neuroradiologists were blinded to the patient identification and diagnosis. After a joint effort, disagreements with the boundary were solved. The ROI segmentations were resampled to match the dimensions of the original images, and both images were saved in.narrd as the input for feature extraction.

The Pyradiomics extractor was customized to calculate and extract the features (10). All built-in filters [wavelet, Laplacian of Gaussian (LoG), square, square root, logarithm, and exponential] were enabled on five image feature classes [first order statistics, shape descriptors, and texture features on the gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), and gray-level size zone matrix (GLSZM)]. Feature definitions and calculation algorithms were available in the PyRadiomics documentation[1].

Machine Learning

The feature importance and the following predictive ML methods were implemented using Python (version 3.7.0) with machine-learning library scikit-learn (version 23.0) (30). All features were standardized through Min-Max scaling. Features with all zero scores were removed. Clinical data (age and gender) were added in constructing the final prediction models.

Feature Importance

The feature importance helped in understanding the importance of the features, since a large number radiomics features with high-dimensional data are difficult to interpret. Three technique approaches were used to identify the important features. First, chi-squared (chi^2) tests were applied in the scikit-learn SelectKBest class to obtain a list of the top 15 best features. Second, the heatmap of correlated features was plotted to identify features highly correlated to predicting targets (glioma grade and biomarker expression) using the seaborn library. Third, a RF classifier was initiated and the in-build feature importance was used to extract the top features.

Predictive Machine Learning Models

Three frequently-used machine-learning based models of LR, SVM, and RF were built for four predictive tasks: (1) glioma grades, (2) Ki67 expression level, (3) GFAP expression level, and (4) S100 expression level in gliomas. Each sub dataset was divided into training and testing sets at a ratio of 4:1 (train_size = 0.8, test_size = 0.2). Principal Component Analysis (PCA) was applied for high-dimension reduction that maps n-dimensional features to k-dimensional features ($n > k$), resulting in brand new orthogonal features. For the unbalanced data in different classes, the synthetic minority over-sampling technique (SMOTE) algorithm was used to oversample the minority class (31). On training set, the grid search with cross-validation was applied for hyper parameters tuning (RF and SVM), and k fold validation was used for LR. The accuracy score was compared with the result from their base models (default settings in scikit-learn) for model selection. The testing set was used for final model evaluation. The performance of the models was evaluated according to accuracy, the area under curve (AUC) of the receiver operating characteristic (ROC), sensitivity, specificity, the positive prediction value (PPV), and the negative predictive value (NPV). According to the AUC and accuracy, the best classifier was chosen for each task.

Statistics

One way-ANOVA or simple t-test was applied to test the differences among gender, age, glioma grade, and the expression levels of the biomarkers. Descriptive statistics was used to summarize the important features through filters and feature classes. All significant levels were tested at 0.05.

[1]https://pyradiomics.readthedocs.io/en/latest/features.html

RESULTS

Subjects and Pathologic Biomarkers

A data set of preoperative MRI and surgical pathologic reports of 420 glioma patients were collected. A total of 51 patients were excluded for not meeting the inclusion criteria. Among these patients, 40 patients were under 18 years old, seven patients had quality issues on their MRI data, and four patients did not have an assigned WHO classification level in their records. The age of the enrolled 369 patients ranged within 18–75 years old (mean age: 45.63 ± 13.22 years old), and consisted of 210 males (age: 46.99 ± 13.24 years old), and 159 females (age: 43.84 ± 13.03 years old). The clinical characteristics of patients and the distribution of the selected biomarkers across glioma grades are presented in **Table 1**.

The expression of GFAP, Ki67, and S100 was reported as follows: 367 patients had GFAP results with four negatives (0 point), 323 positives (1 point), and 35 medium (2 points), or 5 high positives (3 points); 348 patients underwent Ki67 tests, including 96 negatives or low positives (≤5% in tumor cells), and 252 strong positives (>5%); 338 patients underwent S100 tests, which included eight negatives (0 points), 315 positives (1 point), and 15 medium positives (2 points).

There was a significant age difference among male and female patients, as determined by one-way ANOVA [F (1, 367) = 5.17, $P < 0.05$]. Furthermore, there were significant differences in age, gender and tumor volume among glioma grades (WHO I–IV). Moreover, there were significant differences in glioma grade, tumor size, age and gender for the Ki67 expression. However, there were no significant differences in age, gender and glioma grade for S100 and GFAP expression. The t-test and one-way ANOVA results are shown in **Table 2**.

TABLE 1 | Distribution of clinical characteristics and expression levels of IHC biomarkers grouped by glioma WHO grades.

	WHO I	WHO II	WHO III	WHO IV
Total number	5	142	116	106
Mean age (s.d.)	35.4(7.64)	40.65(11.69)	48.29(13.82)	49.87(12.4)
Gender				
Male	2	72	68	68
Female	3	70	48	38
Tumor volume av a (cm^3)	30.8	38.34	46.47	53.81
Ki67 expression level				
0	4	73	15	4
1	0	57	97	98
GFAP expression level				
0	0	1	1	2
1	5	126	98	94
2	0	13	13	9
3	0	2	2	1
S100 expression level				
0	0	3	0	5
1	5	120	104	86
2	0	6	5	4

TABLE 2 | Clinical characteristics vs. glioma grade and expression levels of IHC biomarkers.

	Age	Gender	Tumor volume	Grade
Grade	t = 6.1602 df = 367 p = 1.91e-09	t = −2.2766 df = 367 p = 0.02339	t = 2.5027 df = 355 p = 0.01277	
Ki67	t = 5.6168 df = 346 p = 4.001e-08	F(1,346) = 0.53 p = 0.467	t = 1.5089 df = 336 p = 0.1323	F(1,346) = 124.7 $p < 0.05$
GFAP	t = −0.30242 df = 365 p = 0.7625	F(1,365) = 0.569 p = 0.451	t = −1.1268 df = 354 p = 0.2606	F(1,365) = 0.089 p = 0.77
S100	t = −0.307 df = 336 p = 0.759	F(1,336) = 0.186 p = 0.667	t = 1.639 df = 326 p = 0.1022	F(1,336) = 0.59 p = 0.44

MRI Data Processing and Feature Extraction

A total of 369 original T1C images and their paired segmentation images underwent the feature extraction process using Pyradiomics. The investigators extracted 1,421 radiomics features (14 shape features, 27 first-order intensity statistics features, 68 texture features, 96 square features, 96 square root features, 96 logarithm features, 96 exponential features, 172 LoG features, and 766 wavelet features). After data cleaning, 1,372 features reminded. The data set was normalized by the SKlearn MinMaxScaler.

Features Importance

The investigators obtained the list of the top 15 important features based on the scores obtained from the chi-squared stats between each non-negative feature and the glioma grade, and S100, GFAP, and Ki 67 expression levels. The features and their scores are shown in **Table 3**. The scores ranged within 3.67–44.04. The mean score of the top important features was 9.30, with a standard deviation of 5.83. The frequent top features within the image type were exponential (23), wavelet (22), square (6), square root (3), original (3), gradian (2), and ihp-2D (1). For the feature classes, the frequent top features were divided as follows: glszm (27), glcm (9), glrlm (8), gldm (7), first order (7), and ngtdm (2). The heatmaps of the correlated features for glioma grade and the biomarkers of Ki67, GFAP, and S100 are presented in **Figure 1**. The RF model built-in feature importance is presented in **Figure 2**.

Prediction Machine Learning Models

The performance of the 12 predictive models is presented in **Table 4**. The RF models performed slightly better, when compared to the other models. The comparisons with accuracy and the results are presented below. **Figure 3** shows the AUC_ROC for the RF classifier in sub test sets.

Glioma Grades

The sub data set was randomly split into the training set of 276 cases and the test set of 93 cases. With a PCA retention of 0.95, the PCA process reduced the dimensions

TABLE 3 | Feature importance by chi-square scores.

Prediction	Top pyradiomics imaging feature	Score	Filter	Class
GRADE	Exponential_ngtdm_Coarseness	44.04	Exponential	ngtdm
	Exponential_glszm_LowGrayLevelZoneEmphasis	19.79	Exponential	glszm
	Exponential_glszm_SizeZoneNonUniformityNormalized	16.93	Exponential	glszm
	Exponential_glszm_ZoneEntropy	13.68	Exponential	glszm
	Exponential_glcm_MCC	12.09	Exponential	glcm
	Exponential_glcm_Correlation	11.30	Exponential	glcm
	Exponential_glszm_GrayLevelNonUniformity	10.92	Exponential	glszm
	Exponential_glszm_SmallAreaEmphasis	10.60	Exponential	glszm
	Exponential_glcm_InverseVariance	10.52	Exponential	glcm
	Square_glszm_ZonePercentage	9.80	Square	glszm
	Wavelet-LHL_firstorder_TotalEnergy	9.64	Wavelet-LHL	firstorder
	Exponential_glcm_Imc2	9.59	Exponential	glcm
	Exponential_glszm_GrayLevelNonUniformityNormalized	9.46	Exponential	glszm
	Gradient_firstorder_TotalEnergy	9.03	Gradient	firstorder
	Wavelet-HHL_firstorder_TotalEnergy	8.48	Wavelet-HHL	firstorder
GFAP	Lbp-2D_firstorder_10Percentile	12.38	Lbp-2D	firstorder
	Wavelet-HLH_glrlm_LowGrayLevelRunEmphasis	12.25	Wavelet-HLH	glrlm
	Wavelet-HLH_gldm_LowGrayLevelEmphasis	12.13	Wavelet-HLH	gldm
	Wavelet-HLH_glszm_LowGrayLevelZoneEmphasis	11.79	Wavelet-HLH	glszm
	Wavelet-HHH_gldm_LowGrayLevelEmphasis	11.19	Wavelet-HHH	gldm
	Wavelet-HHH_glrlm_LowGrayLevelRunEmphasis	11.18	Wavelet-HHH	glrlm
	Wavelet-HHL_gldm_LowGrayLevelEmphasis	11.12	Wavelet-HHL	gldm
	Wavelet-HHL_glrlm_LowGrayLevelRunEmphasis	11.08	Wavelet-HHL	glrlm
	Wavelet-HLH_gldm_LargeDependenceLowGrayLevelEmphasis	10.77	Wavelet-HLH	gldm
	Wavelet-HHH_gldm_LargeDependenceLowGrayLevelEmphasis	10.64	Wavelet-HHH	gldm
	Wavelet-HHH_glrlm_LongRunLowGrayLevelEmphasis	9.99	Wavelet-HHH	glrlm
	Wavelet-HHL_gldm_LargeDependenceLowGrayLevelEmphasis	9.89	Wavelet-HHL	gldm
	Wavelet-HLH_glrlm_ShortRunLowGrayLevelEmphasis	9.30	Wavelet-HLH	glrlm
	Wavelet-HHL_glrlm_LongRunLowGrayLevelEmphasis	8.98	Wavelet-HHL	glrlm
	Wavelet-HHH_glrlm_ShortRunLowGrayLevelEmphasis	8.51	Wavelet-HHH	glrlm
S100	Wavelet-LLH_glszm_LargeAreaHighGrayLevelEmphasis	13.65	Wavelet-LLH	glszm
	Wavelet-LLL_glszm_LargeAreaHighGrayLevelEmphasis	10.53	Wavelet-LLL	glszm
	Original_glszm_LargeAreaHighGrayLevelEmphasis	10.45	Original	glszm
	Squareroot_glszm_LargeAreaHighGrayLevelEmphasis	8.44	Squareroot	glszm
	Original_glszm_ZoneVariance	8.08	Original	glszm
	Exponential_firstorder_Energy	7.89	Exponential	firstorder
	Original_glszm_LargeAreaEmphasis	7.87	Original	glszm
	Squareroot_glszm_ZoneVariance	5.93	Squareroot	glszm
	Squareroot_glszm_LargeAreaEmphasis	5.83	Squareroot	glszm
	Exponential_firstorder_TotalEnergy	5.72	Exponential	firstorder
	Wavelet-LHH_glszm_LargeAreaLowGrayLevelEmphasis	5.65	Wavelet-LHH	glszm
	Wavelet-LLH_glszm_ZoneVariance	5.49	Wavelet-LLH	glszm
	Wavelet-LLH_glszm_LargeAreaEmphasis	5.39	Wavelet-LLH	glszm
	Gradient_glszm_LargeAreaLowGrayLevelEmphasis	5.23	Gradient	glszm
	Wavelet-LHL_glszm_LargeAreaHighGrayLevelEmphasis	4.98	Wavelet-LHL	glszm
Ki67	Exponential_ngtdm_Coarseness	18.37	Exponential	ngtdm
	Exponential_glszm_LowGrayLevelZoneEmphasis	8.44	Exponential	glszm
	Exponential_glszm_SizeZoneNonUniformityNormalized	7.75	Exponential	glszm
	Exponential_glszm_ZoneEntropy	6.12	Exponential	glszm
	Exponential_glszm_GrayLevelNonUniformity	4.64	Exponential	glszm
	Exponential_glcm_MCC	4.36	Exponential	glcm
	Square_glszm_SmallAreaLowGrayLevelEmphasis	4.20	Square	glszm

(Continued)

TABLE 3 | Continued

Prediction	Top pyradiomics imaging feature	Score	Filter	Class
	Square_gldm_LowGrayLevelEmphasis	4.12	Square	gldm
	Exponential_glcm_Imc2	4.12	Exponential	glcm
	Square_glrlm_LowGrayLevelRunEmphasis	4.09	Square	glrlm
	Exponential_glcm_InverseVariance	4.02	Exponential	glcm
	Exponential_glszm_GrayLevelNonUniformityNormalized	3.95	Exponential	glszm
	Exponential_glcm_Correlation	3.89	Exponential	glcm
	Square_firstorder_Uniformity	3.69	Square	firstorder
	Square_glcm_MaximumProbability	3.67	Square	glcm

to 37 components, and these remained in the final prediction model of glioma grading. There was a 96:252 class distribution. After SMOTE oversampling, the number of train samples increased to 318. After grid search with cross validation (cv = 5) or K fold validation (n_splits = 5), the selected classifier included: (1) LR (penalty = "l2," C = 1.0), (2) SVM (C = 10, kernel = "rbf," and gamma = 0.1), and (3) RF (min_samples_leaf = 1,min_samples_split = 2, and n_estimators = 100). The RF classifier achieved a satisfying predictive performance (AUC: 0.79, accuracy: 0.81). The average accuracy, sensitivity, specificity and f1 score was 0.81, 0.63, 0.89, and 0.67, respectively.

Ki67 Expression

A total of 348 patients had Ki67 test results, which included 252 low expression levels and 96 high expression levels. There was a 96:252 class distribution. The training set and test set were split into 278 and 70 cases, respectively. After the SMOTE oversampling, the number of train samples increased to 415. With a PCA retention of 0.95, the PCA process reduced the dimensions to 37 components, and there were used for the final prediction model for the Ki_67 expression. After grid search with cross validation (cv = 5) or K fold validation (n_splits = 5), the selected classifier included: (1) LR (penalty = "l2," C = 1.0), (2) SVM (C = 10, kernel = "rbf," and gamma = 0.1), and (3) RF (max_depth = 80, max_features = 3, min_samples_leaf = 4,min_samples_split = 8, and n_estimators = 100). Among these three classifiers, the RF classifier achieved the best predictive performance on the Ki67 expression based on the AUC (0.85), accuracy (0.80), sensitivity (0.91), specificity (0.80), and f1 score (0.85) for the Ki67 high expression.

S100 Expression

A total of 338 patients had S100 test results, which included 323 low expression levels (<2 points) and 15 high expression levels (≥2 points). The class distribution was 323:15. The training set and test set were split into 270 and 68, respectively. After the SMOTE oversampling, the resampled number increased to 518. With a PCA retention of 0.95, the PCA process reduced the dimensions to 38 components, and these were used for the final prediction model for the S100 expression. After grid search with cross validation (cv = 5) or K fold validation (n_splits = 5), the selected classifier included: (1) LR (penalty = "l2," C = 1.0), (2) SVM (C = 1, kernel = "rbf," and gamma = "auto"), and (3) RF (min_samples_leaf = 1,min_samples_split = 2, and n_estimators = 100). Among these classifiers, the RF classifier achieved the best prediction performance on the S100 expression, based on the measurements (AUC: 0.60, accuracy: 0.91, average-weighted sensitivity: 0.88 specificity: 0.91, and f1 score: 0.90). It is noteworthy that the average-weight computes f1 for each class, and returns the average while considering the proportion for each class in the dataset. For S100 low expression levels: accuracy (0.95), sensitivity (0.94), specificity (0.97), and f1 (0.95). For high expression levels: none of the four high expression cases was correctly predicted.

GFAP Expression

A total of 367 patients had a GFAP test. Among these patients, there were 327 low expression levels and 40 high expression levels. The class distribution ratio was 327:40. The training set and test set were split into 293 and 74, respectively. After the SMOTE oversampling, the number of samples increased to 532. With a PCA retention of 0.95, the PCA process reduced the dimensions to 38 components, and those that remained were used for the final prediction model for the GFAP expression. After grid search with cross validation (cv = 5) or K fold validation (n_splits = 5), the selected classifier included: (1) LR (penalty = "l2," C = 1.0), (2) SVM (C = 1, kernel = "rbf," and gamma = "auto"), and (3) RF (min_samples_leaf = 1,min_samples_split = 2, and n_estimators = 100). Among these three classifiers, the RF classifier achieved the best predictive performance on the GFAP expression measured, as follows: AUC (0.72), accuracy (0.81), average-weighted sensitivity (0.74), specificity (0.81), and f1 score (0.76).

DISCUSSION

The machine-learning based radiomics approach was applied to predict glioma grades and the expression levels of pathologic biomarkers Ki67, GFAP, and S100 in low or high. The overall performance of the ML models was satisfactory. The RF algorithm was found to be stable and consistently performed better than LR and SVM. Feature importance varies on predictive tasks, glioma grade or specific protein expression. The most frequent important feature classes were textual and first order statistics.

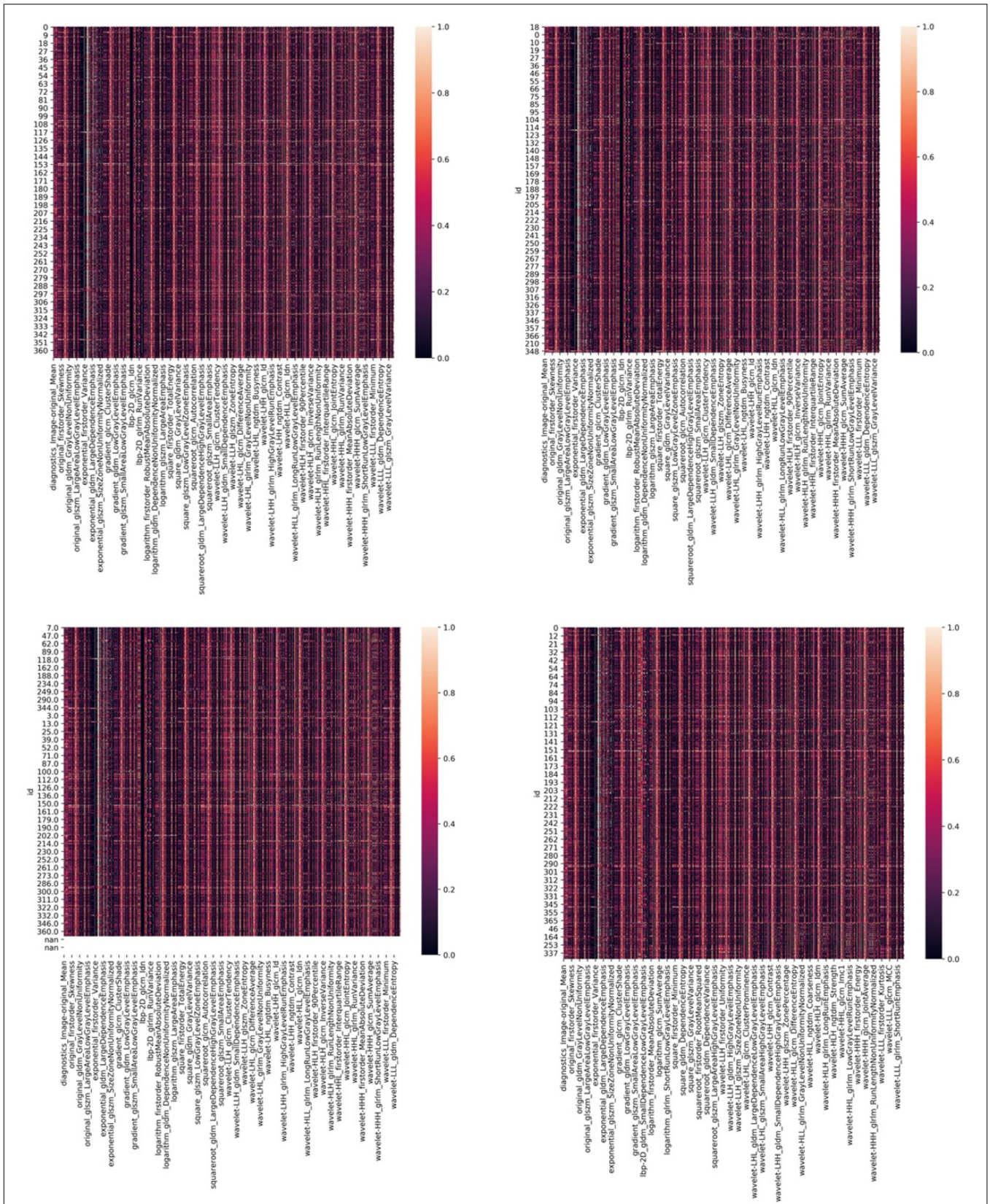


FIGURE 1 | The heatmaps of corelated features for glioma grade and biomarkers of Ki67, GFAP, and S100.

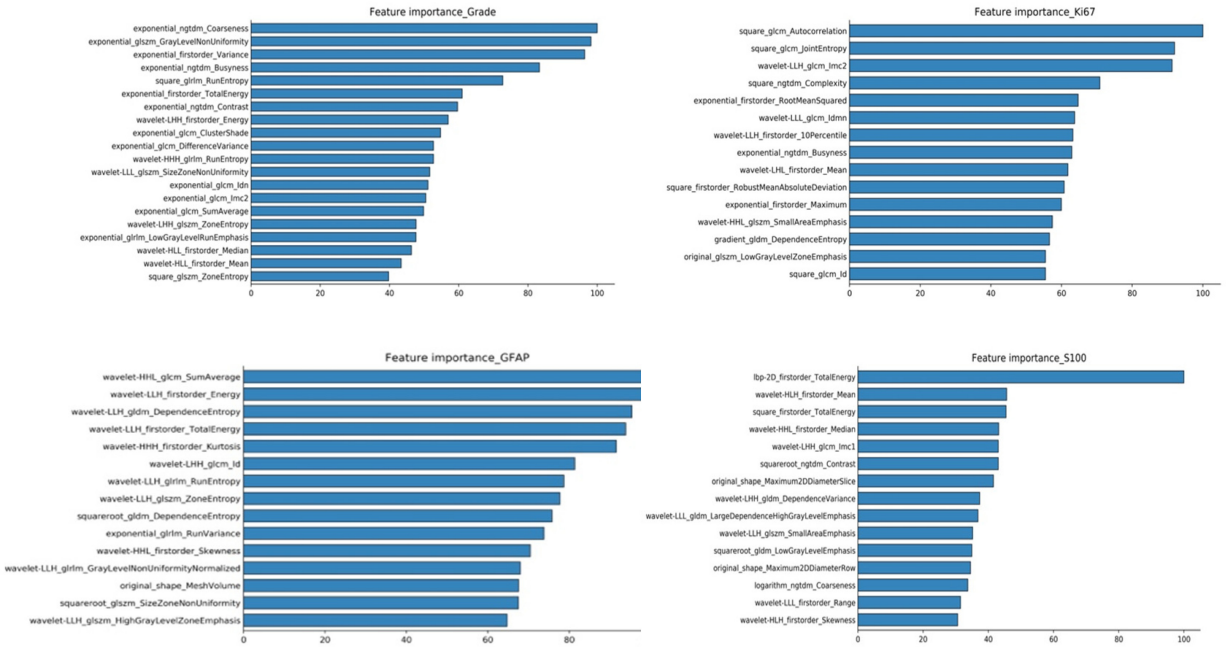


FIGURE 2 | RF model inbuild feature importance for predicting glioma grades and biomarkers of Ki67, GFAP, and S100.

TABLE 4 | The performance of predictive models.

Models	Error rate	True positive rate	True negative rate	AUC	Score (mean accuracy)
Logistic_Ki67	0.22857	0.787234	0.73913	0.799	0.771429
SVM_Ki67	0.25714	0.851064	0.521739	0.748	0.742857
Random Forest_Ki67	0.2	0.914894	0.565217	0.849	0.8
Logistic_GFAP	0.24324	0.615385	0.786885	0.774	0.756757
SVM_GFAP	0.21622	0.153846	0.918033	0.613	0.783784
Random Forest_GFAP	0.18919	0.076923	0.967213	0.718	0.810811
Logistic_S100	0.19118	0	0.859375	0.164	0.808824
SVM_S100	0.11765	0	0.9375	0.48	0.882353
Random Forest_S100	0.08824	0	0.96875	0.604	0.911765

We selected LR, SVM, and RF as classifiers mainly for their popularity. LR, SVM, and RF classifiers can work on non-text data set less than 100K. Whether the data is linearly divisible or not, the linearly separable models (LR, SVM), and the non-linear separable model (RF) are helpful to view the effect and avoid the impact due to poor data. LR shows a higher AUC, in GFAP's prediction model, but performs worst in S100's prediction. Comparing the overall results from three biomarker prediction models, the combination of PCA reduction and RF classification consistently performed best. It suggests a common ML pipeline that may be helpful in standardizing the prediction process of multiple protein expressions.

Also more recently, researchers have demonstrated achievements of deep learning (DL) in the image segmentation and glioma grades prediction (32–37). Convolutional neural networks (CNNs) started outperforming other methods on several high-profile image analysis projects. DL has advantages in computation, as high-performance graphics processing unit (GPU) supports fast computing and less time on modeling. Like a kind of end-to-end learning, DL can automatically extract relevant functions from images, and tasks such as raw data processing and classification can be completed automatically. However, DL is complex and requires thousands of images to start with, otherwise due to a relatively small collection of images like ours, overfitting is more likely. The classic ML methods met our needs and suited the data. RF models performed well for predicting glioma grades and pathologic biomarkers S100, Ki67, and GFAP.

As it is known, the roles of these biomarkers can be complicated and controversial in laboratory experiments (26). In addition to the abilities of predicting tumor phenotypes, radiomics might offer a new approach to evaluate biomarkers, since their differentiation can be identified through the analysis of imaging features. The expression level of Ki67 was significantly correlated with the tumor grade and tumor volume, as well as the patient age and gender. A study once reported that the high level

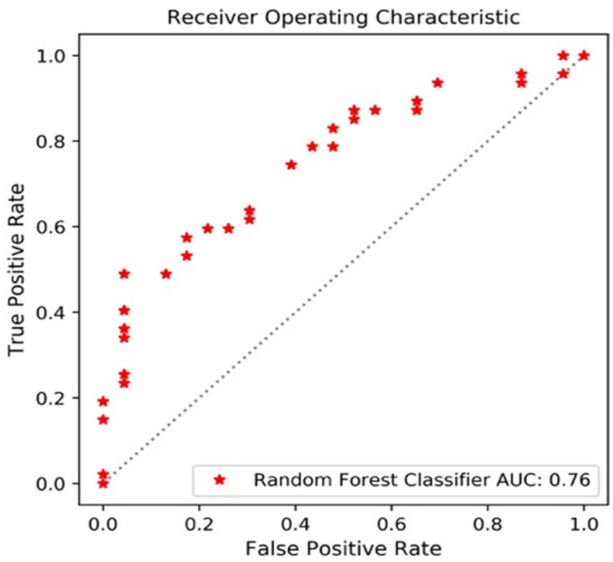


FIGURE 3 | AUC_ROC for the RF classifier.

of Ki-67 expression was correlated to poor overall survival (OS) and progression free survival (PFS) (16). The accurate prediction of high level Ki67 is more meaningful than its low level expression to indicate poor prognosis for glioma patients.

The GFAP has been widely expressed in gliomas. Merely four patients presented as GFAP negative. The majority of the patients (323 of 367, 88%) had GFAP positive (+), and 327 patients with low expression GFAP (90%), combined with four that scored (−), were distributed all over the gliomas grades, including low grade (132, 40%), and high grade (195, 60%). The minority of the patients (40 of 367, 12%) had GFAP medium positive (++) or high positive (+++) distributed in low grade (15, 37.5%) and high grade (25, 62.5%). In the literature, a high GFAP expression is likely to be found in low grade gliomas. The present result was confusing, that is, the high and low expression levels of GFAP were more correlated to high grade gliomas. This result may echo that GFAP is not a direct predictor of low grade gliomas (15, 26). On the classification report of the RF_GFAP model, the accuracy score of predicting a GFAP low expression was up to, while that of predicting high expression levels of GFAP was much lower. The overall prediction performance might not be meaningful, since GFAP was lowly expressed in 90% of patients, and the model could always answer 90% correctly. The same problem was found in the predictive model of S100. It required the rethinking of these two models. There was a need to determine which expression class is more valued. And then, as one solution, the ROC thresholds can tuned, increasing the sensitivity of the favored class.

The interpretation of the predicted results is complex, but may be helpful to understand the molecular mechanisms it underlies. In addition, the investigators selected CE MRI from several typical cases for demonstration, in which the different expression levels of biomarkers exhibited different imaging characteristics (**Figure 4**). For the high expression of S100 case (**Figure 4A**), the tumor exhibited an obvious rosette enhancement, no enhancement of internal necrotic components, and a few edema zones around it, and was diagnosed as glioblastoma (WHO IV grade). In the image of the tumor with a low expression of S100 (**Figure 4B**), the tumor mass effect was obvious, but there was no obvious enhancement, and the surrounding edema was not obvious, which was diagnosed as astrocytoma (WHO II grade).

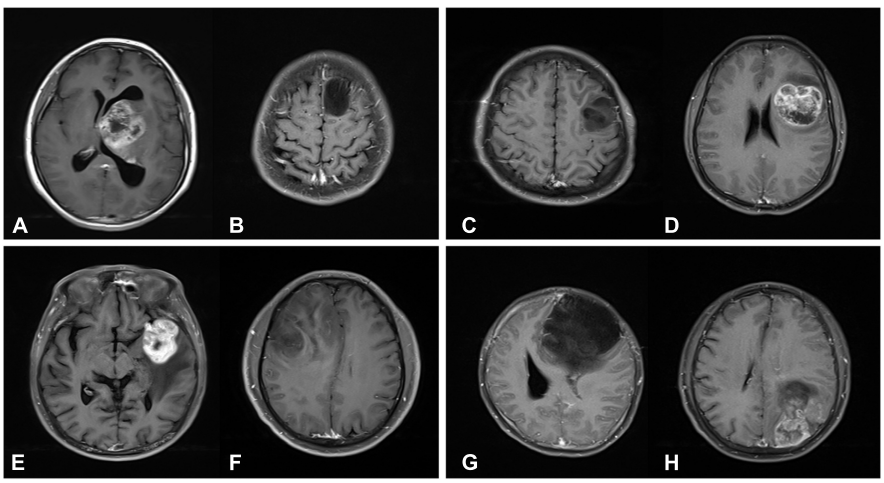


FIGURE 4 | T1-weighted contrast-enhanced MR images. **(A)** A 23-year-old female patient with a grade IV glioma in left thalamus. The expression of S100β is strongly positive (S100β+++). **(B)** A 23-year-old male patient with a grade II glioma in left frontal lobe. The expression of S100β is weakly positive (S100β+). **(C)** A 27-year-old male patient with a grade II glioma in left frontal lobe. The expression of GFAP is strongly positive (GFAP+++). **(D)** A 27-year-old female patient with a grade IV glioma in left frontotemporal lobe. The expression of GFAP is weakly positive (GFAP+). **(E)** A 64-year-old male patient with a grade IV glioma in left frontotemporal lobe. The Ki67 index is 80%. **(F)** A 44-year-old male patient with a grade II glioma in right frontal lobe. The Ki67 index is 80%. **(G)** A 31-year-old female patient with a grade II glioma in left frontal lobe. Genetic test showed that IDH1 was mutant type. **(H)** A 50-year-old male patient with a grade IV glioma in left parietal-occipital lobe. Genetic test showed that IDH1 was wild type.

In this case, the positive correlation appeared as both the S100 and glioma grade moved in the same direction that was contrary to many observations. The study conducted by Wang et al. has proven that S100 is expressed in most gliomas, and that this is an important inducer of CCL2 (19). CCL2 participates in the transport of tumor-associated macrophages (TAM) in gliomas, which affects angiogenesis, invasion, local tumor recurrence and immunosuppression. This may explain the relationship between the degree of tumor enhancement and the expression of S100 in the present cases.

There are some limitations in our study. First, we only used conventional MRI sequences with a default set of tumor features extracted by Pyradiomics. Advanced MRI sequences (e.g., DWI, DKI, MRS, ASL, et al.) can reflect the microstructure and metabolic information of tumors. In future study, we will further investigate the molecular phenotype of gliomas using a multimode magnetic resonance scheme. Second, we only selected 3 common pathologic biomarkers for gliomas from a wide range of biomarkers either current available or under investigation. We have to develop an evaluation plan for other glioma biomarkers and find candidates that can be benefit from radiomics applications. Third, imbalance classes did not reflect the incidences of glioma in real world, where glioblastoma is the most common subtype, and grade I glioma is relatively rare in adults. We used the SMOTE algorithm to balance data, oversampling the minority class, but the differences in data distribution cannot be ignored. In our experiments, before and after the use of SMOTE, AUC was only changed slightly. A larger dataset from multiple sites is expected to complement predictive effects, and the resulting classifiers can be more accurate and stable. Fourth, after PCA reducing feature dimensions, a new set of features was less remained but difficult to interpret. A combination of hierarchical clustering on PCA may help us to select feature more efficiently. At the current stage, a real-world application is out of our scope, but further prospective assessment is warranted. Based on the results we obtained as a reference, we will extend the study to identify the best classifier algorithm and the best set of features to simplify the classification tasks. The standardized computation methods would greatly enhance the reproducibility of radiomics studies, and it may also lead to standardized software solutions available in clinical practice.

In conclusion, the machine-learning based radiomics application provided a non-invasive approach for the prediction of glioma grades and expression levels of multiple pathologic biomarkers, with favorable predictive accuracy and stability. The study also demonstrated the potential of radiomics for pathological assessment and individualized cancer treatment.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of the second Xiangya hospital of central south university. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JL, MG, and SH: conception and design, and provision of study materials or patients. JL and RY: administrative support. MG, SH, XP, XL, and JL: collection and assembly of data. MG, SH, XP, and JL: data analysis and interpretation. All authors: writing and final approval of the manuscript.

FUNDING

This study was funded by the National Natural Science Foundation of China (81671671 and 61971451), the key R&D projects in Hunan Province (2019SK2131), Key Emergency Project of Pneumonia Epidemic of novel coronavirus infection (2020SK3006), and the Guiding Project of Clinical Medical Technology Innovation in Hunan Province (S2018SFYLJS0110).

ACKNOWLEDGMENTS

The authors express their appreciation to Ying Zeng for the acquisition, analysis, and interpretation of data for the work.

REFERENCES

1. Louis D, Ohgaki H, Wiestler O, Cavenee W, Burger P, Jouvet A, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropatholo.* (2007) 114:97–109. doi: 10.1007/978-94-007-1399-4_10
2. Omuro A, DeAngelis L. Glioblastoma and other malignant gliomas: a clinical review. *JAMA.* (2013) 310:1842–50. doi: 10.1001/jama.2013.280319
3. Ostrom QT, Gittleman H, Farah P, Ondracek A, Chen Y, Wolinsky Y, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2006-2010. *Neuro Oncol.* (2013) 15(Suppl. 2):1–56. doi: 10.1093/neuonc/not151
4. James M, Rafay A, Matthew O, Frank L, Misun H. Malignant gliomas: current perspectives in diagnosis, treatment, and early response assessment using advanced quantitative imaging methods. *Cancer Manag Res.* (2014) 6:149–70. doi: 10.2147/cmar.s54726
5. Jackson R, Fuller G, Abi-Said D, Lang F, Gokaslan Z, Shi W, et al. Limitations of stereotactic biopsy in the initial management of gliomas. *Neuro Oncol.* (2001) 3:193–200. doi: 10.1215/15228517-3-3-193
6. Kristensen BW, Priesterbach-Ackley LP, Petersen JK, Wesseling P. Molecular pathology of tumors of the central nervous system. *Ann Oncol.* (2019) 30:1265–78. doi: 10.1093/annonc/mdz164
7. García-Figueiras R, Baleato-González S, Padhani A, Luna-Alcalá A, Vallejo-Casas J, Sala E, et al. How clinical imaging can assess cancer biology. *Insights Into Imaging.* (2019) 10:28. doi: 10.1186/s13244-019-0703-0
8. Villanueva-Meyer JE, Mabray MC, Soonmee C. Current clinical brain tumor imaging. *Neurosurgery.* (2017) 81:3. doi: 10.1093/neuros/nyx103

9. Chaddad A, Kucharczyk M, Daniel P, Sabri S, Jean-Claude B, Niazi T, et al. Radiomics in glioblastoma: current status and challenges facing clinical implementation. *Front Oncol.* (2019) 9:374. doi: 10.3389/fonc.2019.00374
10. Griethuysen JJMV, Fedorov A, Parmar C, Hosny A, Aerts HJWL. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* (2017) 77:e104–7. doi: 10.1158/0008-5472.can-17-0339
11. Zhang B, Chang K, Ramkissoon S, Tanguturi S, Bi WL, Reardon DA, et al. Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro Oncol.* (2016) 19:109–17. doi: 10.1093/neuonc/now121
12. Lu CF, Hsu FT, Hsieh LC, Kao YCJ, Cheng SJ, Hsu BK, et al. Machine learning-based radiomics for molecular subtyping of gliomas. *Clin Cancer Res.* (2018) 24:4429–36. doi: 10.1158/1078-0432.ccr-17-3445
13. Kickingereder P, Bonekamp D, Nowosielski M, Kratz A, Sill M, Burth S, et al. Radiogenomics of glioblastoma: machine learning–based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. *Radiology.* (2016) 2016:161382.
14. Louis DN, Perry A, Reifenberger G, Von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* (2016) 131:803–20. doi: 10.1007/s00401-016-1545-1
15. Paulus W. GFAP, Ki67 and IDH1: perhaps the golden triad of glioma immunohistochemistry. *Acta Neuropathol.* (2009) 118:603. doi: 10.1007/s00401-009-0600-6
16. Chen WJ, He DS, Tang RX, Ren FH, Chen G. Ki-67 is a Valuable prognostic factor in gliomas: evidence from a systematic review and meta-analysis. *Asian Pac J Cancer Prev.* (2015) 16:411–20. doi: 10.7314/apjcp.2015.16.2.411
17. Burger PC, Shibata T, Kleihues P. The use of the monoclonal antibody Ki-67 in the identification of proliferating cells: application to surgical neuropathology. *Am J Surg Pathol.* (1986) 10:611–7. doi: 10.1097/00000478-198609000-00003
18. Torp SH. Diagnostic and prognostic role of Ki67 immunostaining in human astrocytomas using four different antibodies. *Clin Neuropathol.* (2002) 21:252–7.
19. Wang H, Zhang L, Zhang IY, Chen X, Fonseca AD, Wu S, et al. S100B promotes glioma growth through chemoattraction of myeloid-derived macrophages. *Clin Cancer Res An Off J Am Assoc Cancer Res.* (2013) 19:3764–75. doi: 10.1158/1078-0432.ccr-12-3725
20. Hessian PA, Fisher L. The heterodimeric complex of MRP-8 (S100A8) and MRP-14 (S100A9). *Eur J Biochem.* (2003) 268:353–63. doi: 10.1046/j.1432-1033.2001.01894.x
21. Ridinger K. S100A13. Biochemical characterization and subcellular localization in different cell lines. *J Biol Chem.* (2000) 275:8686–94. doi: 10.1074/jbc.275.12.8686
22. Hsu K, Champaiboon C, Guenther BD, Sorenson BS, Khammanivong A, Ross KF, et al. Anti-infective protective properties of S100 calgranulins. *Anti Inflamm Anti Allergy Agents Med Chem.* (2009) 8:290–305. doi: 10.2174/187152309789838975
23. Girolamo PD. Biology of the S100 proteins–Introduction. *Micros Res Tech.* (2003) 60:537–9. doi: 10.1002/jemt.10295
24. Petzold A. Glial fibrillary acidic protein is a body fluid biomarker for glial pathology in human disease. *Brain Res.* (2015) 1600:17–31. doi: 10.1016/j.brainres.2014.12.027
25. Cotrina ML, Chen M, Han X, Iliff J, Ren Z, Sun W, et al. Effects of traumatic brain injury on reactive astrogliosis and seizures in mouse models of Alexander disease. *Brain Res.* (2014) 1582:211–9. doi: 10.1016/j.brainres.2014.07.029
26. Bodegraven EJV, Asperen JVV, Robe PAJ, Hol EM. Importance of GFAP isoform−specific analyses in astrocytoma. *Glia.* (2019) 67:1417–33. doi: 10.1002/glia.23594
27. Yan T, Shuai-Tong Z, Jing-Wei W, Dong D, Xiao-Chun W, Guo-Qiang Y, et al. A radiomics nomogram may improve the prediction of IDH genotype for astrocytoma before surgery. *Eur Radiol.* (2019) 29:3325–37. doi: 10.1007/s00330-019-06056-4
28. Yiming L, Zenghui Q, Kaibin X, Wang K, Fan X, Li S, et al. Radiomic features predict Ki-67 expression level and survival in lower grade gliomas. *J Neuro Oncol.* (2017) 135:317–24. doi: 10.1007/s11060-017-2576-8
29. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage.* (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015
30. Swami A, Jain R. Scikit-learn: machine learning in python. *J Mach Learn Res.* (2013) 12:2825–30.
31. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. *Bmc Bioinformatics.* (2013) 14:1–16. doi: 10.1186/1471-2105-14-106
32. Mzoughi H, Njeh I, Wali A, Slima MB, Mahfoudhe KB. Deep multi-Scale 3D convolutional neural network (CNN) for MRI gliomas brain tumor classification. *J Digit Imaging.* (2020). doi: 10.1007/s10278-020-00347-9 [Epub ahead of print].
33. Ying Z, Ning H, Mathen P, Cheng JY, Krauze AV, Camphausen K, et al. Automated glioma grading on conventional MRI images using deep convolutional neural networks. *Med Phys.* (2020) 47:3044–53. doi: 10.1002/mp.14168
34. Matsui Y, Maruyama T, Nitta M, Saito T, Tsuzuki S, Tamura M, et al. Prediction of lower-grade glioma molecular subtypes using deep learning. *J Neuro Oncol.* (2020) 146:321–7. doi: 10.1007/s11060-019-03376-9
35. Han W, Qin L, Bay C, Chen X, Yu K, Miskin N, et al. Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas. *AJNR Am J Neuroradiol.* (2020) 41:40–8. doi: 10.3174/ajnr.a6365
36. Bangalore Yogananda C, Shah B, Vejdani-Jahromi M, Nalawade S, Murugesan G, Yu F, et al. A novel fully automated MRI-based deep-learning method for classification of IDH mutation status in brain gliomas. *Neuro Oncol.* (2020) 22:402–11. doi: 10.1101/757385
37. Chang K, Bai H, Zhou H, Su C, Bi W, Agbodza E, et al. IDHResidual convolutional neural network for the determination of status in low- and high-grade gliomas from MR imaging. *Clin Cancer Res.* (2018) 24:1073–81. doi: 10.1158/1078-0432.CCR-17-2236

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

METHODS
published: 17 September 2020
doi: 10.3389/fbioe.2020.526814

Multiparametric Color Tendency Analysis (MCTA): A Method to Analyze Several Flow Cytometry Labelings Simultaneously

Andrea Henriques-Pons[1], Carine P. Beatrici[2], Juan Camilo Sánchez-Arcila[3] and Fabricio Alves Barbosa da Silva[2]*

[1] Laboratório de Inovações em Terapias, Ensino e Bioprodutos, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, [2] Scientific Computing Program, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, [3] Department of Natural Sciences, University of California, Merced, Merced, CA, United States

Despite the remarkable evolution of flow cytometers, fluorescent probes, and flow cytometry analysis software, most users still follow the same ways for data analysis. Conventional flow cytometry analysis relies on the creation of dot plot sequences, based on two fluorescence parameters at a time, to evidence phenotypically distinct populations. Thus, reaching conclusions about the biological characteristics of the samples is a fragmented and challenging process. We present here the MCTA (Multiparametric Color Tendency Analysis), a method for data analysis that considers multiple labelings simultaneously, extending and complementing conventional analysis. The MCTA method executes the background fluorescence exclusion, spillover compensation, and a user-defined gating strategy for subpopulation analysis. The results are then presented in conventional FSC x SSC dot plots with statistical data. For each event, the method converts each of the multiple fluorescence colors under analysis into a vector, with longer vectors being attributed to more intense labelings. Then, the MCTA generates a resultant vector, which is therefore mostly influenced by predominant labelings. The radial position of this resultant vector corresponds to a resultant color, making it easy to visualize phenotypic modulations among cellular subpopulations. Besides, it is a deterministic method that quickly assigns a resulting color to all events that obey the gating strategy, with no polymeric regions defined by the user or downsampling. The MCTA application generates a single dot plot showing all events in the FCS file, but a resultant color is attributed only to those that obey the gating strategy. Therefore, it can also help to evidence rare events or unpredicted subpopulations naturally excluded from the regions defined by the user. We believe that the MCTA method adds a new perspective over multiparametric flow cytometry analysis while evidencing modulations of molecular labeling profiles based on multiple fluorescences. Availability and implementation: The instructions for the MCTA application is freely available at https://github.com/flowcytometry/MCTA.

Keywords: flow cytometry, multiparametric data analysis, software, method, t-SNE, MCTA

OPEN ACCESS

Edited by:
Francesco Rundo, STMicroelectronics (Italy), Italy

Reviewed by:
Honglin Jin, Huazhong University of Science and Technology, China
Hirak K. Patra, University College London, United Kingdom

****Correspondence:***
Andrea Henriques-Pons
andreah@ioc.fiocruz.br

Specialty section:
This article was submitted to Biomaterials, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *25 February 2020*
Accepted: *26 August 2020*
Published: *17 September 2020*

Citation:
Henriques-Pons A, Beatrici CP, Sánchez-Arcila JC and da Silva FAB (2020) Multiparametric Color Tendency Analysis (MCTA): A Method to Analyze Several Flow Cytometry Labelings Simultaneously. Front. Bioeng. Biotechnol. 8:526814. doi: 10.3389/fbioe.2020.526814

INTRODUCTION

The flow cytometry technique offers quantitative fluorescence-based data, usually regarding cellular characteristics, at a rate of hundreds of events per second (Shapiro, 2004). Modern flow cytometers detect more than 15 parameters of fluorescence per event, evaluating cellular phenotype, viability, and proliferation; Ca^{++} levels; organelles activity; and much more. All technological advances of flow cytometry resulted in an enormous development in biology and medicine. Still, for reproducible results, all preliminary steps must be carefully planned, considering data generation, data pre-processing and quality control, visualization of results, and final data analysis (Saeys et al., 2016).

Despite the increasing complexity and evolution of multiparametric flow cytometry, conventional data analysis is still based on the evaluation of up to two fluorescence parameters at a time, relying on the creation of multiple dot plots (Mair et al., 2016). The gating strategy also obeys a logical hierarchical sequence of regions drawn manually, and users must reach an experimental conclusion after a fragmented analysis (de Oliveira et al., 2007; Cascabulho et al., 2016). Although flow cytometrists are used to this method, some disadvantages are the manual and imprecise definition of regions (gates), the underestimation and low visibility of rare events, and the fact that minor subpopulations outside the user-defined regions are not considered in the analysis.

To address the lack of exploratory analysis of flow cytometry data and the issues with reproducibility, several algorithms and computational tools have been developed (Pedreira et al., 2019). For instance, some dimensionality reduction algorithms used in flow cytomety data analysis are t-distributed stochastic neighbor embedding (t-SNE) (Toghi Eshghi et al., 2019) and Uniform Manifold Approximation (UMAP)[1]. Both are dimension reduction algorithms that favor the preservation of local distances over global distance. Other methods use clustering algorithms that show spanning trees as a result, like in Spanning-tree Progression Analysis of Density-normalized Events (SPADE) (Mair et al., 2016) and FlowSOM (Van Gassen et al., 2015). In these methods, the cells are represented by nodes connected with the neighbors in the high dimensional data. Therefore, the interconnected nodes are related to phenotypically similar cells, as shown by the CITRUS algorithm (Bruggner et al., 2014). Moreover, subpopulation detection in high-dimensional data can be analyzed using PhenoGraph (Levine et al., 2015). An excellent review of these and other complementary strategies can be found in Weber and Robinson (2016).

Typical workflows in computational tools include data transformation, normalization, filtering, manual or semi-automatic gating, and automatic clustering (Montante and Brinkman, 2019). However, only advanced users with programming skills will be able to go through the analysis process. For regular users, manual efforts remain standard practice, which has been mostly the same for decades, regardless of the commercial analysis software used. To address these issues, we present the MCTA (Multiparametric Color Tendency Analysis). This is an alternative exploratory method that analyzes multiple phenotypic markers simultaneously and evidences complex cellular profiles, different from the multi-step conventional analysis. The method excludes the background range of each channel for each event, according to negative controls, and spillover fluorescence is compensated. Moreover, the gating procedure for subpopulation analysis is done in a single step, including and excluding multiple chosen cellular markers according to the user rationale for cellular identification. Then, to show the phenotypic result, the algorithm attributes a base vector directly proportional to the labeling intensity of each fluorescence parameter to be analyzed, and a resultant vector is calculated for each event. According to the radial position of the resultant vector, a different resultant color is attributed to each event. Therefore, the resultant color is determined by the predominant labeling or labelings under analysis. In the MCTA analysis, biological modulations of experimental target molecules are easily visualized by different resultant colors, and the results are backed up by statistical analysis for data interpretation. Moreover, the MCTA method maps predominant phenotypic profiles on conventional morphology FSC-A vs. SSC-A dot plots, an option that standard multivariate algorithms do not offer.

COMPUTATIONAL METHOD

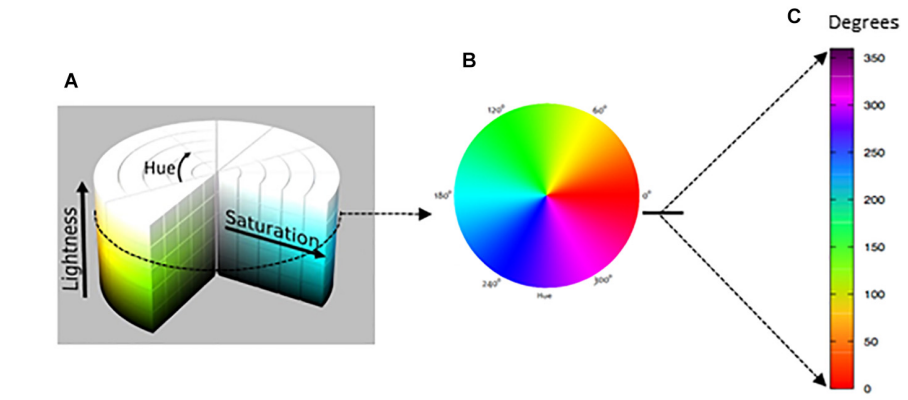
Color Representation in the MCTA Method

The color assignment adopted in this work is based on the HSL representation (Hu et al., 2014) (**Figure 1A**). HSL stands for hue, saturation, and lightness (or luminosity) and consists of a cylindrical-coordinate representation of points in an RGB (red, green, blue) color model. In this representation, the RGB coordinates are geometrically arranged in an attempt to be more intuitive (**Figure 1A**). Developed in the 1970s for computer graphics applications, the HSL is used today in color pickers, image-editing software, and less commonly in image analysis and computer vision (Tsai and Tseng, 2012).

In the general HSL representation, each color is a dot in a cylinder, the angle around the central vertical axis corresponds to "hue," the distance from the axis corresponds to "saturation," and the distance along the axis corresponds to "lightness" (**Figure 1A**). For flow cytometry, the HSL was chosen to represent fluorescence data because each color can be represented as an individual vector with a corresponding angular value. The proportional intensity of each labeling (fluorescence parameter) is represented by the saturation, with longer vectors representing more intense fluorescence labelings. For a bi-dimensional representation, we considered lightness as a constant factor (**Figures 1A,B**). Then, the spectrum was represented as a linear rule with the hue values (fluorescence colors) ranging from 0 to 360° or *0 rad* to 2π *rad* (**Figure 1C**).

Once the MCTA method uses all labeling colors determined by the user simultaneously, each fluorescence parameter must be identified for the vectorial representation per event. For this, our method uses the maximal emission value of each fluorochrome

[1]https://arxiv.org/abs/1802.03426



D

Fluorochrome	Maximal emission	Background cut
FITC	525	10.09
PE	575	17.35
PerCP	675	46.76
APC	660	44.37
APC Cy7	774	13.47

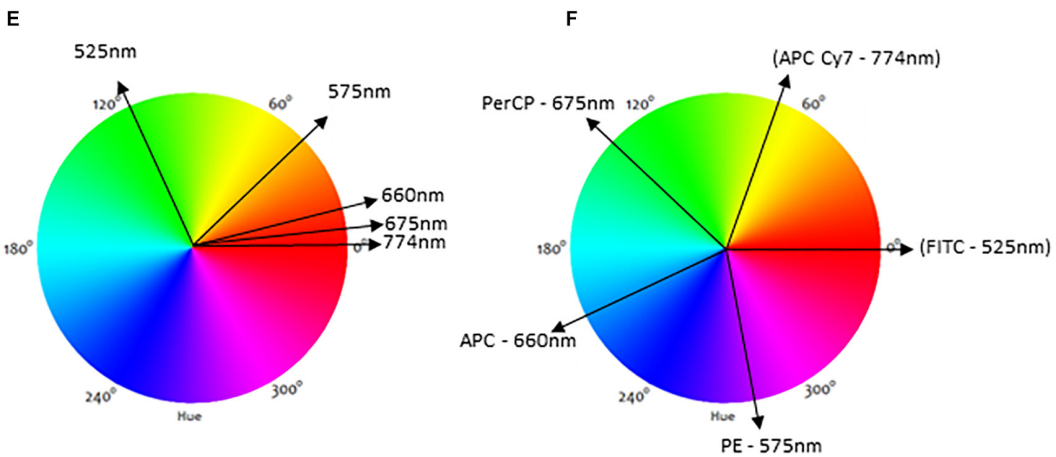


FIGURE 1 | The HSL (hue, saturation, lightness) representation and the basis for data analysis. The conventional HSL representation is a cylinder **(A)**, but for flow cytometry analysis, we considered lightness as a constant factor for a bi-dimensional diagram **(B)**. For the representation next to the dot plots, a linear version of the bidimensional HSL was produced with reference degrees indicated **(C)**. For the analysis, each fluorochrome is indicated by its maximal emission of fluorescence, and the background (negative range) is required **(D)**. When each fluorochrome is represented in a circular bidimensional HSL diagram, different parameters with similar maximal emissions are closely indicated **(E)**. To avoid this overlapping color representation, our method automatically divides the HSL into equal parts **(F)**.

used (**Figure 1D**) and projects a correspondent vector over the bi-dimensional HSL representation (**Figure 1E**). As many fluorochromes have similar or even the same maximal emission value, the MCTA method divides the spectrum (360°) into equal parts automatically (**Figure 1F**). Therefore, in the MCTA method, the color attributed to each fluorescence parameter does not correspond to the real emission of the fluorochrome (**Figure 1F**).

Event Representation in the MCTA Method

The MCTA method computes a resultant vector to each event, using all labeling colors chosen at the same time. This resultant vector determines a resultant color observed in the HSL representation and will be most affected by the predominant labeling(s). This analysis, therefore, aims to show the tendency of labeling(s) with higher median fluorescence intensity (MFI) in each event. When comparing different experimental groups, such as uninfected and pathogen-infected mice, for example, this analysis will quickly show if there was a difference in the repertoire of mostly expressed molecules between samples, showing the cellular profile.

In **Figure 2A**, we illustrate the generation of the resultant vector using an event that was labeled with different fluorochromes, but only two of these parameters were used to generate the resultant vector. Therefore, the spectrum

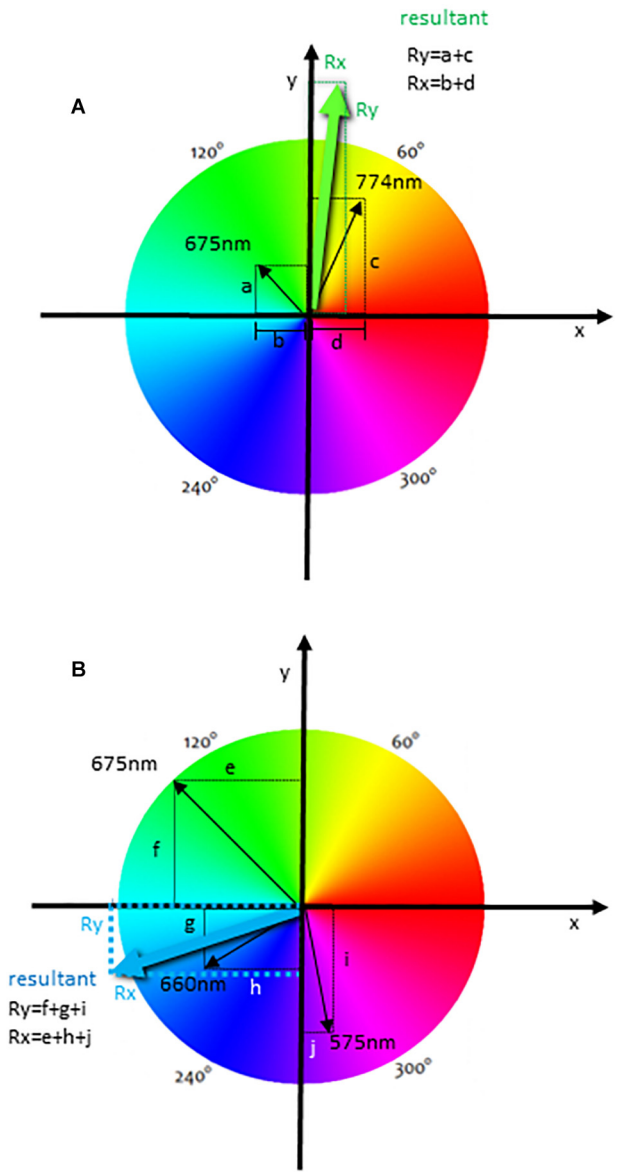


FIGURE 2 | The calculation of a resultant vector using the HSL representation in flow cytometry analysis. The diagrams show one event with two labelings (maximal emissions are 675 and 774 nm) **(A)**, and with three labelings (maximal emissions of 575, 660, and 675 nm) **(B)** used for the resultant vector calculation. Black arrows indicate fluorescence labeling colors, and the length of these arrows is directly proportional to the intensity of each labeling. These vectors project segments over the X- and Y-axis, indicated by letters, and the resultant vector indicating the resultant color is calculated using all segments for each event. The resultant segments are indicated as Rx and Ry.

was divided into equal parts according to the number of fluorescences, but only two vectors were accounted for the resultant color. In this case, we represented the PerCP (maximal emission of 675 nm) and APC Cy7 (774 nm) (**Figure 1D**) labelings. The intensity of each labeling was proportionally represented by the length of each vector (represented in black) (**Figure 2A**). These vectors project segments over the axes X and Y, which are "b" and "a" for the fluorochrome 675 nm and "d" and "c" for the fluorochrome 774 nm (**Figure 2A**). The resultant vector was produced by the segment "d" minus "b" projected over the X-axis Rx (as the segment b is below zero) (Rx) and by the summon of the segments "a" and "c" projected over the Y-axis (Ry) (**Figure 2A**). The calculated segments Rx and Ry determined the resultant vector, indicating the color to be attributed to the event (**Figure 2A**). In another analysis, we illustrate the generation of the resultant vector using three of the experimental labelings (**Figure 2B**), which were 575, 660, and 675 nm (**Figure 1D**). In this case, the resultant vector was generated by the projection of the segments "e," "h," and "j" over the X-axis (Rx) and by the segments "f," "g," and "i" over the Y-axis (Ry). In **Figure 2** we show that the resultant vector for each event was obtained as a function of the angle (hue) of each fluorescence vector and the intensity of individual color labelings.

Formally, for each wavelength L_j we associate one hue value h_j, which is a degree on the color wheel and is directly related to the vector's $\vec{v}_j$ angle. The j-th unit vector is given by:

$$\hat{v}_j = \cos(h_j)\,\hat{x} + \sin(h_j)\,\hat{y} \tag{1}$$

The resultant color is calculated considering all fluorescence channels that were chosen and their intensities (saturation) as bi-dimensional vectors. The final color of the event will be the vectorial sum of all fluorescence colors chosen, with the intensity of the fluorescence channel for each event I_j being taken into account. The resultant vector of the i-th event is given by:

$$\vec{t}_i = \sum_{j=1}^{n_c} I_j\,\hat{v}_j \tag{2}$$

where the I_j is the intensity (saturation) of the j-th channel of the event, n_c is the number of channels and $\hat{v}_j$ is the filter unit vector.

Configuration File

The MCTA method follows required steps for any flow cytometry analysis, which are the compensation of spillover fluorescence (illustrated in **Box 1**) and the definition of fluorescence intensity thresholds (background) (**Figure 1D**). Here we present the MCTA analysis using two sources of FCS files. We used a simple five-color labeling experiment uploaded to the Github repository[2] for a detailed description of the method. Then, to demonstrate the application of the algorithm, we used flow cytometry FCS files available in a public repository[3]. For the results section, we used data that evaluated T lymphocyte response after Staphylococcal enterotoxin B (SEB) stimulation (repository ID FR-FCM-ZZEC, 15 colors) and a panel that identifies human adaptive natural killer (NK) cells (repository ID FR-FCM-ZYY6, 13 colors). It is important to highlight that, for each flow cytometry analysis in the MCTA method, a configuration file is created by the user, applying specific logical and syntax rules depicted in **Figure 3**. This example details the in.dat file containing all required

[2]https://github.com/flowcytometry/MCTA

[3]https://flowrepository.org/public_experiment_representations

BOX 1 | Compensation factors for spillover fluorescences (https://github.com/flowcytometry/MCTA).

	FITC	PE	PerCP	APC	APC Cy7
FITC	100.00	1.40	0.00	0.00	0.00
PE	18.22	100.00	0.00	0.00	0.00
PerCP	2.80	15.42	100.0	0.00	0.00
APC	0.00	0.41	6.86	100.00	14.49
APC Cy7	0.00	0.00	0.00	6.54	100.00

parameters, considering original columns of the acquired FCS file (**Figure 3A** – FCS file) and the fluorescence parameters that must be renumbered to feed the algorithm (**Figure 3A** – column identification).

The first step is the identification of the FCS file under analysis (file name), which is SampleSEB.FCS, in this example (**Figure 3B**). Then, the user enters the total number of parameters of the FCS file and the ones that will be used in the MCTA method. In the example used in **Figure 3B** for lymphoid cells response against SEB, the FCS file had 23 columns (**Figure 3A**); however, only 18 columns were considered. Five columns (identified in red) were not used (time, FCS H, SSC H, and two empty channels) (**Figure 3B**). The next step is the definition of the total number of fluorescence parameters employed by the MCTA method. In this case, we used 15 (**Figure 3B**), which are the renumbered fluorescence parameters shown in the "column identification" (**Figure 3A**). The individual identification of what columns will be used is then entered. Note that the parameters that are shown in red (**Figure 3A**) were not included (columns 2, 4, 6, 17, and 21) (**Figure 3B**). As indicated before, the values of maximal emission of fluorochromes, background (negative range of fluorescence intensity per channel), and compensation factors are used in the MCTA method (**Figure 3B**). These values can be obtained from any flow cytometry analysis software and, in this case, were obtained using FlowJo version 10. Finally, in **Figure 3A**, we see that the renumbered fluorescence channels identified in the column "column identification" represent the labeling of the cellular markers identified in the column "molecule labeled."

One essential step for any flow cytometry analysis is the gating process, which is the definition of cellular subpopulations of interest. The gating procedure in the MCTA method is done including and excluding events that are positive for given cellular markers, and this process is illustrated in **Figure 3B** – resultant definition; gating strategy, **Figure 4**; and **Supplementary Figure 1**. In the example shown in **Figures 3B, 4**, the analysis was done only in the events that were positive for CD3, CD4, CCR7, and CD154 (parameters 5, 6, 9, and 15), illustrating that the analysis was done in activated $CD4^+$ T lymphocytes (**Figure 4**). The indication of exclusion parameters means that the MCTA method will not calculate the resultant color in any of the events that are positive for any of the exclusion parameters. In this case, there will be no resultant vector calculation for events identified as monocytes ($CD14^+$, parameter 4), dead cells (positive events for Live Dead labeling, also parameter 4), naïve T cells ($CD45RA^+$, parameter 11), NK or NKT cells ($CD56^+$, parameter 7), $CD8^+$ T lymphocytes (parameter 8), or T cells expressing IL-4 or IL-21 (parameters 2 and 10, respectively). In this example, although CXCR5 (parameter 13) and PD-1 (parameter 14) are molecules associated with T lymphocyte activation, very few events were expressing either molecule, and we decided to exclude both. In the MCTA method, the exclusion parameters are indicated by the minus sign (**Figure 3B** – resultant definition, gating strategy), and the events that are labeled with any of the exclusion parameters will be represented in black. Finally, after the gating procedure, the syntax to represent the parameter(s) used for the generation of the resultant color is "zero" (**Figure 3B** asterisk – resultant definition). In this particular case, IFN-γ, IL-2, and TNF-α were the parameters used for the MCTA analysis, indicated in the "column identification" as parameter 3, 12, and 1 (**Figures 3A**, **4**). However, in the configuration file, all three parameters are represented as zeros (**Figure 3B** – resultant definition; asterisk). In biological terms, this analysis targets the identification of Th1-responding activated $CD4^+$ T lymphocytes.

RESULTS

The MCTA Analysis

In **Figure 5**, we show the final result of the MCTA analysis, where SEB-activated $CD4^+$ T lymphocytes, according to the gating strategy, are represented as colored events in the FSC x SSC dot plot. The color of each event in **Figure 5A** was attributed according to predominant labeling(s) of the Th1 cytokines selected for the analysis (**Figure 4**). When the configuration file was set up to evaluate Th2 (IL-4-producing) $CD4^+$ T lymphocytes (**Figure 5B**), the analysis considered the inclusion and exclusion parameters shown in **Supplementary Figure 1**, and we observed no colored events. This result means that there were no stimulated $CD4^+$ T cells positive for IL-4 labeling.

At this point, it was clear that the MCTA method was feasible and able to evidence a different pattern of results, showing the combination of multiple labelings for a phenotypic profile of gated subpopulations. However, it was still necessary to identify what parameters mostly influenced the resultant color observed and, for this, we added the statistical analysis. We then illustrated in **Figure 5C**, the statistical analysis of the data shown in **Figure 5A**. The MCTA method calculates conventional statistics that are most useful in flow cytometry analysis. It indicates the total number of events in the file and, regarding the colored events, the number of events that met the gating strategy, geometric mean, geometric standard deviation, and median (**Figure 5C**). The statistical analysis in the MCTA method is not based on the original data of the FCS file; it uses the processed data matrix created after the subtraction of the background and spillover compensation. Therefore, the user can visualize if only one or more fluorescence parameters were predominantly labeled, and what parameter(s) mostly influenced the resultant vector (the resultant color) (**Figure 5C**).

A

FCS file		Identification of algorithm variables			
column number	collumn content	column identification	maximal emission	background cut	molecule labeled
1	event numbering	-	-	-	-
2	time	-	-	-	-
3	FCS A	-	-	-	-
4	FCS H	-	-	-	-
5	SSC A	-	-	-	-
6	SSC H	-	-	-	-
7	FITC	1	515	417	TNF-α
8	PerCP Cy5.5	2	694	578	IL-4
9	V450	3	448	457	IFN-γ
10	BV510	4	510	484	CD14/live dead
11	BV570	5	570	1300	CD3
12	BV605	6	602	1400	CD154
13	BV655	7	655	686	CD56
14	BV711	8	711	1100	CD8
15	BV786	9	786	470	CCR7
16	APC	10	660	484	IL-21
17	empty channel				
18	APC H7	11	786	679	CD45RA
19	PE	12	578	352	IL-2
20	PE efluor 610	13	607	931	CXCR5
21	empty channel				
22	PE Cy7	14	785	294	PD1
23	BUV395	15	395	1200	CD4

B

SampleSEB.fcs → File name

18 → number of columns of the FCS file that were used for the MCTA ("column number")

15 → Number of fluorescences used (from "column identification")

1 3 5 7 8 9 10 11 12 13 14 15 16 18 19 20 22 23 → Identification of the parameters used for the MCTA (from "column number")

515 694 448 510 570 602 655 711 786 660 786 578 607 785 395 → Maximal emission of fluorochromes

417 578 457 484 1300 1400 686 1100 470 484 679 352 931 294 1200 → Background cut

Compensation Table

100.00	2.9011	0.0278	1.7525	3.2841	1.8702	0.3215	0.2103	0.0046	0.00	0.00	0.6055	0.4902	0.0046	0.4424
0.0654	100.00	0.0322	0.086	0.3831	0.5285	11.6198	92.1812	8.5133	3.6023	5.3953	0.429	0.1398	12.0418	0.4911
0.0377	0.00	100.00	16.9397	9.3184	4.5506	0.6474	0.3494	0.0063	0.0063	0.00	0.0188	0.0189	0.00	0.8322
0.7417	0.654	30.7288	100.00	230.2867	150.0982	24.7047	15.2399	0.9364	0.5985	0.1085	2.4305	3.0473	0.2171	-0.1314
0.1097	1.7177	2.9991	0.4057	100.00	80.3195	17.4666	13.5614	0.9052	0.0421	0.00	13.7483	17.29	0.4975	0.3861
2.5982	9.0015	7.6587	2.9984	20.4306	100.00	31.9037	33.8266	5.9893	2.3297	0.3326	4.6589	23.0885	3.9915	10.6992
0.2372	1.5224	1.902	0.7123	1.7489	22.4699	100.00	91.1888	5.971	28.6861	2.017	0.1658	1.596	0.7344	2.571
0.0135	9.4552	1.099	0.1768	0.2442	0.2335	1.2116	100.00	12.4499	0.2518	3.6478	0.009	0.0271	0.5508	0.1032
0.0498	0.083	2.3756	0.3821	0.6258	12.99	0.7332	2.9157	100.00	0.1493	5.8173	0.00	0.00	6.1214	0.00
0.0163	0.7745	0.0136	0.019	0.0663	0.6453	5.174	4.1553	0.3799	100.00	5.7442	0.0081	0.1227	1.0644	0.0901
-1.2814	0.00	-1.2804	-3.8422	-7.7925	-10.4223	-5.1318	-9.0124	8.9604	0.00	100.00	-1.2801	-3.8464	14.0789	-6.5975
0.3235	13.9691	0.1437	0.3236	27.319	14.0529	2.4217	1.3408	0.00	0.1077	0.0718	100.00	80.0103	1.4366	3.1252
0.00	11.0832	0.00	0.00	1.0228	11.1998	2.0228	1.1838	0.0169	0.3051	0.0169	8.9081	100.00	1.682	0.3331
0.4371	1.6247	0.3743	0.4371	1.2693	0.8911	0.125	0.00	5.0015	0.00	4.1199	2.8097	2.5665	100.00	4.8301
0.0037	0.0055	0.0111	0.0037	0.0094	0.0169	0.0129	0.0279	0.0074	0.0055	0.0018	0.0055	0.0074	0.0037	100.00

5 6 9 -13 -14 15 -2 -4 -7 -8 -10 -11 | 0 0 0 → Resultant definition - Included and excluded fluorescence parameters ("collumn identification")

(gating strategy) *

FIGURE 3 | Logical and syntax rules for the configuration files. All flow cytometry files were organized in columns indicated here as "column number" and "column content" under "FCS file" **(A)**. All parameters shown in red were not used in the example depicted. For the production of the configuration file, the original parameters were renumbered, considering only the fluorescence colors, and all other parameters must be entered as indicated **(B)**. In the line identified as "resultant definition," each inclusion and exclusion parameter is indicated, and the parameters used for the MCTA's resultant vector are identified as zeros (**B**, asterisk).

When we analyzed activated $CD8^+$ T lymphocytes (**Figure 6**), we observed resultant color calculations only for cells producing Th1 cytokines (**Figure 6A**), and no cells produced IL-4 (**Figure 6B**). Moreover, when we analyzed the files obtained from the public repository corresponding to control cells incubated with DMSO, we observed no CD4 or CD8 T cells expressing any of the cytokines tested, as no resultant colors were generated (**Supplementary Figure 2**).

Analysis of Rare Populations in the MCTA Method

The MCTA method has important advantages over conventional analysis and other computational tools, such as the quick analysis of many fluorescence parameters simultaneously. Moreover, the resultant vector (color) is shown per event on conventional morphology dot plots, familiar to all flow cytometry users, and the gating strategy is not influenced by regions defined by the user. One advantage is especially important in the MCTA method, which is the fact that all events are shown in the dot plot, colored or not. In conventional flow cytometry analysis, a sequence of manual regions defines the events that will be analyzed. Therefore, minor or unpredicted subpopulations outside the gates defined by the user are automatically excluded from the analysis. In the MCTA method, however, this is not an issue, as all events that obey the gating strategy are shown as colored events, including the ones that would not be in expected drawn regions.

inclusion parameters		
parameter identification	cellular populations	labeling
5	CD3	subpopulation
15	CD4	subpopulation
9	CCR7	activation
6	CD154	activation
exclusion parameters		
parameter identification	cellular populations	labeling
-4 4	CD14/live dead	subpop./dead cells
-11	CD45RA	Naïve T cells
-7	CD56	subpopulation
-8	CD8	subpopulation
-2	IL4	Th2
-10	IL21	Th17
-13	CXCR5	activation
-14	PD-1	activation
resultant		
parameter identification	cellular populations	labeling
3	IFN-γ	Th1
12	IL2	Th1
1	TNF-α	Th1

FIGURE 4 | The gating procedure in the MCTA method. The gating process in the MCTA method is based on the definition of inclusion parameters numerically identified in "parameter identification." In this case, only events positively labeled for CD3, CD4, CCR7, and CD154, identified as parameters 5, 15, 9, and 6 indicated under "cellular subpopulation" and "labeling" will be used in the gating strategy. The exclusion parameters used in this example were identified as –4, –11, –7, –8, to exclude cellular subpopulations and naúve T cells, –2 and –10, to exclude Th2 and Th17 cells, and –13 and –14 as activation T cell markers expressed on few events. The resultant vector was calculated using only the parameters 3, 12, and 1.

Although the MCTA method can identify rare events, it can be challenging to visualize these few colored events in files that have many events; then, we created one interactive dot plot as an additional tool. In this case, the user can determine a narrow range of degrees to observe only the events displayed in that range of resultant colors. This dot plot was named Resultant Plot With Filter, and it excludes uncolored events and events outside the range selected (**Figures 7A–C**). Then, the statistical analysis corresponds only to those events that were included in the range defined (**Figure 7**). In this example, the ranges corresponded to 10 degrees, but the user delimitates the range according to individual results. It is important to highlight that for rare events identification, as in any flow cytometry analysis method, the number of acquired events will be critical for data visualization.

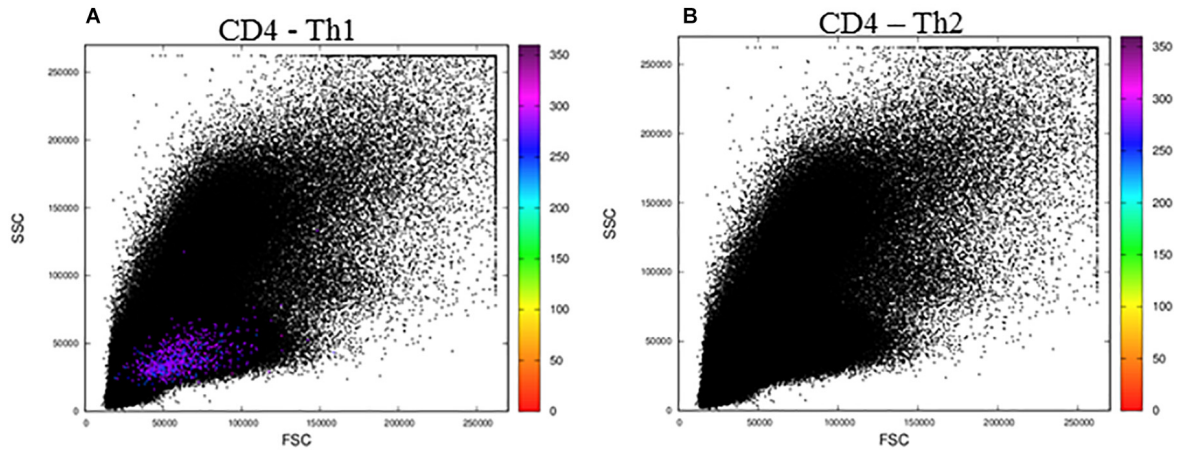
We further challenged the MCTA method using other files downloaded from the repository, and the next experiment evaluated the phenotype of adaptive and conventional NK cells (**Figure 8**). For this analysis, we arranged the fluorescence colors (parameters) as indicated in the **Supplementary Figures 3A–C**, and it shows that the original FCS file had 17 parameters (**Supplementary Figure 3B**). For the analysis of adaptive NK cells, we defined as exclusion parameters the events corresponding to dead cells, monocytes ($CD14^+$) and B lymphocytes ($CD19^+$) (all in parameter 4), T lymphocytes ($CD3^+$ events, parameter 12), and events positive for an adaptive NK cells marker that labeled few events ($CD57^+$ events, parameter 5) (**Supplementary Figure 3B**). The inclusion parameter corresponded to the expression of only NKG2C (adaptive NK cells, parameter 10) (**Supplementary Figure 3B** – indicated by the box "Resultant definition for adaptive NK cells") and for the MCTA analysis, we used a different proposal than the one used by the authors. For the resultant color calculation, we selected the channels that corresponded to two available columns and the expression of CD2, NKp30, CD7, ILT2, Siglec-7, CD56, and NKG2A (parameters 1, 8, 2, 3, 6, 7, 8, 9, 11, and 13, respectively) (**Supplementary Figures 3A,B**). Although we included the available parameters 1 and 8 in the analysis, these variables do not affect the resultant color, as predominant labelings mostly influence the result.

As a result of adaptive NK cells, we observed most events as orange cells (**Figure 8A**) and the statistical analysis, considering geometric mean and median, showed that parameter 6 (CD7 labeling) was the single predominant labeling that accounted for the resultant color in the MCTA analysis (**Figure 8B**).

For conventional NK cells, we used the exclusion parameters shown in **Supplementary Figure 3C** (parameters 4, 12, and 5), and the inclusion parameter was parameter 13 (NKG2A). The MCTA analysis revealed different subpopulations of $NKG2A^+$ cells, with orange, green, and violet as resultant colors (**Figure 8C**). This means that subpopulations of $NKG2A^+$ cells had different predominant markers, which would be difficult to observe in conventional analysis. In this particular case, despite the heterogeneous subpopulations (**Figure 8C**), the statistical analysis showed the parameter (channel) 6 as the predominant labeling channel (**Figure 8D**). It happens because the MCTA analysis is done at the event level, and the statistics give populational results, as any flow cytometry analysis. To find the predominant labeling(s) that generated the different resultant colors (subpopulations), the Resultant Plot With Filter is once more the solution. In this case, the user determines a range of degrees that selects only the events shown in a given color, either orange, green, or violet. Therefore, the statistics will show the predominant labeling(s) for each subpopulation of cells. As shown in **Figure 7**, the statistics will be restricted to each range of degrees for each subpopulation, and the user will be able to identify what molecule or molecules mostly contributed to each resultant color.

DISCUSSION

During the past few decades, we witnessed the development of new software and tools for the analysis of increasingly complex results obtained by flow cytometry. However, most users still follow the same analysis strategies, typically based on the definition of sequences of manual gates on multiple dot plots that show two parameters at a time. Although cytometrists are familiar with this fragmented process, it has many disadvantages that affect data reproducibility and accuracy.



C

Number of total events	292,292.00
Number of events analyzed by the MCTA	864

Channel: 1	Geometric Mean: 2,068.46	Geo Std dev: 4.93	Median: 2,381.29
Channel: 3	Geometric Mean: 2,579.88	Geo Std dev: 4.05	Median: 3,677.94
Channel:12	Geometric Mean: 683.19	Geo Std dev: 4.09	Median: 851.58

FIGURE 5 | The MCTA application and resultant dot plots: the MCTA analysis was done in Th1 **(A)** and Th2 **(B)** $CD4^+$ T lymphocytes obtained from healthy human donors and stimulated *in vitro* with Staphylococcal enterotoxin B (SEB). The MCTA gating strategy considered the events positive for CD3, CD4, CCR7, and CD154 labeling, and colored events correspond to the calculation of a resultant vector (or color) based on the labeling of IFN-γ, IL-2, and TNF-α **(A)** or IL-4 **(B)**. The MCTA statistical analysis is shown **(C)** for the file represented in **(A)**. These FCS files were obtained from the repository (https://flowrepository.org/public_experiment_representations) ID FR-FCM-ZZEC.

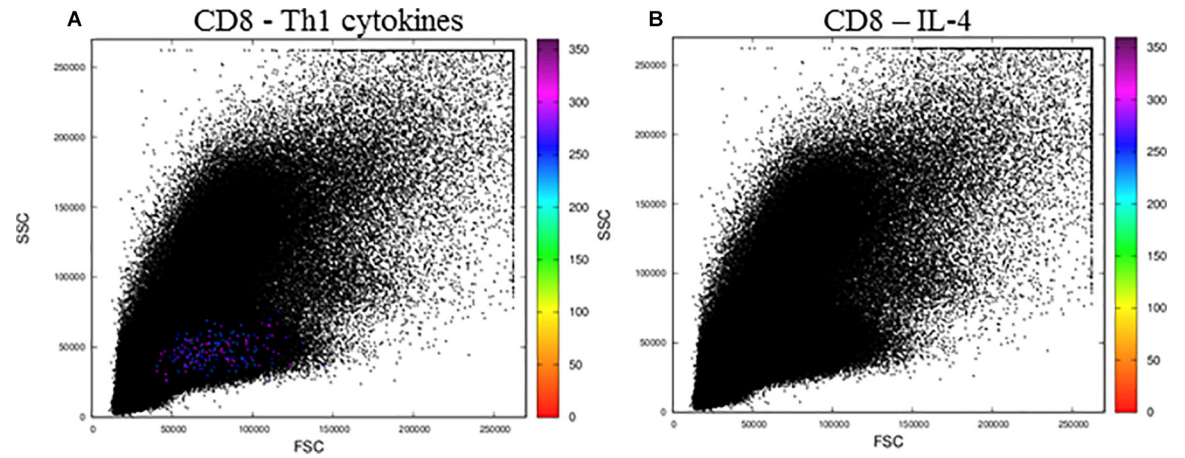


FIGURE 6 | The MCTA application for the analysis of $CD8^+$ T lymphocytes stimulated by SEB. The MCTA analysis was done in Th1 **(A)** or IL-4^+ **(B)** $CD8^+$ T lymphocytes obtained from healthy human donors and stimulated *in vitro* with Staphylococcal enterotoxin B (SEB). Colored events correspond to the calculation of a resultant vector (or color) based on the labeling of IFN-γ, IL-2, and TNF-α **(A)** or IL-4 **(B)** only in the events that are positive for CD3, CD8, CCR7, and CD154. These FCS files were obtained from the repository (https://flowrepository.org/public_experiment_representations) ID FR-FCM-ZZEC.

Here we propose an automated strategy to explore the diversity of cells in flow cytometry data. To the best of our knowledge, the MCTA method is the first algorithm that analyzes multiple labelings simultaneously and maps the results in FSC-A x SSC-A morphology gates, extending the conventional analysis. The process is accessible for average users and can quickly show the resultant color based on multiple fluorescence labelings per sample. However, at this point, we present the rationale of the MCTA method, which is intended to be included in software or packages for flow cytometry analysis in user-friendly interfaces. It is our primary goal to offer users a different perspective on the complexity of their results.

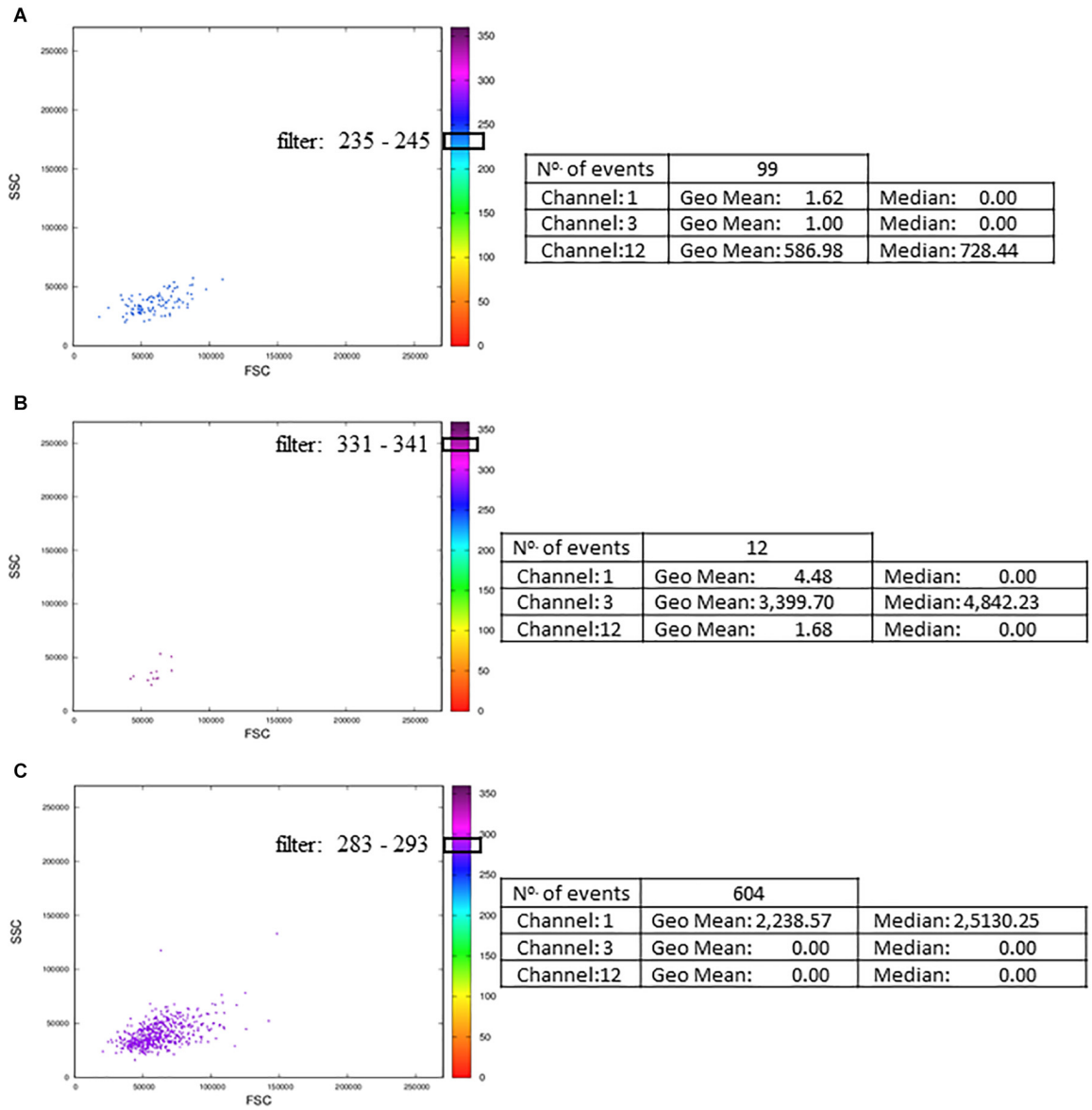
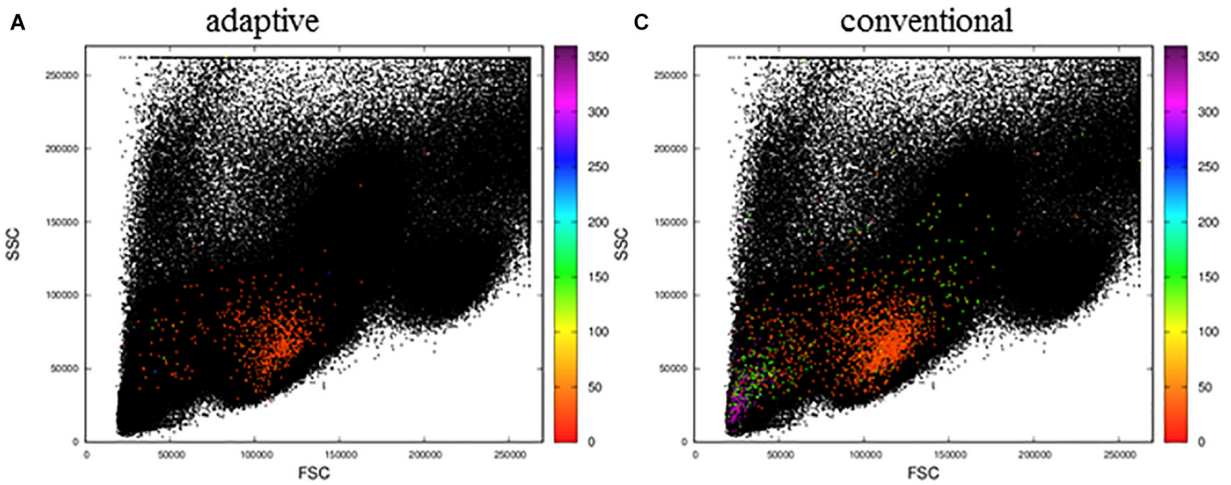


FIGURE 7 | Resultant Plot With Filter: for the Resultant Plot With Filter, the user must define a range of degrees to apply to the MCTA analysis, and these ranges are graphically represented here as block boxes over the degree scale. The analysis was done in Th1 $CD4^+$ T lymphocyte obtained from healthy human donors and stimulated *in vitro* with Staphylococcal enterotoxin B (SEB). The gating strategy considered the events positive for CD3, CD4, CCR7, and CD154 labeling. The resultant vector was calculated based on the labeling of IFN-γ, IL-2, and TNF-α, and the statistical analysis of each range is shown. The examples considered only events with resultant colors in the range of 235 to 245° (**A**, the color range of IL-2), 331 to 341° (**B**, the color range of IFN-γ), and 283 to 293° (**C**, the color range of TNF-α). These FCS files were obtained from the repository (https://flowrepository.org/public_experiment_representations) ID FR-FCM-ZZEC.

This method allows the observation of biological phenomena that could not be identified if any other tool were used. In this way, the application of the MCTA method can guide subsequent analysis choices, allowing as much information as possible to be extracted from biological samples. Moreover, the results are reproducible among collaborators, as long as using the same files and applying the same background and compensation values.

Here, we used FCS files downloaded from a public repository and challenged the method using two sets of data. These experiments employed 13 and 15 fluorescence parameters that were used for the gating strategy and the calculation of the resultant color. In these examples, the resultant color was obtained based on single labeling (IL-4), three labeling colors, or nine colors. Even if the number of parameters used to generate the resultant vector exceeds nine colors, this does not impose a limitation on the MCTA application. Because of the statistical analysis, the users will always identify what marker or markers mainly influenced the resultant vector generated for data interpretation. The Resultant Plot With Filter is another tool



B Adaptive

Channel: 1	NA	Geometric Mean: NA	Geo Std dev: NA	Median: 1.50E-09
Channel: 2	CD2	Geometric Mean: 180.29	Geo Std dev: 3.09	Median: 235.29
Channel: 3	NKp30	Geometric Mean: 370.80	Geo Std dev: 2.96	Median: 443.45
Channel: 6	CD7	Geometric Mean: 6,221.93	Geo Std dev: 2.08	Median: 7,000.59
Channel: 7	ILT2	Geometric Mean: 133.46	Geo Std dev: 3.31	Median: 156.13
Channel: 9	Siglec-7	Geometric Mean: 275.46	Geo Std dev: 3.54	Median: 337.40
Channel:11	CD56	Geometric Mean: 343.67	Geo Std dev: 3.45	Median: 428.36
Channel:13	NKG2A	Geometric Mean: 378.50	Geo Std dev: 3.83	Median: 531.06

D Conventional

Channel: 1	NA	Geometric Mean:	Geo Std dev:	Median: 0.00
Channel: 2	CD2	Geometric Mean: 186.71	Geo Std dev: 3.36	Median: 235.29
Channel: 3	NKp30	Geometric Mean: 363.56	Geo Std dev: 3.03	Median: 438.93
Channel: 6	CD7	Geometric Mean: 5,303.07	Geo Std dev: 3.17	Median: 6,653.65
Channel: 7	ILT2	Geometric Mean: 125.64	Geo Std dev: 5.69	Median: 118.85
Channel: 9	Siglec-7	Geometric Mean: 278.00	Geo Std dev: 3.64	Median: 358.57
Channel:10	NKG2C	Geometric Mean: 197.72	Geo Std dev: 3.67	Median: 268.58
Channel:11	CD56	Geometric Mean: 412.40	Geo Std dev: 3.48	Median: 567.17

FIGURE 8 | Adaptive and conventional NK cells analysis using the MCTA method. The expression of NKG2C was the inclusion parameter for the analysis of adaptive NK cells **(A)**, and the expression of NKG2A **(C)** was used for the study of conventional NK cells. Statistical analysis for adaptive NK cells is shown in **(B)** and for conventional cells is shown in **(D)**. These FCS files were obtained from the repository (https://flowrepository.org/public_experiment_representations) repository ID FR-FCM-ZYY6.

developed to help the users in the interpretation of the results, which is never exclusively based on the visual identification of the color of the events.

One essential aspect of the MCTA method is its determinism. Several flow cytometry methods for data analysis use stochastic algorithms for dimensionality reduction, which may require great processing capacity. To address this problem, it is common to reduce the number of events analyzed to keep running times acceptable (Saeys et al., 2016). Moreover, stochastic methods like t-SNE or UMAP produce different results for each run on the same dataset; therefore, the users should run the algorithm multiple times using the same data to analyze variability and prevailing trends. Furthermore, stochastic methods are typically able to process only a few tens of thousands of events per run, even when implementing additional techniques such as downsampling, hierarchical clustering, or dimensionality reduction. Another obstacle in the use of more complex dimensionality reduction techniques is the setting of the parameters to run the algorithms. Frequently, this inherent complexity leads the users to employ default settings to run their analysis due to the lack of knowledge to change these parameters. In the case of t-SNE for example, the user has to choose values for the perplexity parameter, whose typical values vary between 5 and 50. As a stochastic method, even when using the same perplexity parameter value, the result will variate when comparing different runs. Moreover, it was recently published

that t-SNE can erroneously indicate clusters for homogeneously distributed data, suggesting the wrong number of subgroups or projecting data points that belong to the same subset, as if they belong to different subgroups (Lötsch and Ultsch, 2019).

A deterministic method, like the MCTA algorithm, requires only one run on a dataset and does not demand additional aggregation steps. Furthermore, the complexity of the algorithm described in this work is O(N.*nc*), where N is the number of events, and *nc* is the number of channels. Typically *nc* is fixed (for a specific flow cytometer) and $nc << N$; therefore, the complexity can be approximated by O(N).

Our method can also be easily adapted to explore multiple cores/nodes in parallel, taking full advantage of modern multi-core processors in highly scalable implementations. Indeed, the method proposed here is very efficient computationally, considering both execution time and memory requirements. When we compared the processing times of MCTA vs. t-SNE to generate **Figure 8** using all events, we observed that MCTA spent up to half a minute processing the whole data set. In contrast, the t-SNE analysis required more than 6 h using an average computer (Mac Book pro, 16gb RAM, 2.3 GHz Intel Core i5) and up to 2.5 h in a computational cluster. All these characteristics make the MCTA an ideal method to rapidly evaluate specific questions about cellular phenotype or function. Using the same computer mentioned, it was impossible to run a UMAP dimension reduction technique due to computational restrictions. To date, when we performed a downsample of 15,000 events, we noted a dramatic reduction in the events gated in the MCTA analysis, showing the importance of analyzing the whole set of data. Although downsampling strategies are frequently applied to the data before using dimensional reduction techniques, our observation suggests that data might be lost during downsampling to run t-SNE or UMAP algorithms. Therefore, the identification of rare populations can be impossible.

Although the MCTA method is based on the calculation of a resultant vector according to multiple labelings, events with extreme artifactual fluorescent signals (very high or very low MFI), will not alter the tendency result of the population under analysis. This is true because the MCTA method calculates a resultant color for each event, and extreme artefactual signals will affect only the event itself, not the population under analysis. Moreover, the MCTA method is not indicated to substitute conventional analysis; it is proposed as a new way to show complex phenotypic profiles, complementing and extending conventional analysis. We believe that conventional flow cytometry data analysis, combined with other appropriate computational tools and methods, will help to identify and better describe biological phenomena, leading to more accurate, complete, and reproducible data.

We consider that the MCTA analysis offers results that are not directly comparable with traditional dimensionality reduction techniques as t-SNE and UMAP. In the MCTA method, the analysis is oriented to a specific set of data through the inclusion and exclusion of markers in the gating strategy. On the other hand, t-SNE and UMAP generally use the whole set of fluorochromes available in the data set.

Finally, we believe that the MCTA method can be integrated as a new functionality into flow cytometry analysis software, allowing complementary views and comparisons with well-established methods like t-SNE and UMAP and conventional fragmented analysis.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://flowrepository.org/public_experiment_representations, ID: FR-FCM-ZZEC and https://flowrepository.org/public_experiment_representations repository, ID: FR-FCM-ZYY6.

ETHICS STATEMENT

The animal study was reviewed and approved by A Comissão de Ética no Uso de Animais do Instituto Oswaldo Cruz (CEUA-IOC), certificate number: L006-2015.

AUTHOR CONTRIBUTIONS

CB proposed the HSL representation for the analysis of multiple labelings and wrote the initial version of the code. FS wrote the codes for the generation of dot plots and all statistical representation, made the functionality tests, and assisted in the writing of the text. AH-P did all biological tests and interpretation of MCTA results, wrote the article, and guided the MCTA method implementation according to biological parameters. JS-A reviewed the MCTA code and compared it with other computational tools, and helped writing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (grant numbers 421803/2017-7, 407349/2015-4, and 407711/2012-0) and institutional funding from Instituto Oswaldo Cruz.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Marcelo Meuser Batista and Natalia Vacani Martins for the flow cytometry experiments included in the repository and Dr. Mariana Gandini for testing the MCTA.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.526814/full#supplementary-material

REFERENCES

Bruggner, R. V., Bodenmiller, B., Dill, D. L., Tibshirani, R. J., and Nolan, G. P. (2014). Automated identification of stratifying signatures in cellular subpopulations. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2770–E2777.

Cascabulho, C. M., Beghini, D. G., Meuser-Batista, M., Penido, C., and Henriques-Pons, A. (2016). Chemotaxis and immunoregulatory function of cardiac γδ T cells in dystrophin-deficient mice. *J. Immunol.* 197, 3531–3544. doi: 10.4049/jimmunol.1600335

de Oliveira, G. M., Diniz, R. L., Batista, W. M., Batista, M., Bani Correa, C., de Araújo-Jorge, T. C., et al. (2007). Fas ligand-dependent inflammatory regulation in acute myocarditis induced by *Trypanosoma cruzi* infection. *Am. J. Pathol.* 171, 79–86. doi: 10.2353/ajpath.2007.060643

Hu, G., Pan, Z., Zhang, M., Chen, D., Yang, W., and Chen, J. (2014). An interactive method for generating harmonious color schemes. *Color Res. Appl.* 39, 70–78. doi: 10.1002/col.21762

Levine, J. H., Simonds, E. F., Bendall, S. C., Davis, K. L., Amir, E. A., Tadmor, M. D., et al. (2015). Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* 162, 184–197. doi: 10.1016/j.cell.2015.05.047

Lötsch, J., and Ultsch, A. (2019). Current projection methods-induced biases at subgroup detection for machine-learning based data-analysis of biomedical data. *Int. J. Mol. Sci.* 21:79. doi: 10.3390/ijms21010079

Mair, F., Hartmann, F. J., Mrdjen, D., Tosevski, V., Krieg, C., and Becher, B. (2016). The end of gating? An introduction to automated analysis of high dimensional cytometry data. *Eur. J. Immunol.* 46, 34–43. doi: 10.1002/eji.201545774

Montante, S. M., and Brinkman, R. R. (2019). Flow cytometry data analysis: recent tools and algorithms. *Int. J. Lab. Hematol.* 41(Suppl. 1), 56–62. doi: 10.1111/ijlh.13016

Pedreira, C. E., Costa, E. S. D., Lecrevise, Q., Grigore, G., Fluxa, R., Verde, J., et al. (2019). From big flow cytometry datasets to smart diagnostic strategies: the EuroFlow approach. *J. Immunol. Methods* 475:112631. doi: 10.1016/j.jim.2019.07.003

Saeys, Y., Van Gassen, S., and Lambrecht, B. N. (2016). Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nat. Rev. Immunol.* 16, 449–462. doi: 10.1038/nri.2016.56

Shapiro, H. M. (2004). The evolution of cytometers. *Cytomet. A* 58, 13–20. doi: 10.1002/cyto.a.10111

Toghi Eshghi, S., Au-Yeung, A., Takahashi, C., Bolen, C. R., Nyachienga, M. N., Lear, S. P., et al. (2019). Quantitative comparison of conventional and t-SNE-guided gating analyses. *Front. Immunol.* 10:1194. doi: 10.3389/fimmu.2019.01194

Tsai, S.-H., and Tseng, Y.-H. (2012). A novel color detection method based on HSL color space for robotic soccer competition. *Comput. Math. Appl.* 64, 1291–1300. doi: 10.1016/j.camwa.2012.03.073

Van Gassen, S., Callebaut, B., Van Helden, M. J., Lambrecht, B. N., Demeester, P., Dhaene, T., et al. (2015). FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data. *Cytomet. A* 87, 636–645. doi: 10.1002/cyto.a.22625

Weber, L. M., and Robinson, M. D. (2016). Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. *Cytomet. A* 89, 1084–1096. doi: 10.1002/cyto.a.23030

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 20 October 2020
doi: 10.3389/fonc.2020.580698

Deciphering Cancer Cell Behavior From Motility and Shape Features: Peer Prediction and Dynamic Selection to Support Cancer Diagnosis and Therapy

Michele D'Orazio[1], Francesca Corsi[2,3], Arianna Mencattini[1], Davide Di Giuseppe[1], Maria Colomba Comes[1], Paola Casti[1], Joanna Filippi[1], Corrado Di Natale[1], Lina Ghibelli[3] and Eugenio Martinelli[1]**

[1] Department of Electronic Engineering, University of Rome "Tor Vergata," Rome, Italy, [2] Department of Chemical Science and Technologies, University of Rome "Tor Vergata," Rome, Italy, [3] Department of Biology, University of Rome "Tor Vergata," Rome, Italy

OPEN ACCESS

Edited by:
Francesco Rundo,
STMicroelectronics (Italy), Italy

Reviewed by:
Umberto Malapelle,
University of Naples Federico II, Italy
Francesca Trenta,
University of Catania, Italy

***Correspondence:**
Francesca Corsi
francesca.corsi@uniroma2.it
Eugenio Martinelli
martinelli@ing.uniroma2.it

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 06 July 2020
Accepted: 08 September 2020
Published: 20 October 2020

Citation:
D'Orazio M, Corsi F, Mencattini A, Di Giuseppe D, Colomba Comes M, Casti P, Filippi J, Di Natale C, Ghibelli L and Martinelli E (2020) Deciphering Cancer Cell Behavior From Motility and Shape Features: Peer Prediction and Dynamic Selection to Support Cancer Diagnosis and Therapy. Front. Oncol. 10:580698. doi: 10.3389/fonc.2020.580698

Cell motility varies according to intrinsic features and microenvironmental stimuli, being a signature of underlying biological phenomena. The heterogeneity in cell response, due to multilevel cell diversity especially relevant in cancer, poses a challenge in identifying the biological scenario from cell trajectories. We propose here a novel peer prediction strategy among cell trajectories, deciphering cell state (tumor vs. nontumor), tumor stage, and response to the anticancer drug etoposide, based on morphology and motility features, solving the strong heterogeneity of individual cell properties. The proposed approach first barcodes cell trajectories, then automatically selects the good ones for optimal model construction (good teacher and test sample selection), and finally extracts a collective response from the heterogeneous populations *via* cooperative learning approaches, discriminating with high accuracy prostate noncancer vs. cancer cells of high vs. low malignancy. Comparison with standard classification methods validates our approach, which therefore represents a promising tool for addressing clinically relevant issues in cancer diagnosis and therapy, e.g., detection of potentially metastatic cells and anticancer drug screening.

Keywords: machine learning, cell motility, peer prediction, dynamic feature selection, cancer heterogeneity, metastatic cancer cell detection, drug screening

INTRODUCTION

The ability of cells to coordinately move is indispensable in many biological processes, such as tissue morphogenesis and repair, cancer progression, and invasion (i.e., metastasis spreading) (1). Cell movements vary according to intrinsic features and microenvironmental conditions, possibly being a signature of underlying biological phenomena. A straightforward simplification is that, for instance, healthy cells move differently from tumor cells, especially when they undergo the epigenetic changes leading to epithelial-to-mesenchymal transition, a phenomenon that provides new motility ability to cancer cells allowing metastatic spreading (2). Motility is hardly described by mere molecular markers, and therefore this important issue requires different approaches to be properly addressed.

Classifying cells according to their behavior in terms of coordinated motility needs facing the problem of cell heterogeneity; cells apparently identical by morphological criteria may behave differently because of fundamental differences in genetic or epigenetic asset, the stage of cell cycle or differentiation, in cell–cell or cell–environment interaction, etc., parameters that, although assessable by single molecular labeling, continuously change in time and combinations, being thus impossible to describe in classical molecular terms. Heterogeneity in cell response thus represents a big limitation to identify the underlying biological scenario from cell motility; nevertheless, such heterogeneity allows extracting behavioral rules to finalize the automatic understanding, for example, of cell state (e.g., tumor vs. nontumor), tumor stage (e.g., metastatic vs. nonmetastatic), response to anticancer drugs, etc. To this purpose, label-free (3) fluorescence time-lapse microscopy (TLM) and special purpose video data analysis tools (4–7) are providing promising novel, nonmolecular, dynamic approaches.

We present here a novel methodology to conduct massive analysis of cell motility in different *in vitro*–controlled conditions that combines TLM and label-free imaging, with cell tracking, quantitative representation of trajectories, and novel machine learning (ML) strategies within peer prediction framework. Peer prediction strategies acquired much interest in many contexts such as assignments in massive open online courses and in collecting feedback about a new service (8). Such algorithms use reports from multiple participants to score their contributions in settings in which there is no way to verify the quality of response (9). Cell systems, where a unique correct response for cell behavior is not expected, represent therefore an unconventional and challenging environment for peer prediction paradigm extension.

Optimization of ML strategies and adaptation to cell motility investigation need the identification of the correct learning examples. Differently from other social contexts (10, 11), none of the cells and related trajectories can be judged by experts, both because it cannot be practically done and because the heterogeneity of cell behavior and the massive number of cells make it impossible to extract the "truth" at sight. Because the acquired samples (cells) are not labeled by experts, cell trajectories would directly inherit the same label assigned to the entire experiment, i.e., cells moving in a control experiment would be assumed to behave in a unique, similar way. However, this assumption is generally invalid. The intrinsic data heterogeneity forbids the direct assignment of a unique label to all the cells, impeding to represent a cell population as a unique behavioral entity. Hence, the selection of samples for model construction becomes the core of the ML problem.

In the present work, we address the problem of learning a classification model from cell trajectories and related descriptors (peers) using a novel strategy. First, inspired by a previous approach (12), all cell trajectories are "barcoded" during model construction; however, only some of the barcoded trajectories are assigned the role of trainers (hereafter denoted as "the good teachers") because only a certain number of cell trajectories can be used to construct the good model. Second, not all cell trajectories in the test set are used for testing because not all of them represent the global target (e.g., a unique behavior for the same cell line or the same reaction to a given stimulus). The presence of a collective response phenomenon forces the approach to automatically identify peers in the test set, with high agreement in terms of the same descriptors selected in the training step.

Regarding the descriptors selection, only some features extracted from each cell trajectory can be assigned a "discriminatory role" because not all features are likely to be simultaneously relevant for all groups of cells. As an example, in a group of cells moving toward a target cell, e.g., immune–cancer cross-talk (13, 14), speed and directional persistence are needed to model their collective motion; on the other hand, in a group of cells interacting with a target cell, e.g., immune cells killing a cancer cell (15, 16), mean interaction time and track curvature have proved to be specifically tailored for the phenomenon quantification. In particular, in this work, we extended and applied a dynamic feature selection (DFS) procedure (17, 18), selecting, in an unsupervised way, the optimal feature set extracted from the training set for each new test sample; this will be used to build a classifier for the test label prediction.

Of importance, in addition to the model construction, in our approach the novelty includes the decision-making step. In *in vitro* experiments, cells naturally cluster before reaching the confluence; consensus strategies can be exploited to acquire a unique decision for the cluster. In this regard, we applied two distinct cooperative learning criteria, inspired by collective phenomena and peer influence studies (11); on the one hand, we applied a majority voting procedure to all the labels assigned by the classifier to the cell trajectories selected for that cluster; on the other, we summed up all the scores assigned to each category of the cells belonging to the same cluster and assigned the class with the largest total score to the cluster. We refer to the two criteria as majority voting criterion (maj-vot) and maximum trustiness criterion (max-trust).

MATERIALS AND METHODS

Video Acquisition Details

The videos were acquired with a custom small-scale inverted microscope (19). In order to have control on acquisition methods and light exposure, a custom firmware was developed in MATLAB 2017a®. We acquired images at one frame per minute with 6 h of total experimental time (12 h in the LNCaP case). The images have a field of view of 1.2-mm width by 1.0-mm height and a theoretical spatial resolution of 0.33 μm/px.

We recorded two videos per treatment condition in RWPE-1 and PC-3 prostate cell experiments and four videos for the control case in the LNCaP cells.

Cell Culture Details

Human prostate cancer cells, PC-3 and LNCaP cell lines (ATCC, Rockville, MD), were grown in RPMI 1640 medium, supplemented with 10% fetal bovine serum, 1% L-glutamine (2 mg/mL), and 1% penicillin/streptomycin (100 IU/mL) (Euroclone).

Nonneoplastic, immortalized human prostatic epithelial cells, RWPE-1 (ATCC, Rockville, MD) were grown in keratinocyte serum-free medium (K-SFM), supplemented with 1% penicillin/streptomycin (100 IU/mL), 50 μg/mL bovine pituitary extract, and 5 ng/mL epidermal growth factor (Life Technologies, Barcelona, Spain).

Cells were grown at 37°C in a humidified atmosphere of 5% CO_2 in air. In each experiment 40,000 cells/mL were seeded in 35-mm Petri dishes (Jetbiofil). Seventy-two hours postseeding, cells were treated with the chemotherapeutic drug etoposide (Sigma-Aldrich), a topoisomerase II inhibitor, at the final concentrations of 0.5, 1, or 5 μM and immediately analyzed with TLM.

Method for Automatic Cell Behavior Classification

Step 1. Cell Localization and Tracking

The method is focused on the use of a previously validated cell tracking tool, Cell-Hunter, which has been tested in prostate cancer cell automatic tracking (12, 19), immune–cancer cell crosstalk studies (16), and recently in red blood cell plasticity analysis (20). The software automatically locates cells with a radius within a given range provided by the user and tracks them providing a predetermined maximum displacement allowed.

Step 2. Automatic Cell Clustering Identification

Cells naturally cluster when they are put in *in vitro* culture, a primitive status before moving toward confluence. Cells move according to the cluster they belong, promoting different roles according to the cell stage, age, drug absorption, etc. The automatic identification of the clusters each cell belongs to is performed through image analysis algorithms involving image binarization and morphological operators (12). The technique is based on the localization of individual cells by performing the segmentation of circular objects using the Circular Hough Transform (CHT) (21) set according to the mean estimated radius of cells involved. Each detected cell is represented as a white circular object. By using an accumulation criterion, consisting of the overlapping of the cell nuclei detected along all the frames and normalizing by the maximum value, a gray-scale map is obtained, in which higher intensity values locate cells with limited motility frame by frame and thus higher probability to stay in that position during movement. By applying pixel intensity thresholding using the Otsu criterion (21) and then morphological operators refining (21), a rough binary (black and white) image representation of each cluster is obtained. The boundary extraction of the detected regions represents cluster contours.

Step 3. Feature Extraction

Each cell is characterized in terms of its kinematics and shape dynamics. To do this, we identified some quantitative descriptors to characterize the dynamics of cell movement. In addition, shape descriptors are also considered to characterize the morphodynamics during movement. Further mathematical details of the two sets of descriptors can be found in the following subparagraph.

Cell morphology feature extraction

The shape extraction process is described in **Supplementary Figure 1**. We used the position of the cell trajectory to correctly focus the window containing the cell under study for every frame (**Supplementary Figure 1A**). We obtained an initial contour applying a CHT (22, 23) with a high sensitivity and a maximum radius smaller than the radius expected from the first object found (**Supplementary Figure 1B**). We took the perimeter of the smallest convex polygon (convex hull) containing the union of all the found circles (**Supplementary Figure 1C**) as starting contour for an active contours algorithm (24) that gave us the final result (**Supplementary Figure 1D**).

Looking at time-lapse videos, we observed that cells in their motion change eccentricity, perimeter, and area. They also change solidity when making pseudopodia. Furthermore, nonneoplastic cells (RWPE-1) are smaller than the others, and the milder neoplastic cells (LNCaP) have a higher eccentricity on average. These considerations led us to consider as significant features eccentricity, area, perimeter, and solidity (25).

(a) Eccentricity is defined as

$$\text{eccentricity} = \frac{d_f}{D_M} \tag{1}$$

where d_f is the distance between the foci, and D_M is the major axis length;

(b) Area is defined as

$$\text{area} = \sum_i \sum_j f\left(i,j\right) \tag{2}$$

where $f\left(i,j\right)$ is 1 for $\left(i,j\right)$ in the region of interest and zero elsewhere;

(c) Perimeter

$$\text{perimeter} = \sum_i \sum_j g\left(i,j\right) \tag{3}$$

where $g\left(i,j\right)$ is 1 over the pixels that have at least one neighbor (in 8-connection) with zero value and zero elsewhere;

(d) Solidity (26) is defined as

$$\text{solidity} = \frac{\text{area}}{\text{area convex hull}} \tag{4}$$

where the convex hull is the smallest convex polygon that contain the region.

To exploit the dynamic of these descriptors for each cell over time, we performed the following statistics: mean, standard deviation, skewness, kurtosis, Shannon entropy, and signal entropy.

Cell motility feature extraction

In order to have statistical significance of the extracted features, we discarded all the trajectories, which lasted <50 min (50 time points). Cell position at each time point is affected by errors:

discretization error, which is linked to the dimension of each pixel (0.66 μm/px) and the optical resolution ($R \cong 0.8\ \mu m$). Another source of error occurs when the algorithm does not find the cell, assigning the previous position to the cell, thus resulting in jumps in the trajectories. We reduced this error with a smoothing spline approximation (27). On the new set of coordinates, $(x_s(t_k), y_s(t_k))$, we computed the following parameters for their already proven informative content (19):
(i) Tangential speed norm

$$v(t_k) = \sqrt{\left(\frac{x_s(t_{k+1}) - x_s(t_k)}{(t_{k+1} - t_k)}\right)^2 + \left(\frac{y_s(t_{k+1}) - y_s(t_k)}{(t_{k+1} - t_k)}\right)^2} \quad (5)$$

(ii) Track curvature $\chi(t_k)$

$$\chi(t_k) = \frac{|x_s' y_s'' - y_s' x_s''|}{\left[(x_s')^2 + (y_s')^2\right]^{\frac{3}{2}}} \quad (6)$$

(iii) Turning angle $\vartheta(t_k)$

$$\vartheta(t_k) = \tan^{-1}\left(\frac{v_x}{v_y}\right) \quad (7)$$

(iv) Angular velocity, computed as the ratio between the magnitude of the velocity and the distance from the center of the trajectory.

$$\omega(t_k) = \frac{v(t_k)}{R(t_k)} \quad (8)$$

where $R(t_k) = \sqrt[2]{(x_s(t_k) - x_c)^2 + (y_s(t_k) - y_c)^2}$, $x_c = \frac{1}{N}\sum_{k=1}^{N} x_s(t_k)$ and $y_c = \frac{1}{N}\sum_{k=1}^{N} y_s(t_k)$.
(v) Diffusion coefficient

$$D = 4^{-1} \cdot e^{y_0} \quad (9)$$

where y_0 is the y-axis intercept estimated form a linear fit in log space of the mean square displacement (28).
(vi) Directional persistence, defined as the ratio between the initial and the final point and the real length of the track.

$$p = \frac{\left\|\underline{x}_s(t_f) - \underline{x}_s(t_i)\right\|}{L} \quad (10)$$

where $L = \sum_k \left\|\underline{x}_s(t_k) - \underline{x}_s(t_{k-1})\right\|$.

From each time-varying feature, we extracted the following high-level statistical descriptors: mean, standard deviation, skewness, kurtosis, and signal Shannon entropy. In conclusion, we collected 24 shape descriptors and 39 motility features.

Kinematics and shape features allow excluding some trajectories from the whole analysis through unsupervised outlier detection. Such step is required because of some false tracks extracted by the cell tracking software. Misdetected trajectories may be related to false cells localization (for example, out-of-focus cells) or to tracks that exit the field of view and are linked to new cells entering the scene.

It is straightforward to note that optimal descriptors depend not only on the task, but also on the training and testing samples. For this reason, we selected a wide set of descriptors commonly used for evaluating cell behavior from motility and shape analysis. The assumption underlying the selection needs to be able to monitor different aspects of cell motility, such as speed, curvature, turning angle, persistence, etc., as well as synthetic descriptors of shape dynamics along time.

Step 4. Good Teacher Selection

Let us consider a set of training samples, $\boldsymbol{F} = \begin{vmatrix} \boldsymbol{f}_1 \\ \vdots \\ \boldsymbol{f}_T \end{vmatrix}$, with T as the number of training samples, and $\boldsymbol{f^j_k} = \left\{f^j_{k1}, f^j_{k2}, \ldots, f^j_{kM}\right\}$ the subarray of descriptors for the kth cell in the jth cluster $\boldsymbol{C_j}$, $j = 1, \ldots N_C$ with N_C being the number of clusters in the training set.

First, the algorithm automatically selects a subset of descriptors, $\tilde{\boldsymbol{F}} \subset \boldsymbol{F}$, with T rows and $M' < M$ columns (descriptors) such that a maximum value is obtained for a given criterion Ψ_1 applied to the set $\tilde{\boldsymbol{F}}$. The suboptimal criterion Ψ_1 used here is the maximum area under the curve (AUC) values (29) obtained in all the associated binary problems in a multiclass context (in an all-vs.-all classification strategy validated on the training set). The AUC is a metric of separability for a given descriptor with respect to the output label of different classes. The higher the AUC value (bounded in [0,1]), the higher the discrimination capability of the descriptor.

Then, a subset of training samples, namely, $\boldsymbol{F}' \subset \tilde{\boldsymbol{F}}$ with $T' < T$ rows and M' columns, is extracted by taking the training samples whose descriptors fall within a tuned range (i.e., percentile $[th_1, th_2]$) independently calculated in each video. Formally, $[th_1, th_2]$ allows keeping all the observations whose cumulative distribution function is between th_1 and th_2.

The threshold values set in each experiment are listed in **Table 1**, rows 1 and 2.

Step 5. Test Sample Selection

By using the same descriptors selected in Step 4, a similar refining procedure is applied to the test cell trajectories, by using an independent range of elimination, namely $[th_3, th_4]$, leading to test samples indicated with $\boldsymbol{H}$. **Table 1**, rows 3 and 4, lists the values for percentiles th_3 and th_4. Good teacher and test sample selection procedures represent the forerunner of the peer prediction paradigm.

Step 6. Dynamic Feature Selection

After training and test data have been collected, namely, $\boldsymbol{G}$ and $\boldsymbol{H}$, descriptors are finely selected by using a DFS procedure. DFS applies three distinct criteria. The first supervised criterion, Fisher criterion in **Figure 1**, selects features that correlate with the output in training set, according to a limit value $th_{1\text{dfs}}$. The second and third criteria are unsupervised and use two distinct approaches. In the second, the Mahalanobis criterion (**Figure 1**), features in the test set whose Mahalanobis distance from features in the training set is under a given limit threshold $th_{2\text{dfs}}$ are selected. In the third criterion (**Figure 1**), the maximum posterior

TABLE 1 | List of algorithm parameters setting used in the experiments for performance assessment.

Algorithm parameters setting			
Symbol	**Description**	**Shape features**	**Motility features**
th_1	Lower bound percentile good teacher selection	0.2	0.2
th_2	Upper bound percentile good teacher selection	0.8	0.8
th_3	Lower bound percentile test sample selection	0.1	0.1
th_4	Upper bound percentile test sample selection	0.9	0.9
th_{1dfs}	DFS Fisher criterion threshold	1.0	0.9
th_{2dfs}	DFS minimum Mahalanobis criterion threshold	0.1	0.1
th_{3dfs}	DFS maximum probability criterion threshold	1.0	0.2
p	Confidence value for stepwise feature selection	0.3	0.3

A brief description of each parameter is also included.

probability of feature values in test to belong to the distribution of values in training set over all the classes is calculated; features with probability values higher than a given threshold $th_{3\text{dfs}}$ are kept. Further mathematical details can be found in Mosciano et al. (18). Features that satisfied all the three criteria are then selected. The extension we propose here with respect to the standard DFS (18, 30) is the inclusion of a preliminary supervised selection performed at the beginning of the model construction based on stepwise feature selection procedure (31) applied on the training samples. The fact that motion models may vary within the same experiment (12) implies the necessity to extract many kinematics descriptors. The modification to standard DFS allows limiting the initial set of descriptors to a maximum effective set. The p-value of the F test (32) used for the acceptation of a feature in the selection process, indicated with p, is an algorithm parameter. Values of $th_{1\text{dfs}}$, $th_{2\text{dfs}}$, $th_{3\text{dfs}}$, and p are listed in **Table 1**, rows 5–8.

Finally, we may indicate with $\overline{\boldsymbol{G}}$ and $\overline{\boldsymbol{H}}$ the refined sets for model construction and automatic classification.

Step 7. Classification Model

Model construction is performed considering three distinct classification models: linear discriminant analysis (LDA) (33), support vector machine (SVM) (34), and K-nearest neighbor (KNN) (35).

LDA finds a linear combination of features (input data) to separate two or more classes of objects or events. In this work, LDA naturally produces as an outcome not only the class label but also an associated posterior probability to belong to the class. According to this, given a test set $\overline{\boldsymbol{H}}$, the LDA model provides for each class $c_{\overline{\boldsymbol{H}}}$ a score value $y_{\overline{\boldsymbol{H}}}$. Such values are used in the cooperative strategies as shown below.

SVM presents one of the most robust prediction methods, based on the statistical learning framework. An SVM model is a representation of the examples as points in new prediction space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall. The SVM algorithm may be turned into nonlinear classification model by using a nonlinear kernel, commonly radial basis function. In this work, we used SVM with linear kernel for harmonization with the LDA competitive method.

KNN is a nonparametric method in which the input consists of the K-closest training examples ($K = 5$ in this work) in the feature space (input data), whereas the output is a class membership. An object is assigned to the class most common among that of its KNN training samples. A standard metric for representing neighborhood is the Euclidean distance, which is used in our work.

Step 8. Cooperative Learning

In the test set, all the cell trajectories associated to a cluster are individually scored through $y_{\overline{\boldsymbol{H}}}$ and labeled through $c_{\overline{\boldsymbol{H}}}$. Under the need to provide a unique decision, i.e., a unique proof of concept to the underlying biological hypothesis, the approach allows aggregating the labels and the scores of the trajectories belonging to the same cluster, using cooperative decision-making strategies. In details, we considered two distinct independent criteria that are used as alternatives. On the one hand, counting of labels $c_{\overline{\boldsymbol{H}}}$ assigned to each class in the cluster is applied, and the class with the majority of labels is finally assigned to the cluster, the *majority voting criterion*. On the other hand, the sum of scores $y_{\overline{\boldsymbol{H}}}$ assigned to each class computed over the cluster is used to assign the class with the highest score, the *maximum trustiness criterion*.

The two criteria are inspired by two different considerations. First, *maj-vot* represents the logic of consensus based on the agreement among artificial labelers (cell trajectories in test). This is in line with the assumption of a unique collective underlying phenomenon in a given experiment. On the other hand, the *max-trust* criterion considers all the scores assigned to the entire cluster giving strength not only to artificial labelers in agreement (majority voting paradigm) but rather to all labelers in the cluster, even those not in agreement. In other words, the latter criterion applied a more democratic principle, giving voice also to minority cell behavior with high scores. Cooperative learning approaches represent the final step of the peer prediction paradigm, in which final decision is taken among peers, after the elimination of abnormal or deviated responders (test samples rejected).

Experimental Setup

Three prostate cell lines were chosen to test the validity of the proposed methodology: RWPE-1 (nonneoplastic cells), LNCaP (neoplastic cells), and PC3 (metastatic neoplastic cells), representing healthy, tumor, and highly aggressive tumor cell phenotypes, respectively. RWPE-1 and PC3 were treated with the chemotherapy agent etoposide at different concentrations (0.5 and 5 μM for RWPE-1, 1 and 5 μM for PC3). RWPE-1 and PC3 were also acquired in control conditions (i.e., no drug). Therefore, for RWPE-1 and PC3, we collected six videos (two ones for each condition), and for LNCaP, we collected four replicated experiments in control condition (globally 16 videos).

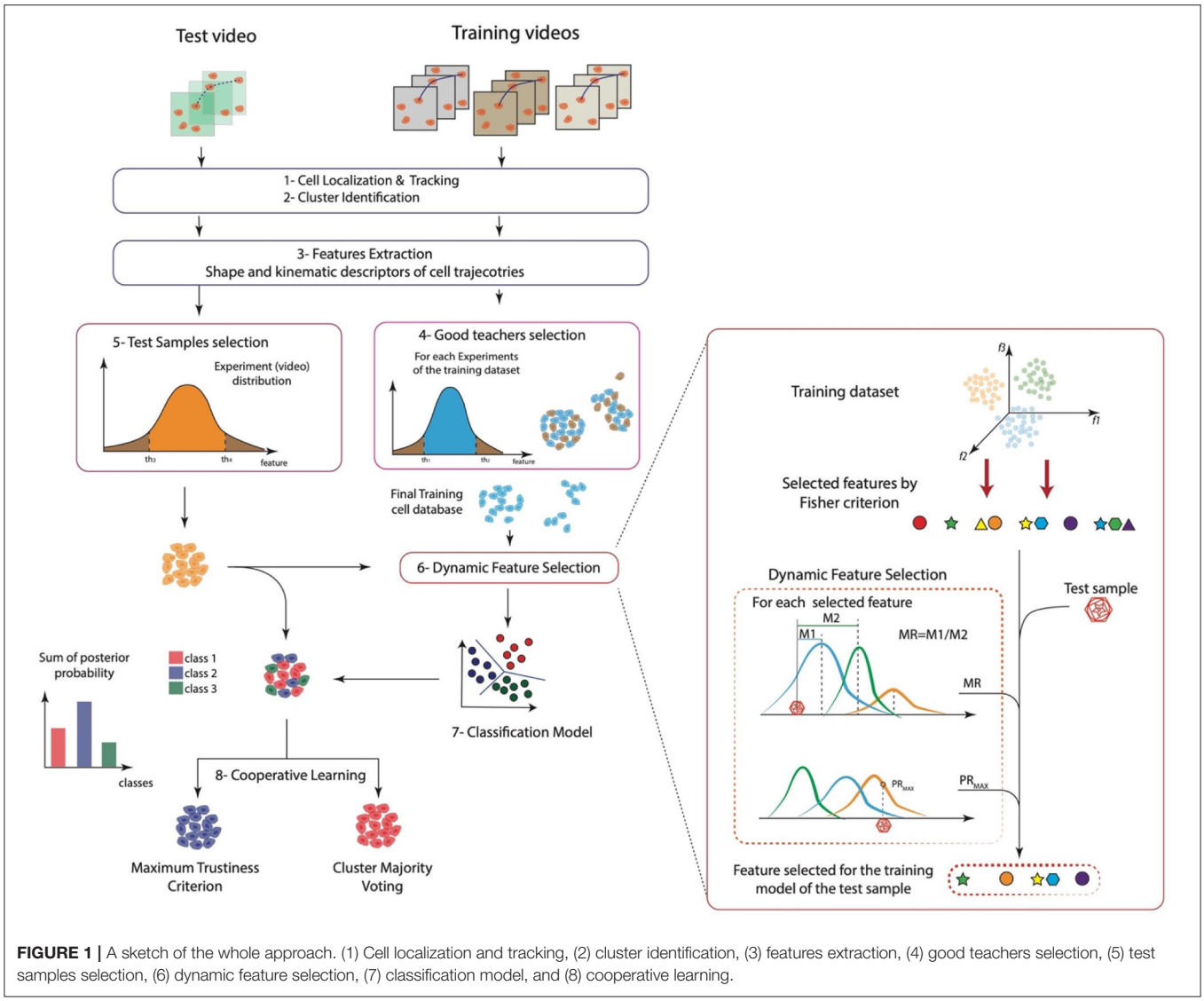


FIGURE 1 | A sketch of the whole approach. (1) Cell localization and tracking, (2) cluster identification, (3) features extraction, (4) good teachers selection, (5) test samples selection, (6) dynamic feature selection, (7) classification model, and (8) cooperative learning.

In order to demonstrate the effectiveness and the general validity of the approach, we ran a leave-one-experiment-out validation procedure, holding out an experiment at a time for testing and using the remaining for training the method. Despite the low number of available experiments, results are very promising, in relation to the challenging identified setup. On the other hand, under the assumption of the intrinsic heterogeneity of the cell behavior in a given group of nominally identical cells, we performed cooperative learning by maj-vot and max-trust criteria applied at cluster level.

An example of clustered cells for the three cell lines is shown in **Supplementary Figure 2**. The color bar indicates the time varying cross the trajectory. Four distinct cell shapes and positions along the corresponding trajectories are also shown. As immediately observed, cell appearance is very heterogeneous, both among the same cell line and along the trajectory of the same cell. This fact demonstrates the difficulty to extract synthetic descriptors from trajectories and construct a model on them for recognizing changes in the cell behavior.

Quantification and Statistical Analysis

To evaluate the performances of all the classification models, a cross validation procedure has been applied.

RESULTS

Setting of the Proposed Approach

In this work, we present a general method to analyze and discriminate cell behavior in controlled *in vitro*–cultured environments. The proposed approach can be divided into eight key steps: (1) cell localization and tracking, (2) automatic cell clustering identification, (3) cell morphology and motility feature extraction, (4) good teacher selection, (5) test samples selection, (6) DFS, (7) classification model, and finally (8) cooperative learning. A schematic representation of the whole approach is

reported in **Figure 1**. Briefly, the method exploited a previously validated cell tracking tool (Cell-Hunter) to automatically locate and track cells. Each cell is then identified as a member of a cell cluster by image analysis algorithms (12) and characterized in terms of kinematics and shape descriptors. To this aim, quantitative descriptors to characterize cell morphology and motility over time were extracted. Good teacher and test sample selection procedures were then applied to retain only those cell trajectories considered as good trainers and good samples, respectively, to construct the model. After training and data collection, DFS further finely selected only those features satisfying the Fisher criterion, Mahalanobis criterion, and the maximum posterior probability, excluding all abnormal behaviors. Model construction was then performed, and two cooperative learning techniques, i.e., the maj-vot and the max-trust, were implemented to ultimately extract a unique collective cell response. Further details on the proposed method for automatic cell behavior classification are reported in Materials and Methods.

Three prostate cell lines were chosen to test the validity of the proposed methodology: RWPE-1 (nonneoplastic cells), LNCaP (neoplastic cells), and PC3 (metastatic neoplastic cells), representing healthy, tumor, and highly aggressive tumor cell phenotypes, respectively, treated with increasing doses of the chemotherapy agent etoposide. Among chemotherapeutics, etoposide was selected because of its well-known effect on both cell shape (**Figure 2**) and motility (19); i.e., it affects the features extracted for the classification method.

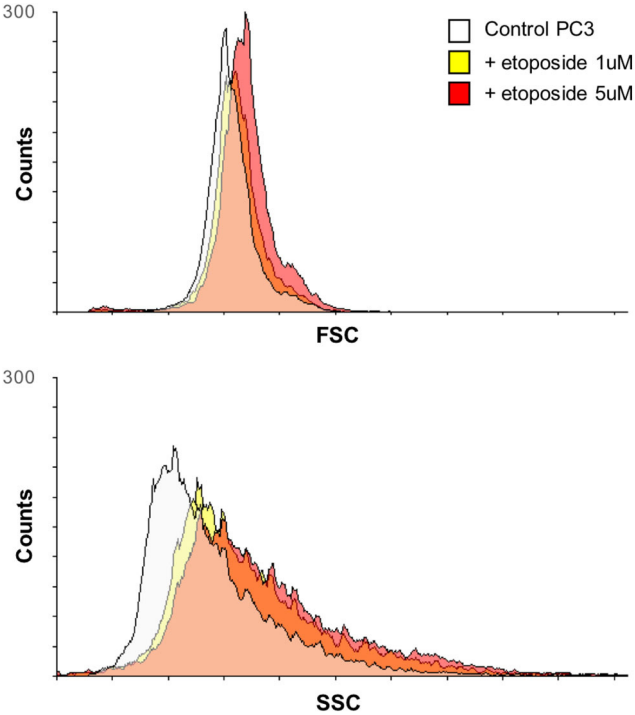


FIGURE 2 | Variation over time of PC3 cell morphology after treatment with etoposide assessed by flow cytometry. Overlays of the forward scatter (FSC) and the side scatter (SSC) of PC3 cells, before and after treatment with etoposide (1 or 5 μM) for 12 h, are reported; the two parameters relate to cell size and granularity, respectively.

Setting of Algorithm Parameters

Quantitative results of the test have been assessed using different indices; balanced accuracy and unbalanced accuracy (ACC_b and ACC, respectively) were computed over the confusion matrix related to the classification results. We reported the results computed over each single-cell trajectory tested (single-cell result) and the results achieved using the cooperative learning strategies. In particular, we show results referred to the maj-vot and to the max-trust criteria. Furthermore, the results were compared with those obtained using standard classification strategies or with the elimination of specific algorithms blocks, such as data test and good teacher (training and/or test) selection. In this way, we demonstrated the validity not only of the whole approach, but also the improvement introduced by each sub-block.

Table 1 lists the parameters values for each test performed for system performance assessment. The values have been estimated by an optimization procedure run on a repeated subsampling of the training set.

Classification Results: The Proposed Method Reached Accuracy Values of 95%

We validated the approach on the automatic recognition of the three different prostate cell lines tested (RWPE-1, LNCaP, and PC3). We separated the results obtained using only shape or motility features, in order to appreciate the relevance of the two groups of descriptors for the task.

In **Figure 3**, we included the confusion matrices using the SVM classifier related to (A, D) single-cell result, (B, E) maj-vot result performed at cluster level, and (C, F) max-trust result performed at cluster level, for shape (A–C) and motility (D–F) features, respectively.

In detail, by using shape descriptors, we obtained accuracy values, ACC (ACC_b), equal to 94.4% (91.8%) for the single-cell result, 95.1% (93.4%) for the maj-vot result, and 94.6% (92.6%) for max-trust result. The highest accuracy values are obtained for RWPE-1 and LNCaP cells. PC3 cells, instead, are misclassified in more than 10% of cases. Nevertheless, the classification error always occurs in the LNCaP class and never in that of the RWPE-1, underlying the great validity of the novel approach that when it fails, it misclassifies only between the two tumor classes (metastatic vs. nonmetastatic tumor cells), in accordance with the heterogeneity-characterizing tumors.

By using only motility features, instead, we obtained lower (although still very promising) accuracy values, ACC (ACC_b), equal to 86.7% (83.5%) for the single-cell result, 91.6% (89.0%) for the maj-vot result, and 91.4% (88.8%) for max-trust result.

The use of shape descriptors therefore improves the global recognition accuracy with respect to motility features. This is a further demonstration of the potential of video analysis in TLM toward the possibility to combine spatiotemporal properties in morphokinetic studies.

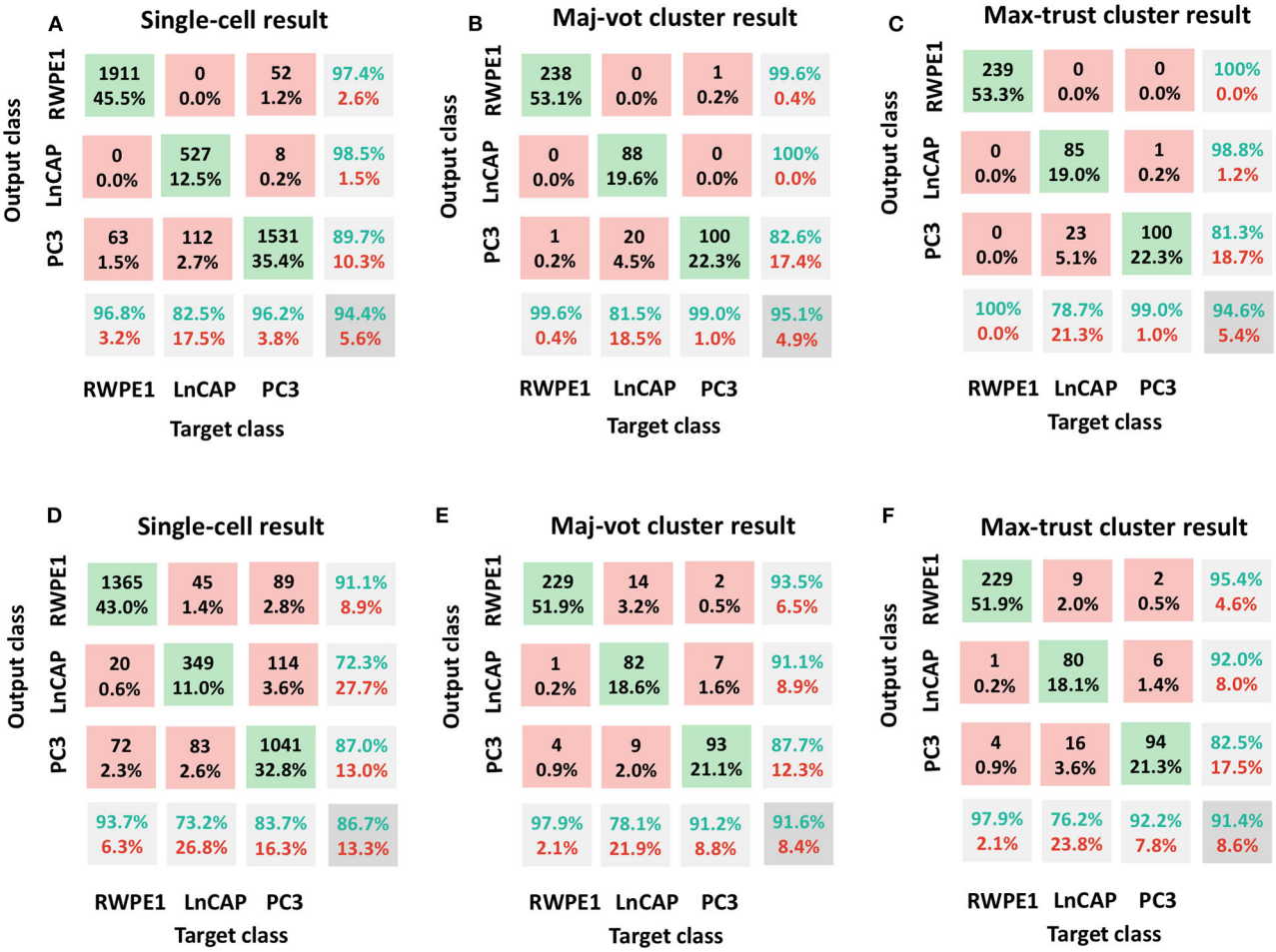


FIGURE 3 | Classification results: RWPE1, LNCaP, and PC3 cell lines. Shape features in **(A–C)** and motility features **(D–F)**. **(A,D)** Single-cell result, **(B,E)** maj-vot cluster-level result, **(C,F)** max-trust cluster-level result. In **(A,D)**, numbers indicate the cells tested, whereas in **(B,C,E,F)**, numbers indicate the number of clusters. Green accuracy values represent the true-positive results for each class, whereas the values in the pink boxes indicate the number of false positives (upper triangular part of the confusion matrix) and false negatives (the lower triangular part of the matrix). The values in the gray box represent the total accuracy.

The Crucial Role of the Good Teacher and Test Sample Selection to Maximize the Classification Performance

In this section, we evaluated the results of the proposed approach based on the three distinct classification models: LDA, SVM, and KNN.

Classification results are shown in **Table 2**. As it can be noted, the three classifiers produced similar results (above all LDA and SVM); cell classification according to the phenotype is effectively solved by the proposed synergic approach. In light of this, LDA remains the simplest model achieving almost the highest performance, to the advantage of an increased architecture and easier interpretation of the results.

In order to demonstrate the crucial role of the good teacher and test sample selection, we conducted two specific tests. First, we totally removed the good teacher selection procedure (Step 4) from the strategies and reported the results of a model constructed on the entire training dataset and the test conducted on all the samples in each cluster. Second, we removed the test sample selection (Step 5); namely, we only selected good trainers but not good samples for testing the results. Numerical results are shown in **Table 2**, columns D and E.

First, we observed that using shape descriptors, performance is higher. This is due to the fact that although cell shape changes during movement, as observed from **Supplementary Figure 2**, and that etoposide administration deeply affects cell shape, this variation is smaller than that existing among distinct cell lines. Therefore, the impact of data selection is strong, but not crucial (we obtained even accuracy values of 88 and 92% without the application of the novel strategies). Data selection, instead, acquires a primary role in the case of motility descriptors; indeed, it increases the accuracy values even by more than 10%.

To classify cell types based on motility features, selection of appropriate cell trajectories results pivotal; indeed, some aspects of cell behavior can be relevant for identifying a certain phenomenon, but less important for a different task. In light

TABLE 2 | Comparative results in terms of balanced accuracy (ACC_b) and accuracy (ACC) of classification.

		(A)	(B)	(C)	(D)	(E)
Shape features		**Proposed approach**	**SVM**	**KNN**	**No good teacher selection**	**No data test selection**
ACC_b	Single-cell	91.94%	92.19%	88.51%	77.19%	85.15%
ACC		94.31%	93.63%	90.10%	79.84%	87.43%
ACC_b	Maj-vote	93.36%	93.15%	92.53%	84.29%	89.95%
ACC		95.09%	95.09%	94.64%	87.53%	91.40%
ACC_b	Max-trust	93.52%	92.55%	92.57%	86.02%	90.26%
ACC		95.31%	94.64%	94.64%	88.36%	91.61%
Motility features		**Proposed approach**	**SVM**	**KNN**	**No good teacher selection**	**No data test selection**
ACC_b	Single-cell	83.51%	85.84%	80.20%	70.15%	69.70%
ACC		86.69%	88.74%	86.06%	74.62%	74.30%
ACC_b	Maj-vote	89.04%	90.54%	86.21%	77.91%	76.56%
ACC		91.61%	92.52%	89.57%	83.43%	81.21%
ACC_b	Max-trust	88.74%	90.10%	88.47%	80.61%	80.20%
ACC		91.38%	92.06%	91.16%	85.25%	83.64%

Top part shows shape features, and bottom part shows motility features. The results of the proposed approach using LDA (A) are compared with the use of (B) support vector machine (SVM) with linear kernel, (C) K-nearest neighbor (KNN) with K = 5, and the cases (D) with no data selection and (E) that in which the good data selection was not applied to the test set.

of this, **Figure 4** shows some examples of clusters and related trajectories for the three cell lines. Using different colors for cell candidates, we could discriminate among cell trajectories extracted through the Cell-Hunter software (cyan) and tracks extracted using the test sample selection approach (green). As can be observed, in most cases, cell trajectories selected for the scope of classification as good test samples lie at the boundary of the cluster (this is particularly evident for RWPE-1 and PC3 cells), suggesting that the behavior of cells within the cluster has a less discriminative role in this case study.

DISCUSSION

In this work, we present a novel methodology combining TLM with cell tracking, providing a quantitative representation of trajectories and novel ML strategies, within peer prediction paradigm. This allows classifying cells in the categories of nontumor, tumor with no metastatic power, and tumor with high metastatic power, on the basis of cell behavior in terms of variations over time of cell morphology and motility.

As any methodology based on ML, we had to consider that such investigations need the identification of the correct learning examples (8). This is a hard task because of the dramatic heterogeneity of cell response, even among apparently similar cells, because of intrinsic different genetic and/or epigenetic assets and extrinsic environmental conditionings. Peer prediction protocol was implemented here to solve the strong heterogeneity of individual cell properties and activity, which renders difficult to represent a cell population as a unique behavioral entity.

To this purpose, during model construction, good teacher selection (12) was applied to cell trajectories; i.e., only those cell trajectories considered as good trainers were selected to construct the good model. The good teacher selection strategy acts therefore as a sort of candidate selection and can be used to visually investigate the role of each selected cell within any cell cluster. Selection was performed again in the testing phase: the test sample selection, indeed, allows excluding cell trajectories not complying with the representative behavior of the examined population, excluding the "noncanonical" behaviors to maximize the classification performances. Importantly, this methodology paves the way to future studies including those cells that behave differently, which could, nonetheless, represent second, third, etc., subpopulations in a heterogeneous mixture. The analysis of the currently labeled but excluded peers, in fact, would be crucial, for example, to investigate the heterogeneous genetic and epigenetic nature of cells within real biological systems, distinguishing between subpopulations. This is especially important in tumors, known to be composed of different cancer cell subpopulations. This is a paramount issue, because cancer cell heterogeneity is a main reason why therapies fail. Importantly, there are presently no straightforward ways to point out diversity. Therefore, the development and validation of the present tool, providing a mean to "barcoding" the different cancer populations, would find immediate application in clinics, with important diagnostic improvements.

To build the classifier for the test label prediction, we then combined the good teacher–good test sample selection strategies to a novel use of the DFS approach (17, 18); extracted features are dynamically selected according to the testing set characteristics. This is allowed by the novel paradigm of autonomy, in which good test samples suggest the optimal descriptors to teachers for optimal working. In line with a social peer prediction paradigm, it is the responders, and not the masters of service, who decide which aspects to judge in service quality assessment.

Through the combination of a novel good teacher–good test sample selection strategies and dynamic features selection

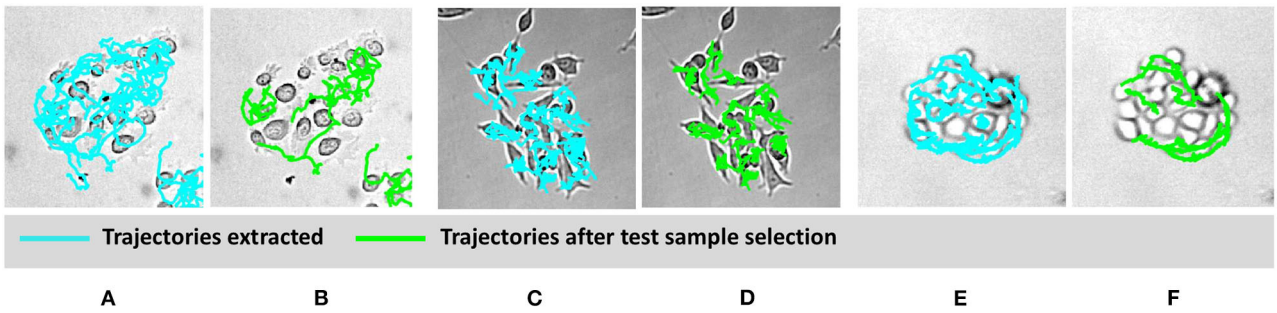


FIGURE 4 | Visual example of selected cell trajectories. RWPE1 **(A,B)**, LNCaP **(C,D)**, PC3 at drug concentration of 1 μM **(E,F)**. The cyan trajectories are those extracted by the cell-tracking software for all the cells in the experiment. The green trajectories are those selected in the good sample selection.

approach for optimal model construction, we were thus able to automatically select cell trajectories for both learning and testing, by excluding cells with noncanonical behavior. The implementation of two cooperative learning techniques based on distinct peer agreement rules ultimately demonstrated the existence of a collective response rather than a collection of individual responses, finally allowing our classifier to get accuracy values of even 95% for shape descriptors.

In this regard, the use of shape together with kinematic descriptors represents a further novelty of the proposed approach. Investigation of cell morphology lost importance over time because of its impossibility of being quantifiable, therefore being not objective and not objectified. In the present work, instead, we demonstrated that the use of shape descriptors improves the global recognition accuracy of the model with respect to only motility features, thus combining spatiotemporal properties in morphokinetics studies.

The promising results achieved strongly suggest that after implementation, for example, extending the study on a larger sample of tumor cell lines, the proposed model could represent a novel tool in understanding cancer, thereby facilitating diagnosis and therapy. Indeed, the proposed predictive system may be employed in diagnostics as a fast method to identify cancer cells possessing a potential metastatic behavior and classify the type, stage, and aggressiveness of a tumor, in addition to the traditional diagnostic biomarkers screened after biopsy. To this purpose, several chemotherapeutics may be rapidly tested on patients' tumor cells, to gain information from the therapy-promoted behavioral changes; this may allow classifying patients' cells according to their aggressiveness, i.e., identifying cells metastatic potential. Noteworthy, our approach accurately correlating cell physical aspects (such as morphology and motility) to cell phenotypes may also be employed to associate different cell motilities to corresponding diverse cancer driver mutations, thus not only predicting cancer cell predisposition to therapies, but also inferring information on oncogenes and/or tumor suppressors role in cancer genesis and progression.

As far as therapy is concerned, instead, the predictive model may be used as an innovative drug screening platform, to identify effective anticancer biomodulating agents (36). Indeed, sets of chemotherapeutics may be tested on aggressive tumor cells, allowing selecting those able to remodulate cell behavior, e.g., shifting cancer cells in a less malignant or even in the nontumor category (phenotypic reversion). The proposed model would therefore allow identifying those drugs able to matter-of-factually "normalize" cancer cell behavior, even allowing case-by-case analyses for personalized therapy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

EM, MD'O, LG, and FC designed the experiments. FC prepared and characterized the biological samples. AM, MD'O, MC, PC, CD, and EM performed the data analysis. MD'O, JF, and DD collected the experimental videos. MD'O, AM, FC, LG, and EM wrote the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2020.580698/full#supplementary-material

REFERENCES

1. Friedl P, Gilmour D. Collective cell migration in morphogenesis, regeneration and cancer. *Nat Rev Mol Cell Biol.* (2009) 10:445–57. doi: 10.1038/nrm2720
2. Corallino S, Malabarba M, Zobel M, Di Fiore P, Scita G. Epithelial-to-mesenchymal plasticity harnesses endocytic circuitries. *Front Oncol.* (2015) 5:45. doi: 10.3389/fonc.2015.00045
3. Mobiny A, Lu H, Nguyen H, Roysam B, Varadarajan N. Automated classification of apoptosis in phase contrast microscopy using

capsule network. *IEEE Trans Med Imaging.* (2020) 39:1–10. doi: 10.1109/TMI.2019.2918181
4. Forcina G, Conlon M, Wells A, Cao J, Dixon S. Systematic quantification of population cell death kinetics in mammalian cells. *Cell Syst*. (2017) 4:600–10.e6. doi: 10.1016/j.cels.2017.05.002
5. Artymovich K, Patel K, Szybut C, Garay PM, O'Callaghan T, Dale T, et al. CellPlayerTM kinetic proliferation assay. *Assay Ess Biosci*. (2016) 1–17.
6. O'Clair L, Roddy M, Tikhonenko M, Syzbut C, Bevan N, Schroeder K, et al. Quantification of cell migration and invasion using the IncuCyteTMChemotaxis assay. *Ess Biosci*. (2015) 1–5.
7. O'Clair L, Artymovich K, Roddy M, Appledorn DM. Quantification of cytotoxicity using the IncuCyte ® cytotoxicity Assay. *Ess Biosci*. (2017) 1–5.
8. Agarwal A, Mandal D, Parkes D, Shah N. Peer prediction with heterogeneous users. *ACM Trans Econ Comput.* (2020) 8:1–34. doi: 10.1145/3381519
9. Liu Y, Helmbold DP. *Online Learning Using Only Peer Assessment*. Palermo, ML: Research Press (2019). p. 1–25.
10. James S, Lanham E, Mak-Hau V, Pan L, Wilkin T, Wood-Bradley G. Identifying items for moderation in a peer assessment framework. *Knowl Based Syst.* (2018) 162:211–9. doi: 10.1016/j.knosys.2018.05.032
11. Zhang Z, Liu L, Wang H, Li J, Hu D, Yan J, et al. Collective behavior learning by differentiating personal preference from peer influence. *Knowl Based Syst.* (2018) 159:233–43. doi: 10.1016/j.knosys.2018.06.027
12. Comes M, Mencattini A, Di Giuseppe D, Filippi J, D'Orazio M, Casti P, et al. A camera sensors-based system to study drug effects on *in vitro* motility: the case of PC-3 prostate cancer cells. *Sensors.* (2020) 20:1531. doi: 10.3390/s20051531
13. Biselli E, Agliari E, Barra A, Bertani F, Gerardino A, De Ninno A, et al. Organs on chip approach: a tool to evaluate cancer -immune cells interactions. *Sci Rep.* (2017) 7:1–12. doi: 10.1038/s41598-017-13070-3
14. Vacchelli E, Ma Y, Baracco EE, Sistigu A, Enot DP, Pietrocola F. Chemotherapy-induced antitumor immunity requires formyl peptide receptor 1. *Science.* (2015) 350:972–8. doi: 10.1126/science.aad0779
15. Comes M, Casti P, Mencattini A, Di Giuseppe D, Mermet-Meillon F, De Ninno A, et al. The influence of spatial and temporal resolutions on the analysis of cell-cell interaction: a systematic study for time-lapse microscopy applications. *Sci Rep.* (2019) 9:6789. doi: 10.1038/s41598-019-42475-5
16. Nguyen M, De Ninno A, Mencattini A, Mermet-Meillon F, Fornabaio G, Evans SS, et al. Dissecting effects of anti-cancer drugs and of cancer-associated fibroblasts by on-chip reconstitution of immunocompetent tumor microenvironments. *Cell Rep.* (2018) 25:3884–93.e3. doi: 10.1016/j.celrep.2018.12.015
17. Magna G, Mosciano F, Martinelli E, Di Natale C. Unsupervised On-Line Selection of Training Features for a robust classification with drifting and faulty gas sensors. *Sens Actuat B Chem*. (2018) 258:1242–51. doi: 10.1016/j.snb.2017.12.005
18. Mosciano F, Mencattini A, Ringeval F, Schuller B, Martinelli E, Di Natale C. An array of physical sensors and an adaptive regression strategy for emotion recognition in a noisy scenario. *Sens Actuat A Phys*. (2017) 267:48–59. doi: 10.1016/j.sna.2017.09.056
19. Di Giuseppe D, Corsi F, Mencattini A, Comes M, Casti P, Di Natale C, et al. Learning cancer-related drug efficacy exploiting consensus in coordinated motility within cell clusters. *IEEE Transact Biomed Eng.* (2019) 66:2882–8. doi: 10.1109/TBME.2019.2897825
20. Rizzuto JSV, Mencattini A, Álvarez-González B, Ortega MA, Ramon-Azcon J, Martinelli E, et al. Microfluidic filtering unit for the evaluation of RBC mechanical properties (Rare haemolytic anaemia model). In: *Proceedings of XXXVII Annual Conference of The Spanish Society of Biomedical Engineering* Santander. (2019). p. 72–5.
21. Gonzalez RC, Woods RE. *Digital Image Processing*. 2nd ed. Boston, MA: Addison-Wesley Longman Publishing Co., Inc. (2001).
22. Atherton TJ, Kerbyson DJ. Size invariant circle detection. *Image Vis Comput*. (1999) 17:795–803. doi: 10.1016/S0262-8856(98)00160-7
23. Yuen H, Princen J, Illingworth J, Kittler J. Comparative study of hough transform methods for circle finding. *Image Vis Comput*. (1990) 8:71–7. doi: 10.1016/0262-8856(90)90059-E
24. Chan TF, Vese LA. Active contours without edges. *IEEE Trans Image Process.* (2001) 10:266–77. doi: 10.1109/83.902291
25. Pratt WK. *Digital Image Processing: PIKS Inside*, 3rd Edn. (1994).
26. Zdilla M, Hatfield S, McLean K, Cyrus L, Laslo J, Lambert H. Circularity, solidity, axes of a best fit ellipse, aspect ratio, and roundness of the foramen ovale. *J Craniof Surg.* (2016) 27:222–8. doi: 10.1097/SCS.0000000000002285
27. Craven GWP. Smoothing noisy data with spline functions. *Numer Math*. (1979) 31:373–403. doi: 10.1007/BF01404567
28. Sbalzarini IF, Koumoutsakos P. Feature point tracking and trajectory analysis for video imaging in cell biology. *J Struct Biol*. (2005) 151:182–95. doi: 10.1016/j.jsb.2005.06.002
29. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. (1982) 143:29–36. doi: 10.1148/radiology.143.1.7063747
30. Lianou A, Mencattini A, Catini A, Di Natale C, Nychas G, Martinelli E, et al. Online feature selection for robust classification of the microbiological quality of traditional vanilla cream by means of multispectral imaging. *Sensors.* (2019) 19:4071. doi: 10.3390/s19194071
31. Draper NR, Smith H. *Applied Regression Analysis*. Hoboken, NJ: Wiley-Interscience (1998).
32. Bevington PR, Robinson DKJ, Blair M, Mallinckrodt AJ, McKay S. Data reduction and error analysis for physical sciences. *Comput Phys*. (1993) 7:415–6. doi: 10.1063/1.4823194
33. Davies E. *Machine Vision: Theory, Algorithms, Practicalities*. Amsterdam: Elsevier (2004).
34. Schölkopf B, Sung K, Burges C, Girosi F, Niyogi P, Poggio T, et al. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Transact Signal Process*. (1997) 45:2758–65. doi: 10.1109/78.650102
35. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat.* (1992) 46:175–85. doi: 10.1080/00031305.1992.10475879
36. Thomas S, Schelker R, Klobuch S, Zaiss S, Troppmann M, Rehli M, et al. Biomodulatory therapy induces complete molecular remission in chemorefractory acute myeloid leukemia. *Haematologica.* (2014) 100:e4–6. doi: 10.3324/haematol.2014.115055

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 30 October 2020
doi: 10.3389/fbioe.2020.534592

Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI

Guilherme Folego[1,2], Marina Weiler[3], Raphael F. Casseb[4], Ramon Pires[1] and Anderson Rocha[1] for the Alzheimer's Disease Neuroimaging Initiative[†] and the Australian Imaging Biomarkers and Lifestyle flagship study of aging[‡]*

[1] Institute of Computing, University of Campinas, Campinas, Brazil, [2] CPQD, Campinas, Brazil, [3] Laboratory of Behavioral Neuroscience, National Institute on Aging, National Institutes of Health, Intramural Research Program (NIA/NIH/IRP), Baltimore, MD, United States, [4] Seaman Family MR Research Center, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

OPEN ACCESS

Edited by:
Concetto Spampinato, University of Catania, Italy

Reviewed by:
Amit Alexander, National Institute of Pharmaceutical Education and Research, India
Carmelo Pino, University of Catania, Italy

***Correspondence:**
Guilherme Folego
gfolego@gmail.com

[†]Some data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report

[‡]Some data used in the preparation of this article were obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of aging (AIBL) funded by the Commonwealth Scientific and Industrial Research Organization (CSIRO) which was made available at the ADNI database

Specialty section:
This article was submitted to Biomaterials, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *13 February 2020*
Accepted: *18 September 2020*
Published: *30 October 2020*

The projected burden of dementia by Alzheimer's disease (AD) represents a looming healthcare crisis as the population of most countries grows older. Although there is currently no cure, it is possible to treat symptoms of dementia. Early diagnosis is paramount to the development and success of interventions, and neuroimaging represents one of the most promising areas for early detection of AD. We aimed to deploy advanced deep learning methods to determine whether they can extract useful AD biomarkers from structural magnetic resonance imaging (sMRI) and classify brain images into AD, mild cognitive impairment (MCI), and cognitively normal (CN) groups. We tailored and trained Convolutional Neural Networks (CNNs) on sMRIs of the brain from datasets available in online databases. Our proposed method, ADNet, was evaluated on the CADDementia challenge and outperformed several approaches in the prior art. The method's configuration with machine-learning domain adaptation, ADNet-DA, reached 52.3% accuracy. Contributions of our study include devising a deep learning system that is entirely automatic and comparatively fast, presenting competitive results without using any patient's domain-specific knowledge about the disease. We were able to implement an end-to-end CNN system to classify subjects into AD, MCI, or CN groups, reflecting the identification of distinctive elements in brain images. In this context, our system represents a promising tool in finding biomarkers to help with the diagnosis of AD and, eventually, many other diseases.

Keywords: Alzheimer's disease, computer aided diagnosis, artificial intelligence, computer vision, deep learning, convolutional neural networks, image classification, magnetic resonance imaging

1. INTRODUCTION

Dementia by Alzheimer's disease (AD) is characterized by multiple cognitive problems, including difficulties in memory, executive functions, language, and visuospatial skills. The most significant risk for AD is aging—there is almost a 15-fold increase in the prevalence of dementia between the ages of 60 and 85 years (Evans et al., 1989). The projected burden of the disease represents a looming healthcare crisis as the population of most industrialized countries continues to grow older. Although there is still no cure, it is possible to treat both cognitive and behavioral symptoms of AD.

The early diagnosis of the disease is paramount for interventions, and clinical trials in AD tend to enroll subjects at earlier time-points before neuronal degeneration has achieved a particular stage

and treatment is often more effective. In this context, neuroimaging is one of the most promising areas of research for early detection of AD, as the progressive degeneration of brain structures can be seen as a dramatic cerebral shrinkage in structural magnetic resonance imaging (sMRI).

Thus far, works in this area have recurrently considered only a small number of subjects and images, often with curated data (i.e., reviewed, prepared, and organized by experts), such as ADNI's Standardized MRI Data Sets (Wyman et al., 2013). Additionally, with the lack of a standard evaluation protocol, each study employed its criteria, with its own random data split. The lack of standardization limits the comparison between different methods and usually overestimates performance in a real-world scenario. When data is not readily preprocessed and comes from different sources, this situation is even more problematic.

A recent and extensive review (Wen et al., 2020) indicated that a reasonable number of studies using convolutional neural networks (CNNs) for AD either present evident data leakage problems, or offer scarce explanation for the validation method to ensure that data leakage has not occurred. Data leakage possibilities only emphasize the need for an independent set of images for evaluation. In addition to this review, Wen et al. (2020) proposed a standard framework for rigorous performance assessment, using data from ADNI, AIBL (Ellis et al., 2009), and OASIS (Marcus et al., 2007, 2009).

For fair comparisons between different methods, a few challenges with standard protocols and hidden test labels were launched, such as the CADDementia challenge (Bron et al., 2015). Although presenting a good classifier—63.0% accuracy in classifying MRI images into cognitively normal (CN), mild cognitive impairment (MCI), and AD patients—, the winning method (Sørensen et al., 2017) used transductive inference to calculate hippocampal shape scores, requiring the CADDementia test data to be calculated, which deviates from the original proposal of applying the algorithm in the clinical setting. Additionally, their pipeline failed to process three scans from the CADDementia test set, requiring manual intervention. The analysis of each subject took 19 h of computation time. The second-best team (Wachinger and Reuter, 2016) employed a domain-adaptation approach, and optimization was done on the union of ADNI and CADDementia training sets, with equal weights for each sample. The analysis of each subject took 17.4 h of computation time.

Among the available machine learning methods, CNNs have been increasingly used in the Alzheimer's biomarker identification task, given its power to learn discriminative representations hierarchically in an automated fashion. Most studies employing CNNs in the context of AD used 2D inputs, whereas studies that used 3D inputs focused basically on binary classifications. The few works that trained a 3D CNN for the multiclass CN/MCI/AD classification evaluated their performances with cross-validation on ADNI data and considered only networks comparable to our smallest proposed architecture. All of this context highlights the novelty in our research, as we optimized very deep 3D CNNs, with up to 22 layers, for the multiclass diagnosis task, and evaluated our performance on the CADDementia challenge, with unknown labels, making our results much more reliably applicable in a real-world setting.

A previous work that employed the deep learning approach in the context of AD (Korolev et al., 2017) designed 3D CNNs based on smaller versions of VGG (Simonyan and Zisserman, 2014) and ResNet (He et al., 2016) architectures. However, only binary classification tasks were considered, which were evaluated using cross-validation on ADNI. The first group to successfully propose a deep-learning approach to the CADDementia challenge (Dolph et al., 2017) extracted features such as cortical thickness, surface area, volumetric measurements, and texture. These values were used to greedily layer-wise train a stacked auto-encoder with three hidden layers, achieving competitive results. Using the whole brain as input, Hosseini-Asl et al. (2018) employed a small stack of unsupervised 3D convolutional auto-encoders (3D-CAE), evaluating with cross-validation on ADNI. One of the works more closely related to ours (Esmaeilzadeh et al., 2018) optimized small 3D CNNs, similar to our most basic model, and considered the multiclass classification task. However, the performance was measured using ADNI cross-validation, hindering better comparisons with our method. It is worth noting that they reported a classification accuracy of 61.1%, but observed overfitting in training data.

In more recent work, Abrol et al. (2020) developed 3D CNNs based on the ResNet architecture and experimented with several binary and multiclass tasks. They used the ADNI data to create a training set, used for cross-validation, and a small test set. Even though their results were promising, no further comparisons with different datasets or standardized evaluation frameworks were made. Interestingly, their experiments also presented overfitting in training data. Also recently, Mehmood et al. (2020) adapted the VGG architecture to create a 2D siamese CNN. They evaluated their model using training and test split in the OASIS dataset, presenting compelling results. However, given their proposed data flow chart, it is possible that their data augmentation technique also introduced data leakage.

In our research, we relied upon a 3D CNN with data primarily provided by ADNI (Mueller et al., 2005) and evaluated on the CADDementia challenge (Bron et al., 2015). Our solution also includes an accountable mechanism to allow us to understand its decisions. Our experiments were conducted in a scenario similar to real-world conditions, in which a CAD system is used on a dataset that is different from the one used for training.

The main challenges and contributions of our research included devising a deep-learning solution completely automatic and comparatively fast, while also presenting competitive results without using any domain-specific knowledge. Our method, named ADNet, yields considerable gains in accuracy, outperforming several other systems in the prior art, all of which require prior knowledge of the disease, such as specific regions of interest from input images. Alternatively, our system does not require any manual intervention, clinical information, or *a priori* selected brain regions. The main reason for not using any information from the disease is to empower the system to automatically learn and extract relevant patterns from regions of the brain, and eventually enable it to support current diagnosis standards for known or new diseases. In addition, it runs 80

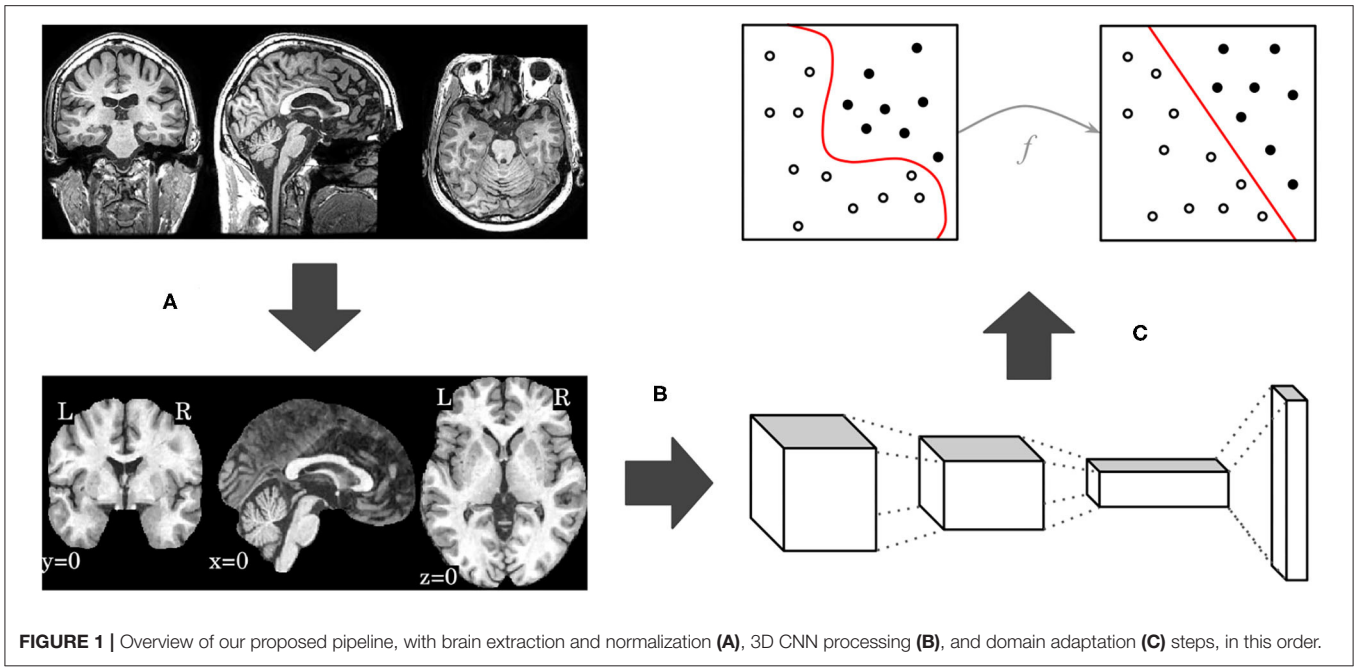


FIGURE 1 | Overview of our proposed pipeline, with brain extraction and normalization **(A)**, 3D CNN processing **(B)**, and domain adaptation **(C)** steps, in this order.

times faster than the state of the art (Sørensen et al., 2017), on average.

Our generated ADNet and ADNet-DA models, as well as supporting code, are publicly available[1] to be used or trained on new data. With this work, we are releasing one of the first models ready to use, encouraging open science and reproducible research, while also setting a starting point for researchers working with 3D MRIs.

2. METHODOLOGY

In this work, we propose an end-to-end deep 3D CNN for the multiclass AD biomarker identification task, using the whole image volume as input. Our pipeline, displayed in **Figure 1**, is composed of three main steps: brain extraction and normalization (**Figure 1A**), 3D CNN processing (**Figure 1B**), and domain adaptation (**Figure 1C**). This section provides details of our pipeline, including image preprocessing, CNN architectures, and optimization techniques.

2.1. Brain Extraction and Normalization

Optimizing deep-learning systems using sMRIs in their original space requires the systems to learn discriminative patterns invariant to several transformations, demanding larger models and an even larger number of samples, with all expected variations. By registering our images to a standard template, we can expect similar structures to be roughly in the same spatial location, allowing us to handle the entire image at once and automatically determine the most important regions of interest.

We used the Advanced Normalization Tools (ANTs; Avants et al., 2009) to extract and normalize brain images. Our pipeline was based on previously defined scripts[2] (Avants et al., 2011; Tustison and Avants, 2013), and we made use of the provided default parameters, including transformation types, sequence, and metrics. Essentially, our brain extraction and normalization pipeline comprised the following steps: image intensity winsorizing, bias field correction, another winsorizing step, translation alignment, rigid transform, affine transform, deformable symmetric normalization (SyN), application of brain mask from the atlas, and range normalization.

As we used registered brains in our research, we opted for a less rigid and less linear atlas, allowing for some degree of variation during the registration process. This atlas also had a high spatial resolution, so finer details would not be lost in the process. As such, the Montreal Neurological Institute (MNI) 152 International Consortium for Brain Mapping (ICBM) 2009c Nonlinear Asymmetric $1 \times 1 \times 1\ mm^3$ (Collins et al., 1999; Fonov et al., 2009, 2011) atlas was chosen.

After the brain extraction and normalization process, the output image has the same dimensions as the atlas (i.e., $193 \times 229 \times 193$). From all of these $8,530,021$ voxels, only $1,886,574$ (22%) of them are not zero. Since the brain is enclosed in a smaller region inside the image, we removed the border dimensions that contained no information, resulting in a final image of $145 \times 182 \times 155$. This new space represents 48% of the original volume, reducing sparsity from 78 to 54%. Finally, we used the training set to compute mean and variance, then used them across all sets to normalize the data to zero mean and unit variance. Given that the used datasets did not fit in main memory, we adopted a single-pass online mean and estimated variance algorithm (Welford, 1962).

[1] https://github.com/gfolego/alzheimers

[2] Specifically, scripts *antsBrainExtraction.sh* and *antsRegistrationSyNQuick.sh* of version 2.1.0.

TABLE 1 | CNN architectures evaluated in this study.

Architecture	Layers	Parameters (in millions)
LeNet-5	7	0.3
VGG 2048	11	89.8
VGG 512	11	26.8
GoogLeNet	22	14.6
ResNet A	18	33.0
ResNet B	18	33.2

Our main challenge was putting together a registration pipeline, including the adopted atlas, that provided useful and meaningful results in a reasonable time, while maintaining high resolution images.

2.2. Convolutional Neural Network

We describe here the convolutional neural network architectures we adopted and the modifications we performed to each. These networks were initially designed for 2D color images, and we are dealing with 3D grayscale MRIs. Thus, the most natural adaptation was to convert all 2D operations, such as convolution or pooling, to 3D. Given these adaptations, we were unable to employ a transfer learning approach (Sharif Razavian et al., 2014) with the original networks directly.

A common attribute to all considered architectures is that spatial dimension is reduced as information flows to deeper layers. Spatial dimensionality reduction is usually achieved with max-pooling layers, or with more substantial strides in convolutional layers. To accommodate different data shapes that were not necessarily divisible by two, we adopted an *ad-hoc* approach by zero-padding each layer as needed, so no information was lost. We also added batch normalization (Ioffe and Szegedy, 2015) to every convolutional and fully-connected layers. All activation functions were rectified linear units (ReLU; Nair and Hinton, 2010), except for the classification output, which was a softmax function. Finally, the number of layers varied according to the adopted network standard. **Table 1** shows the CNN architectures evaluated, taking into account the original approach for layer counting in each network.

We started with a small network, based on the LeNet-5 architecture (LeCun et al., 1989; Lecun et al., 1998). Because this network is significantly older than the others, it required the largest modifications. This network was composed of the following layers: convolution, subsampling, convolution, subsampling, fully connected (originally implemented as convolutional), fully connected, and output. As subsampling layers had learnable parameters, we converted them to convolutions, with filter (kernel) size and stride equal to $2 \times 2 \times 2$, thus keeping the subsampling behavior. The main difference was in the connection between the first subsampling and second convolutional layers, for which the particular arrangement in the original work was converted to a dropout layer with a probability of 40%. Similar to the original architecture, if we had adapted the last convolutional layer to match the previous layer's output size, it would have had 120 feature maps with a kernel size of $34 \times 43 \times 36$, seriously increasing the number of parameters. To mitigate these issues, we adapted those kernels to $5 \times 5 \times 5$ and added a global average pooling layer immediately following, similar to GoogLeNet (Szegedy et al., 2015) and ResNet (He et al., 2016). Naturally, the last layer contained three units (one for each class), with a softmax function activation.

The Visual Geometry Group (VGG) proposed deep CNN architectures, achieving second place in the classification task at ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014 (Simonyan and Zisserman, 2014). They designed very uniform architectures ranging from 11 (configuration A) to 19 (configuration E) weight layers, i.e., considering only convolutional and fully-connected layers. Due to its uniformity, mostly with filters of size 3×3, the VGG architecture is considerably large. The first layers, with the original input dimensions, consumed a large GPU memory, while the last layers, with dense connections, generated several parameters. Since our input data were already quite large when compared to traditional 2D images, we adapted the VGG network configuration A by halving all numbers of filters in convolutional layers, and all numbers of units in fully-connected layers, while keeping filters sizes of $3 \times 3 \times 3$ and dropout rate at 50%. Even after reducing the network size, the first fully-connected layer of our adapted VGG-A, with 2,048 units, accounted for 78, 643, 200 (88%) parameters. For comparison, **Table 1** also includes our VGG-A with 512 units in the fully-connected layers.

While VGG achieved second place in ILSVRC-2014, GoogLeNet secured first place in the classification task (Szegedy et al., 2015), proposing an architecture named Inception. The basic idea was to increase both depth and width while keeping computational requirements constrained. This approach led to a deeper model with fewer parameters and better performance. We adapted directly from their GoogLeNet architecture, i.e., only discarding the local response normalization (Krizhevsky et al., 2012) layer and the auxiliary networks. We also adjusted the last average pooling layer, following the output shape of the previous layer, and kept the dropout rate at 40%. In this architecture, the number of layers came from depth, where single convolutional or fully-connected layers counted as one, while inception modules counted as two. However, each inception module had six individual internal convolutional layers.

In ILSVRC-2015, Residual Network (He et al., 2016) won first place in classification, localization, and detection tasks. These researchers wanted to understand whether learning better networks meant simply stacking more layers. With this study, they found the degradation problem, where traditional models similar to VGG stopped improving performance after a certain number of layers, and even started getting worse afterwards. To overcome this problem, they proposed the residual function, which is the basic building block of a Residual Network (ResNet). We adapted ResNet directly from the non-bottleneck 18-layer architecture, in which shortcuts with increasing dimensions were either (A) identity shortcuts, i.e., padding with zero, or (B) projection shortcuts, i.e., convolutions with $1 \times 1 \times 1$ filter (kernel) size. Similarly to VGG, the number of layers came from convolutional and fully-connected

layers, with projection convolutions not considered in the layer count.

In summary, we adopted four CNN architecture designs, namely, LeNet-5, VGG, GoogLeNet, and ResNet. LeNet-5 is considerably older and smaller, so it shall have a lower probability of overfitting. The VGG network is known for its uniformity, which makes it relatively simple to adapt, inspect, and use for many different tasks; however, this characteristic also makes it significant in the number of parameters and in hardware requirements. These drawbacks were addressed in both GoogLeNet and ResNet architectures, which also adopted very specific building blocks, making it possible to extract more complex patterns from data, while also increasing the number of layers and reducing the number of parameters. The idea was to explore different architectures and understand how they would behave in the AD task.

To avoid overfitting, we adopted regularization with L1 and L2 norms. In L1, this effect is achieved by minimizing the absolute values of the weights, while in L2 this is done with their squared values. In principle, L2 norm tends to produce diffuse and small numbers, while L1 tends to produce sparse numbers. This property makes L1 particularly well-suited to handle noisy data, acting as a feature selection algorithm, which could help us better visualize and explain what the CNN has learned. However, in general, L2 can be expected to provide superior results over L1.

All network architectures and their optimization were implemented using upstream (i.e., the latest version from the code repository) Lasagne (Dieleman et al., 2015), which is a deep learning framework based on Theano (Al-Rfou et al., 2016). At the time this research was carried out, we used a development version of Lasagne 0.2, and a development version of Theano 0.9.0, with Python 2.7.6, CUDA 7.5, and CuDNN 5. Additionally, we used SciKit-learn 0.18.1 (Pedregosa et al., 2011) and NumPy 1.11.3 (van der Walt et al., 2011).

2.3. Domain Adaptation

In addition to brain processing and CNN pipelines, we considered a domain adaptation approach. In our method, we trained a system using one dataset and evaluated it on a different dataset (i.e., CADDementia). Even though they are related, such differences mean that the source data distribution could be different from the target data distribution. Thus, it should be possible to further improve the results by adapting the previously-trained system to the new dataset, even if using only a small number of samples from this target domain. This scenario, also known as cross dataset validation, is more closely related to a real-world scenario, in which data sources will most likely be different between training and actual usage. Additionally, this is a more reliable way of assessing generalization capabilities of a machine learning algorithm.

In our domain adaptation approach, we started by using our previously-optimized CNN to extract features from the complete target dataset (i.e., CADDementia), in one of the last CNN layers. After, we normalized these features to zero mean and unit variance, using only the target training set to compute the parameters. With the normalized data, we optimized a one-vs.-rest logistic regression (McCullagh, 1984) on the complete target training set. In order to find the best parameters for this classifier, we used grid search with leave-one-out cross-validation. Then, we finally had a system that was enhanced for the target domain, making it possible to output improved classification probabilities for each sample in the target domain. This pipeline is similar to a transfer-learning approach (Sharif Razavian et al., 2014).

TABLE 2 | Datasets considered in this study.

Dataset	Number of MRI images	References
ADNI (ADNI1, ADNIGO, ADNI2)	18,303	Mueller et al., 2005; Beckett et al., 2015
AIBL	1,098	Ellis et al., 2009
CADDementia	384	Bron et al., 2015
MIRIAD	708	Malone et al., 2013
OASIS	3,056	Marcus et al., 2007, 2009

3. EXPERIMENTAL SETUP

Given that training a CNN from scratch usually requires massive amounts of data, we gathered as many different imaging sources as possible. We collected an AD sMRI dataset comprising 23,165 images. Below, we describe our optimization approach, including associated parameters.

3.1. Data

In our data collection process, we considered the datasets indicated in **Table 2**. ADNI1 originally included three participant groups: CN, MCI, and AD. Starting in ADNIGO, the MCI stage was split into early MCI (eMCI) and late MCI (lMCI). Later, in ADNI2, a subjective memory complaint (SMC) group was added (Beckett et al., 2015). Similarly to ADNI1, both AIBL and CADDementia sets were composed of CN, MCI, and AD stages, whereas both MIRIAD and OASIS sets contained only CN and AD.

Since one of our main goals for this research was achieving a good result in the CADDementia challenge, we adopted only equivalent diagnoses. As such, eMCI and lMCI stages were grouped along with MCI, and SMC was not considered. From these datasets, we downloaded all available raw T_1-weighted sMRI scans associated with Alzheimer's, i.e., we did not consider any pre- or post-processed image.

To isolate possible confounding factors, we made a distinction between MP-RAGE and IR-SPGR/IR-FSPGR sequences (Lin et al., 2006; Jack et al., 2008), and aggregated different data sources and sequence techniques in steps. While all ADNI sets had both MP-RAGE and IR-SPGR/IR-FSPGR, AIBL and OASIS had only MP-RAGE, and MIRIAD had only IR-FSPGR. The resulting datasets are described in **Table 3**, and detailed in **Table 4**.

For each dataset, we created training, validation, and test splits. In Dataset 1, we randomly split the corresponding subjects, trying to keep the original age, sex, and diagnostic stratification across each set, with 70% of subjects for training, 10% for

TABLE 3 | Datasets assembled in this study.

Resulting dataset	Overarching sets	MP-RAGE only	Final number of MRI images
Dataset 1	ADNI1	Yes	9,149
Dataset 2	All ADNI	Yes	15,885
Dataset 3	All ADNI	No	18,303
Dataset 4	All	No	23,165

validation, and 20% for testing. In each subsequent dataset, we first assigned images from previous subjects to the respective set, then we proceeded with the stratified random split considering only new subjects.

3.2. Metrics and Optimization

The primary evaluation measure we considered herein was classification accuracy, which is the number of correctly classified samples divided by the number of all samples. Even though this performance value does not take into account class priors, the challenge organization deemed class sizes insignificantly different, therefore regarding this metric as a better approach for overall classification accuracy. Additionally, the receiver operating characteristic (ROC) curve and the respective area under the curve (AUC) were also considered, as they provide metrics that are independent of the threshold chosen for classification. Also, since AUC does not traditionally depend on class sizes, we adopted an AUC measure that does not rely on class priors. Finally, the true positive fraction (TPF), the number of correctly classified samples of a given class divided by the number of all samples from that class, was calculated for each class. According to the authors, TPFs for diseases (AD and MCI) can be interpreted as the two-class sensitivity, while TPF for CN corresponds to the two-class specificity.

As we optimized and trained our networks, we compared them and selected the best ones using the average of TPFs, since it more closely relates to the accuracy and does not depend on class priors. To perform the training process, we used Adam optimizer (Kingma and Ba, 2014) with default parameters (i.e., $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$). With a small sample of images, we empirically decided to begin with a learning rate of $\alpha = 10^{-4}$, and settled to a batch size of three (for VGG architectures) or nine (for all the others), mainly due to GPU memory limitations, even though we only used GPUs with 12 GB of dedicated memory. Finally, we adopted Glorot uniform initialization (Glorot and Bengio, 2010) with scaling factor of $\sqrt{2}$.

4. RESULTS AND DISCUSSIONS

Fairness, accountability, and transparency (FAT) have become increasingly essential aspects of machine learning (Goodman and Flaxman, 2016). For example, laudable efforts include explaining algorithmic decisions, making an effort to understand sources of error and uncertainty, and creating auditable systems (Diakopoulos et al., 2017).

Given the expectations described above, we will now discuss further details of our study. We better describe our optimization process, specifying the steps to handle overfitting problems. Then, we report performance results, including previously described metrics, along with efficiency measurements. Finally, we discuss our best CNN model, providing further insights into its functionality, and how it processes data to make predictions.

4.1. Optimization

As stated earlier, we determined the initial learning rate of $\alpha = 10^{-4}$, and varied some configurations in each architecture to achieve the best accuracy in the CADDementia training set. These options included regularization with L1 and L2 norms, regularization strength λ, number of units in fully-connected layers, dropout rates, batch size, and multi-class hinge loss (instead of the traditional categorical cross-entropy loss).

The parameters for regularization strength, number of units, and dropout rates were also used for regularization, acting as trade-offs between model complexity and bias, thus managing the probability to overfit. Overfitting was a significant concern for us due to the large size of our networks and a relatively small amount of data. The different batch size was an experiment to compare the behavior of all networks with the same batch size of three. Given that support vector machine (SVM; Cortes and Vapnik, 1995) classifiers usually present reasonable results, and have successfully been used to identify Alzheimer's biomarker previously (Magnin et al., 2009), we also experimented with the multi-class hinge loss.

In general, we varied regularization strength λ in powers of 10, between 10^{-5} and 10^{2}, number of units in fully-connected layers in powers of 2, between 32 and 2,048, and dropout rate with steps of 10 percentage points, between 40 and 90%, including 95, 99, and 99.9%. Note that some networks had specific parameters, i.e., these variations did not apply to all evaluated architectures. We followed a greedy approach, by first tuning regularization strength with L2 norm, followed by several units, and then dropout rates. Next, we evaluated batches of size three for all networks, L1 norm, multi-class hinge loss, and, finally, larger datasets.

To form a balanced batch, the same number of samples was consistently selected from each class. In each epoch, we randomly sampled each class, limited by the class with fewer images. We worked with a total batch size of either three or nine samples, depending on the network architecture.

We observed that, at some point, most networks underfit or overfit, presenting erratic metrics, with high variations between epochs. To overcome this issue, we applied the early stopping to interrupt the optimization before the model began to overfit. After 50 epochs without further improvement in average TPF over the validation set, the training was stopped. The model was optimized for up to 200 epochs.

4.2. Performance

We first analyzed the efficiency of our processing pipeline, divided into brain image, CNN, and domain adaptation stages.

Figure 2 shows a histogram and a kernel density estimation of the execution time of brain extraction and normalization steps

TABLE 4 | Datasets summaries: number of subjects, number of images, descriptive age statistics, image-wise percentage of females (vs. males) and image-wise percentage of 1.5 T field strength (vs. 3.0 T).

Dataset	Subjs.	Group	Images	Age (years)				Female (%)	1.5 T (%)
				Med	Avg ± Std	Min	Max		
Dset. 1	845	All	9,149	76.6	76.3 ± 6.9	54.6	93.0	42.2	82.2
		CN	2,701	76.7	77.2 ± 5.1	60.0	92.8	50.2	80.5
		MCI	4,845	76.5	76.0 ± 7.4	54.6	90.9	35.3	83.0
		AD	1,603	76.5	76.1 ± 7.9	55.2	93.0	49.5	82.5
Dset. 1 Train.	591	All	6,314	76.5	76.2 ± 6.9	54.6	93.0	43.4	82.6
		CN	1,809	77.2	77.3 ± 4.9	60.0	90.8	49.5	81.3
		MCI	3,399	76.1	75.7 ± 7.3	54.6	90.9	36.3	83.0
		AD	1,106	75.9	76.1 ± 7.9	55.2	93.0	55.3	83.5
Dset. 1 Val.	84	All	951	76.4	75.8 ± 6.8	56.2	89.2	40.5	82.8
		CN	301	75.7	76.5 ± 4.8	65.2	88.6	58.5	79.7
		MCI	501	78.2	76.7 ± 6.7	56.2	89.2	28.5	83.8
		AD	149	72.0	71.2 ± 8.6	56.5	85.0	44.3	85.2
Dset. 1 Test	170	All	1,884	77.2	77.0 ± 6.9	56.7	92.8	38.7	80.4
		CN	591	76.2	77.2 ± 5.6	63.3	92.8	47.9	78.5
		MCI	945	77.7	76.5 ± 7.8	56.7	90.9	35.1	82.4
		AD	348	79.7	78.0 ± 6.3	63.1	87.7	33.0	78.2
Dset. 2	1 503	All	15,885	75.8	75.4 ± 7.3	54.6	95.8	44.0	53.3
		CN	4,646	76.8	76.9 ± 5.8	56.3	95.8	50.0	56.5
		MCI	8,940	75.0	74.6 ± 7.7	54.6	93.5	40.0	50.5
		AD	2,299	76.4	75.8 ± 7.8	55.2	93.0	47.5	57.5
Dset. 3	1 715	All	18,303	75.8	75.5 ± 7.4	54.6	95.8	43.5	48.2
		CN	5,361	76.7	76.9 ± 6.0	56.3	95.8	50.0	52.5
		MCI	10,306	75.0	74.6 ± 7.7	54.6	93.6	39.5	45.5
		AD	2,636	76.2	75.8 ± 7.9	55.2	93.0	45.9	50.2
Dset. 4	2 984	All	23,165	75.0	73.5 ± 11.7	18.0	98.0	46.5	55.5
		CN	8,462	75.0	71.3 ± 16.1	18.0	97.0	53.9	62.8
		MCI	10,460	75.0	74.7 ± 7.7	54.6	96.0	39.6	45.1
		AD	4,243	75.4	75.3 ± 7.9	55.0	98.0	48.4	66.3
CADD. Train.	30	All	30	65.0	65.2 ± 6.9	54.0	80.0	43.3	0.0
		CN	12	62.0	62.3 ± 6.1	55.0	79.0	25.0	0.0
		MCI	9	68.0	68.0 ± 8.2	54.0	80.0	44.4	0.0
		AD	9	67.0	66.1 ± 5.0	57.0	75.0	66.7	0.0
CADD. Test	354	All	354	65.0	65.1 ± 7.8	46.0	88.0	39.8	0.0

AD, Alzheimer's disease; Avg, average; CADD., CADDementia; CN, cognitively normal; Dset., dataset; Max, maximum; Med, median; MCI, mild cognitive impairment; Min, minimum; Std, standard deviation; Subjs., subjects; Train., training; Val., validation.

for Dataset 4, which is our largest, containing 23,165 volumes. Interestingly, only 151 images (0.7%) took longer than 25 min to process. Each process used two cores in a shared cluster of commodity hardware, such as Intel® Xeon® CPU E5645 at 2.40 GHz, and around 2 GB of RAM.

To train our CNNs, we used three different models of NVIDIA GPUs: GeForce GTX TITAN X (Maxwell microarchitecture), Tesla K40c, and Tesla K80. Training usually lasted for about 100 epochs, taking around 4 days to complete. We performed a total of 121 experiments. Inference time for our best network (VGG 512) was <1 s. The grid search for domain adaptation was completed in under one minute, and the classification of all 354 samples from CADDementia test set was accomplished in about 1 ms.

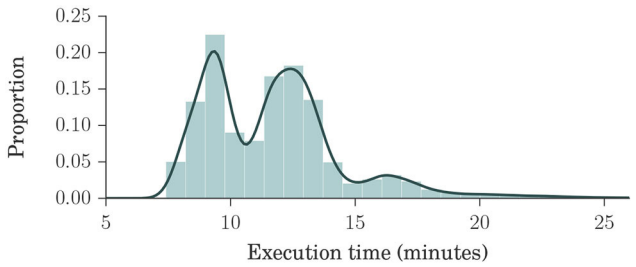


FIGURE 2 | Histogram and kernel density estimation plots of brain extraction and normalization times for Dataset 4, in minutes. Dataset 4 is our largest, composed of ADNI1, ADNIGO, ADNI2, AIBL, MIRIAD, and OASIS datasets, with a total of 23,165 volumes. These plots show that the processing times ranged mostly between 7 and 15 min, with an average of about 12 min, indicating that our method is fast.

TABLE 5 | Performance results (average true positive fraction, labeled avgTPF) of our best CNN architectures and respective configurations found in optimization experiments.

Architecture	avgTPF (%)	Norm	λ	Dropout (%)
LeNet-5	56.5	L2	10^{-2}	40
VGG 512	75.9	L2	10^{-4}	50
GoogLeNet	58.3	L1	10^{-3}	80
ResNet B	60.2	L2	10^{-2}	–

TABLE 6 | Multiple performance results of our best CNN, in percentage.

Model	Dataset	Split	Accuracy	TPF			AUC			
				CN	MCI	AD	All	CN	MCI	AD
ADNet	Dataset 1	Train.	60.6	89.6	36.7	86.8	87.9	90.3	80.6	88.8
		Val.	44.1	71.1	22.4	62.4	68.9	72.2	56.9	72.5
		Test	43.6	67.3	21.1	64.7	68.0	73.9	57.0	68.9
ADNet	CADD	Train.	76.7	83.3	55.6	88.9	90.3	92.1	83.1	96.3
		Test	51.4	77.5	27.9	46.6	68.5	70.5	61.2	73.6
ADNet -DA	CADD	Train.*	76.7	75.0	55.6	100.0	88.5	90.7	79.4	95.8
		Train.	90.0	83.3	88.9	100.0	98.0	95.8	97.9	100.0
		Test	52.3	68.2	37.7	49.5	70.9	72.8	60.5	79.0

AD, Alzheimer's disease; AUC, area under the receiver operating characteristic curve; CN, cognitively normal; CADD, CADDementia; CNN, convolutional neural network; MCI, mild cognitive impairment; Train., training; TPF, true-positive fraction; Val., validation; Train., leave-one-out cross-validation results.*

In summary, our method is expected to provide a response in <15 min, with extreme cases taking a little longer than 2 h. This processing time contrasts with the current best method in CADDementia challenge, which requires 19 h of computation (Sørensen et al., 2017). In other words, our method is nearly 10× faster, considering the worst-case scenario, or almost 80× faster, on average.

Regarding performance metrics in terms of results, we present our best configuration for each network architecture in **Table 5**. The best VGG had 512 units in each fully-connected layer, and the best ResNet used the projection shortcut (B). We also included our main optimization metric—average TPF (avgTPF)—for the training set of CADDementia, in which the top value was 75.9%, translating to 76.7% in accuracy. All of these results were found while optimizing the networks with Dataset 1.

As initially expected, L2 norm provided the best results for almost all architectures. The best GoogLeNet using L2 achieved 57.4% average TPF, close to the one using L1 (58.3%), while the L1 norm performed considerably worse for the other networks. ResNet with identity shortcuts (A) achieved 57.4%, which is slightly inferior to the projection shortcut (B), with 60.2%, a similar difference found in the original work (He et al., 2016). We hypothesize that deeper architectures did not achieve the highest scores because they tend to do better in more massive datasets, which we did not have.

A batch size of three (instead of nine) only produced significantly worse results, indicating that our best VGG model could potentially achieve even better results if we used GPUs with larger memory or a multi-GPU framework implementation. Similarly, multi-class hinge loss did not improve our results. Most surprisingly, Dataset 1, our smallest, presented the best performances, and Dataset 2 achieved an average TPF as high as 72.2%. We hypothesize that this happened due to the higher diversity of data sources and conditions in more massive sets, indicating that a smaller but more cohesive dataset should be sufficient for optimization.

VGG 512 was our best network model, and the respective performance metrics are shown in **Table 6**. We named our CNN approach ADNet (Alzheimer's Disease Network), with the domain adaptation method ADNet-DA, and submitted our prediction scores to the CADDementia challenge. Currently, there are 48 different submissions, ours included[3]. Similar to previous results (Esmaeilzadeh et al., 2018; Abrol et al., 2020), we also observed some overfitting in the training data. However, the performance differences in Dataset 1 between validation and test sets were small, indicating that we appropriately mitigated this problem.

In general, ADNet presented promising results in the CADDementia training set. The low TPF in the MCI group was expected, since classifying MCI subjects solely on their sMRI represents a challenge (Albert et al., 2011), also observed in a number results of CADDementia submissions (Bron et al., 2015). However, the decrease in MCI and AD TPFs between training and test sets was higher than expected. As such, this method achieved an interesting two-class specificity, with a modest two-class sensitivity, meaning it is better suited for determining healthy patients. Regarding accuracy in the test set, ADNet ranked 25th, tied with two other systems, outperforming 22 submissions. Besides, this result was only statistically different, with a 95% confidence interval from the first and the last three systems. Considering that we were the first group that did not use any domain-specific information for this task, we

[3]https://caddementia.grand-challenge.org/results_all/ [Online; accessed 2020-05-14].

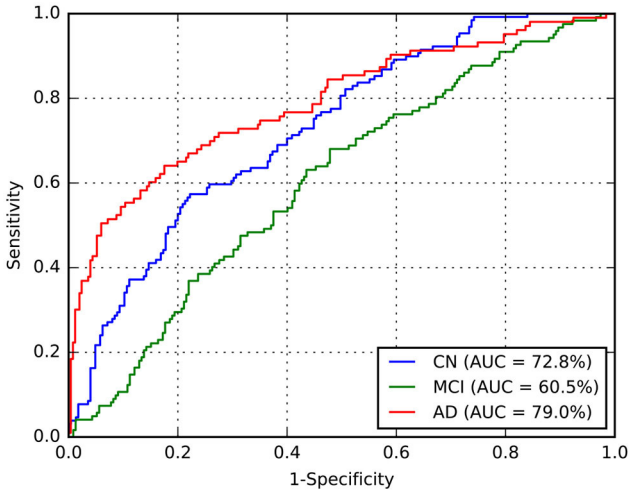


FIGURE 3 | Receiver operating characteristic (ROC) curve for ADNet-DA, provided by CADDementia. AD, Alzheimer's disease; AUC, area under the receiver operating characteristic curve; CN, cognitively normal; MCI, mild cognitive impairment.

can claim that our CNN method was able to learn meaningful patterns automatically.

As for the domain adaptation approach, we extracted 512 features from the second-to-last layer of ADNet, and then we performed a grid search on the parameters of a logistic regression classifier. Using the best parameters found (most importantly, $C = 0.001$), we optimized this classifier on the complete training set and applied it to output classification probabilities for each sample from the challenge. We also submitted these predictions to CADDementia, naming it ADNet-DA (ADNet with domain adaption). The corresponding results are also indicated in **Table 6**. This method ranked 21st, outperforming 27 submissions, with a statistical difference from the first and the last four systems.

Considering this approach, we reported the leave-one-out cross-validation results in the training set while performing a grid search, and also the results in this same set after the last optimization with all training samples. As expected, developing and evaluating a system on the same data overestimated its generalization performance; however, even our cross-validation attempt did not significantly improve our estimations for the test set. In comparison with ADNet, ADNet-DA improved both MCI and AD TPFs, while decreasing CN TPF, with an overall improvement of almost one percentage point in accuracy. These results indicate that domain adaptation was indeed an important technique. The corresponding ROC for CADDementia test set is displayed in **Figure 3**, and the respective confusion matrix is in **Table 7**.

Though Dolph et al. (2017) pioneered deep learning on this challenge, we are the first (to our knowledge) to propose an end-to-end training deep 3D CNN for the multiclass AD biomarker identification task in CADDementia. One of their systems ranked 7th, with 56.8% accuracy, while the other ranked 25th, tied with ADNet on 51.4%. Our ADNet-DA method was able to outperform a deep-learning system that uses domain-specific information, which demonstrates the effectiveness of the approach proposed in this work.

TABLE 7 | Confusion matrix (in percentage) for ADNet-DA, provided by CADDementia.

		Prediction		
		CN	**MCI**	**AD**
Actual	**CN**	68.2	25.6	6.2
	MCI	51.6	37.7	10.7
	AD	29.1	21.4	49.5

Values are adjusted relative to the actual class, i.e., divided by the row sum. This way, the main diagonal represents the true positive fraction (TPF) for each class. AD, Alzheimer's disease; CN, cognitively normal; MCI, mild cognitive impairment.

4.3. Accountability

Understanding the decision-making process of a machine-learning algorithm has become crucial lately, especially in medicine. For practical application, an algorithm must present good performance results and also demonstrate how predictions are generated. The explicability requirement has become even more critical in recent years with rules such as the General Data Protection Regulation (GDPR), which also brought explainable artificial intelligence (XAI) to the spotlight (Goodman and Flaxman, 2016).

Explaining what and how a neural network has learned is an open problem, with a rapidly evolving research field. In order to better understand what our model is analyzing in brain images and how it is done, we experimented with a number of visualization approaches, considering the most used techniques in accountable machine learning for neural networks. Some of these approaches were also recently explored by Rieke et al. (2018).

Similarly to Krizhevsky et al. (2012), we analyzed the filters from our first convolutional layer. While their kernels were of size $11 \times 11 \times 3$, presenting some interesting smooth and colorful patterns, our kernels are $3 \times 3 \times 3$ in grayscale, producing less than ideal images for visualization.

Another traditional approach for visualization is to show outputs of activation functions from the network, after processing an input. Activation is simply the result of a mathematical function. These outputs represent some of the initial patterns that the network learned to be the most relevant for this task. These outputs are then non-linearly combined with additional and more complex patterns before the final classification.

Occlusion is a technique to visualize how and where the input image affects the result of the network. The basic idea is to systematically hide (occlude) some regions of the input image, preventing the network from becoming activated in these specific regions, and then storing the probabilities output. Given a class of interest, for instance AD, it is possible to create a heatmap with the corresponding prediction for each occluded region, where most important regions will present highest impact (with low probability), due to the occlusion. This technique was initially proposed by Zeiler and Fergus (2014).

There are different ways to hide a region of the input image and avoid activations in a network. The most straightforward and most direct would be to set input values to their respective averages, which, in our case, is zero. Considering images in a range from zero to 255, it is possible to occlude with the average value (gray), with zero (black), with 255 (white), and even more sophisticated approaches, such as different forms of noise.

Finally, we investigated an approach that more closely related to the actual output decision of the network. For that, we calculated the gradient of the network concerning the input, which is used to update the network's internal parameters. These gradients may also be plotted and interpreted as how much the output is affected by changes in input values; however, this simple approach produces rather noisy visualizations. An improvement to this technique, called deconvolution, was proposed (Zeiler and Fergus, 2014) and can be interpreted as reversing the operations performed by the network. Even though this is an interesting approach, the guided backpropagation method (Springenberg et al., 2014) produces even sharper visualizations. Interestingly, guided backpropagation combines calculations from both backpropagation and deconvolution, resulting in more detailed images.

Figure 4 illustrates activated areas for each group, where brighter regions mean a larger effect on the prediction output. For the CN group, we can see activations distributed in a diffuse pattern, but mainly restricted to the cortex in the right temporal lobe (predominantly in the medial temporal gyrus and the parahippocampal gyrus), the central portion of the occipital lobe, the posterior cingulum, and the posterior parietal cortex. For MCI, activations occurred mainly in the left posterior parietal cortex, the right anterior cingulum, and the right dorsolateral prefrontal cortex. For AD, more significant activations were detected in the left posterior parietal cortex, right temporal pole, cerebellum, and more diffusively in the spherical surface of the brain.

The diffuse pattern of activations in all groups (mainly in temporal and posterior regions of the brain) can be interpreted in the context of neuroimaging findings in the field of AD. Although no single structure can differentiate AD patients from CN subjects, atrophy in temporal regions seems to be an inevitable process in the disease. The medial temporal lobe regions might be the first affected in the course of the disease, presenting very early signs of neurodegeneration (Karas et al., 2004), which correlate with clinical symptoms even in the prodromal stage, i.e., MCI (Frisoni et al., 2010). As in pathophysiological aspects, the temporal regions mainly present intracellular aggregates of hyperphosphorylated tau protein, which are associated with reduced gray matter density (Thomann et al., 2009). Another signature of AD, extracellular amyloid β-protein (Aβ) deposition in the form of plaques, is mainly observed in the midline regions (posterior cingulate and medial prefrontal cortices), and parietal areas. Longitudinal studies have shown that these areas not only atrophy at the mild stage of AD (Weiler et al., 2015), but they continue to degenerate at a rate of about 2–4% per year (Thompson et al., 2003; Leow et al., 2009). Thus, we were not surprised to find larger activations in those regions classically affected by the disease.

For our last visualization technique, our motivation was to understand how our data samples were spatially distributed within internal feature representations of our network, in order to determine whether these representations were really helpful to discriminate between each class. To plot our data from this high-dimensional space, we first projected them into two dimensions using the t-distributed stochastic neighbor embedding (t-SNE; Maaten and Hinton, 2008), with principal component analysis (PCA) initialization. Considering the outputs from a specific layer of our network, we generated an embedding with all training and test data in CADDementia, and then colored training samples according to each respective class. It is important to remark that this projection did not use label information from training data, which was used solely to color our plots.

First, we extracted features from the second-to-last layer of our network, traditionally used to transfer learning and domain adaptation, with 512 dimensions. Then, we considered the final layer from ADNet that outputs classification probabilities, with 3 dimensions, and the probability outputs from ADNet-DA. Resulting embeddings are present in **Figure 5**. Considering ADNet, even though t-SNE (Maaten and Hinton, 2008) did not use any label information, training data points were better grouped in an internal feature representation space rather than in the probability output space, indicating that the softmax classifier used in the network did not perform as well as it could. From these plots, we can also see that probabilities from ADNet-DA are better distributed in comparison with ADNet, especially for the AD group, while there was a smaller confusion for MCI.

5. CONCLUSIONS

Using data from ADNI, we optimized a 3D CNN with the whole brain image as input and achieved the best accuracy with a network architecture based on VGG (Simonyan and Zisserman, 2014). Our method, named ADNet, outperformed several other systems in the prior art. Additionally, our method with domain adaptation, called ADNet-DA, reached 52.3% in accuracy on the CADDementia challenge test set, outperforming most of the submissions to this challenge. Our approach is completely automatic (i.e., does not require additional information input and manual intervention), and is considerably fast (around 80 times faster than CADDementia winning methods).

Importantly, whereas all other submissions used prior information from the disease (e.g., hippocampal volume, demographic information), our method did not use any domain-specific knowledge from AD. For that reason, we believe it could be applied to other disorders that could benefit from the CAD system using sMRI as input data. We understand that our approach can be used to find meaningful patterns within data, corroborate previous findings by specialists, assist in diagnosis scenarios, and eventually help identify patterns for diseases other than AD. Our conclusions are supported by our explainable artificial intelligence (XAI) techniques, including accountability visualizations.

Future work could investigate XAI techniques to understand brain regions involved in the decision-making process, and cross-match highlighted regions with specialists knowledge, to see how

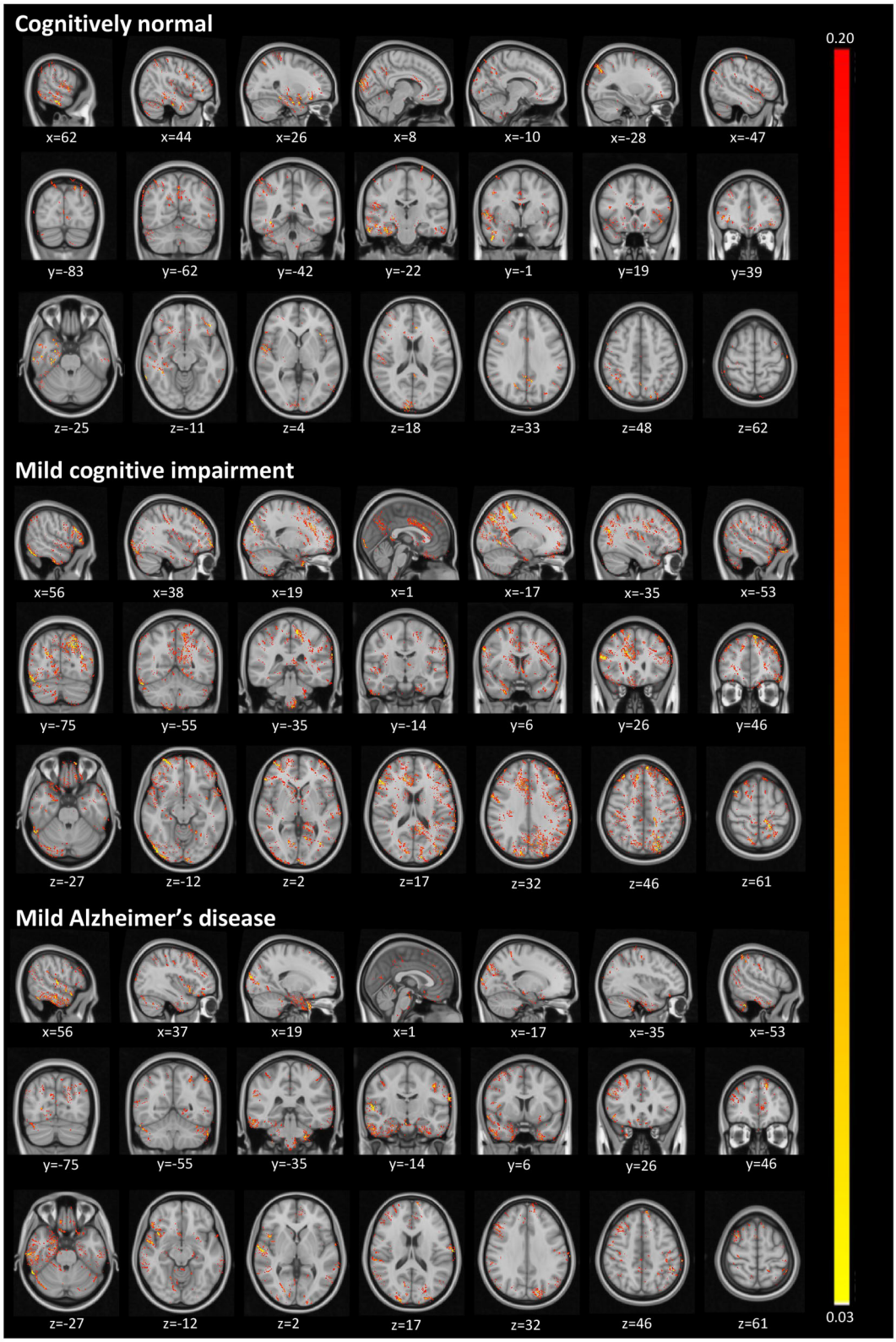


FIGURE 4 | Activated regions from the guided backpropagation method for each group. Activations are displayed in *hot* colormap overlaid onto the MNI template. *Hotter* regions mean a more significant effect on the prediction output. Colors only represent the relative importance of each voxel, having no direct meaning associated to their absolute values.

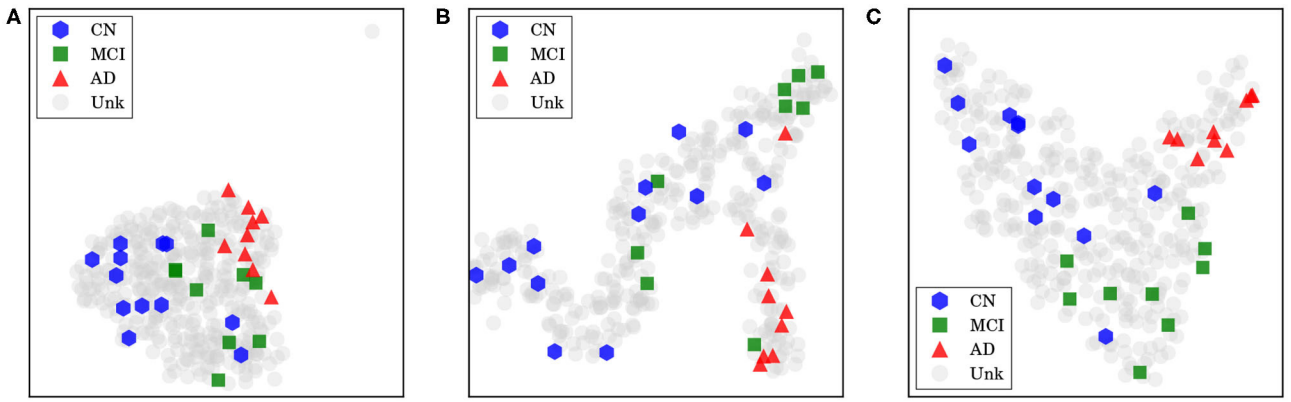


FIGURE 5 | Features and probabilities visualizations with 2D t-SNE projections. AD, Alzheimer's disease; CN, cognitively normal; MCI, mild cognitive impairment; Unk, unknown. **(A)** Visualization with 2D t-SNE projections using 512-dimensional features extracted from the second-to-last layer. **(B)** Visualization with 2D t-SNE projections using the 3-dimensional ADNet probabilities. **(C)** Visualization with 2D t-SNE projections using the 3-dimensional ADNet-DA probabilities.

one can complement the other in refining the technique. Finally, it would be interesting to incorporate patients' history data to enrich the information present in MRIs, to drive the decision process and to tie it with patients' backgrounds.

DATA AVAILABILITY STATEMENT

Generated datasets are a compilation of public data, which require specific data use agreements. Full information on accessing this data can be found in the acknowledgments of this article.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

AR and MW coordinated this research and helped in defining our method and experimental design. RC assisted with the neurological area and with the interpretation of our explainable AI visualizations. GF developed our method and dealing directly with the data and the experiments. RP contributed to the manuscript writing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES, DeepEyes Grant), by the Brazilian National Council for Scientific and Technological Development (CNPq, Grant #304497/2018-5), and by the Intramural Research Program of the National Institute on Aging, National Institutes of Health (NIA/NIH), through grant NIH/NIA/IRP #1-ZIA-AG000340, and Open access publication fees were partially supported by CAPES DeepEyes, CNPq 304497/2018-5, and FAPESP 2017/12646-3.

ACKNOWLEDGMENTS

We would like to thank Lucas David for reviewing early versions of this paper, as well as Alan Godoy and Recod Lab members for insightful discussions.

The authors appreciate the effort from researchers and patients in collecting and providing all the data used in our study. We additionally thank infrastructure support for running our experiments. Each group is explicitly acknowledged next, in alphabetical order.

This research is part of the first author's master's thesis, through the Institute of Computing, University of Campinas, Campinas, Brazil (Folego, 2018).

ADNI

Some data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds

to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (https://fnih.org/). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

CADDementia

Some data used in the preparation of this article were obtained from the CADDementia challenge (https://caddementia.grand-challenge.org/).

Microsoft Azure

Cloud computing resources were provided by a Microsoft Azure for Research award (https://www.microsoft.com/en-us/research/academic-program/microsoft-azure-for-research/).

MIRIAD

Some data used in the preparation of this article were obtained from the MIRIAD database (http://miriad.drc.ion.ucl.ac.uk). The MIRIAD investigators did not participate in analysis or writing of this report. The MIRIAD dataset is made available through the support of the UK Alzheimer's Society (https://www.alzheimers.org.uk/) (Grant RF116). The original data collection was funded through an unrestricted educational grant from GlaxoSmithKline (Grant 6GKC).

NVIDIA

A Tesla K40 used for this research was donated by the NVIDIA Corporation.

OASIS

Grant numbers P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, and R01 MH56584.

REFERENCES

Abrol, A., Bhattarai, M., Fedorov, A., Du, Y., Plis, S., and Calhoun, V. (2020). Deep residual learning for neuroimaging: an application to predict progression to Alzheimer's disease. *J. Neurosci. Methods* 339:108701. doi: 10.1016/j.jneumeth.2020.108701

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement*. 7, 270–279. doi: 10.1016/j.jalz.2011.03.008

Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., Ballas, N., et al. (2016). Theano: a Python framework for fast computation of mathematical expressions. *ArXiv e-prints*. *arXiv*:1605.02688.

Avants, B. B., Tustison, N., and Song, G. (2009). Advanced normalization tools (ANTs). *Insight J*. 2, 1–35. Available online at: http://hdl.handle.net/10380/3113

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* 54, 2033–2044. doi: 10.1016/j.neuroimage.2010.09.025

Beckett, L. A., Donohue, M. C., Wang, C., Aisen, P., Harvey, D. J., and Saito, N. (2015). The Alzheimer's disease neuroimaging initiative phase 2: increasing the length, breadth, and depth of our understanding. *Alzheimer's Dement*. 11, 823–831. doi: 10.1016/j.jalz.2015.05.004

Bron, E. E., Smits, M., van der Flier, W. M., Vrenken, H., Barkhof, F., Scheltens, P., et al. (2015). Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural mri: the CADDementia challenge. *NeuroImage* 111, 562–579. doi: 10.1016/j.neuroimage.2015.01.048

Collins, D. L., Zijdenbos, A. P., Baaré, W. F., and Evans, A. C. (1999). "Animal+insect: improved cortical structure segmentation," in *Information Processing in Medical Imaging 16th International Conference, IPMI'99*, eds A. Kuba, M. Šáamal, and A. Todd-Pokropek (Visegrád: Springer), 210–223. doi: 10.1007/3-540-48714-X_16

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn*. 20, 273–297. doi: 10.1007/BF00994018

Diakopoulos, N., Friedler, S., Arenas, M., Barocas, S., Hay, M., Howe, B., et al. (2017). *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*. FAT/ML. (Accessed May 14, 2020).

Dieleman, S., Schlüter, J., Raffel, C., Olson, E., Sønderby, S. K., Nouri, D., et al. (2015). *Lasagne: First Release*. Geneva: Zenodo. doi: 10.5281/zenodo.27878

Dolph, C. V., Alam, M., Shboul, Z., Samad, M. D., and Iftekharuddin, K. M. (2017). "Deep learning of texture and structural features for multiclass Alzheimer's disease classification," in *International Joint Conference on Neural Networks (IJCNN)* (Anchorage, AK), 2259–2266. doi: 10.1109/IJCNN.2017.7966129

Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., et al. (2009). The Australian imaging, biomarkers and lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int. Psychogeriatr*. 21, 672–687. doi: 10.1017/S1041610209009405

Esmaeilzadeh, S., Belivanis, D. I., Pohl, K. M., and Adeli, E. (2018). "End-to-end Alzheimer's disease diagnosis and biomarker identification," in *Machine Learning in Medical Imaging*, eds Y. Shi, H.-I. Suk, and M. Liu (Cham: Springer International Publishing), 337–345. doi: 10.1007/978-3-030-00919-9_39

Evans, D. A., Funkenstein, H. H., Albert, M. S., Scherr, P. A., Cook, N. R., Chown, M. J., et al. (1989). Prevalence of Alzheimer's disease in a community population of older persons: higher than previously reported. *JAMA* 262, 2551–2556. doi: 10.1001/jama.1989.03430180093036

Folego, G. (2018). *Adnet: computer-aided diagnosis for Alzheimer's disease using whole-brain 3D convolutional neural network* (Master's thesis). University of Campinas, Campinas, Brazil.

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., and Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage* 54, 313–327. doi: 10.1016/j.neuroimage.2010.07.033

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C., and Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage* 47(Suppl. 1):S102. doi: 10.1016/S1053-8119(09)70884-5

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., and Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol*. 6, 67–77. doi: 10.1038/nrneurol.2009.215

Glorot, X., and Bengio, Y. (2010). "Understanding the difficulty of training deep feedforward neural networks," in *13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, eds Y. W. Teh and M. Titterington (Sardinia: PMLR), 249–256

Goodman, B., and Flaxman, S. (2016). European Union regulations on algorithmic decision-making and a "right to explanation". *ArXiv e-prints*. doi: 10.1609/aimag.v38i3.2741

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV), 770–778. doi: 10.1109/CVPR.2016.90

Hosseini-Asl, E., Ghazal, M., Mahmoud, A., Aslantas, A., Shalaby, A. M., Casanova, M. F., et al. (2018). Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Front. Biosci*. 23, 584–596. doi: 10.2741/4606

Ioffe, S., and Szegedy, C. (2015). "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *The 32nd International Conference on Machine Learning (ICML 2015)*, eds D. Blei and F. Bach (Lille), 448–456.

Jack, C. R. Jr., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27, 685–691. doi: 10.1002/jmri.21049

Karas, G., Scheltens, P., Rombouts, S., Visser, P., van Schijndel, R., Fox, N., et al. (2004). Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. *NeuroImage* 23, 708–716. doi: 10.1016/j.neuroimage.2004.07.006

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv*. *arXiv:*1412.6980.

Korolev, S., Safiullin, A., Belyaev, M., and Dodonova, Y. (2017). "Residual and plain convolutional neural networks for 3D brain MRI classification," in *IEEE International Symposium on Biomedical Imaging* (Melbourne, VIC), 835–838. doi: 10.1109/ISBI.2017.7950647

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* 25, eds F. Pereira, C. J. Burges, L. Bottou, and K. Q. Weinberger (Stateline; Lake Tahoe, NV: Curran Associates, Inc.), 1097–1105.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., et al. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comput*. 1, 541–551. doi: 10.1162/neco.1989.1.4.541

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Leow, A. D., Yanovsky, I., Parikshak, N., Hua, X., Lee, S., Toga, A. W., et al. (2009). Alzheimer's disease neuroimaging initiative: a one-year follow up study using tensor-based morphometry correlating degenerative rates, biomarkers and cognition. *NeuroImage* 45, 645–655. doi: 10.1016/j.neuroimage.2009.01.004

Lin, C., Watson, R., Ward, H., Rydberg, C., Witte, R., and Bernstein, M. (2006). MP-RAGE compared to 3D IR SPGR for optimal T1 contrast and image quality in the brain at 3T. *Int. Soc. Magn. Reson. Med*. 14:981. Available online at: https://cds.ismrm.org/ismrm-2006/files/00981.pdf

Maaten, L. V. D., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res*. 9, 2579–2605. Available online at: https://www.jmlr.org/papers/v9/vandermaaten08a.html

Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., et al. (2009). Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology* 51, 73–83. doi: 10.1007/s00234-008-0463-x

Malone, I. B., Cash, D., Ridgway, G. R., MacManus, D. G., Ourselin, S., Fox, N. C., et al. (2013). MIRIAD-Public release of a multiple time point Alzheimer's MR imaging dataset. *NeuroImage* 70, 33–36. doi: 10.1016/j.neuroimage.2012.12.044

Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2009). Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. *J. Cogn. Neurosci*. 22, 2677–2684. doi: 10.1162/jocn.2009.21407

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2007). Open access series of imaging studies (oasis): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci*. 19, 1498–1507. doi: 10.1162/jocn.2007.19.9.1498

McCullagh, P. (1984). Generalized linear models. *Eur. J. Oper. Res*. 16, 285–292. doi: 10.1016/0377-2217(84)90282-0

Mehmood, A., Maqsood, M., Bashir, M., and Shuyuan, Y. (2020). A deep Siamese convolution neural network for multi-class classification of Alzheimer disease. *Brain Sci*. 10:84. doi: 10.3390/brainsci10020084

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., et al. (2005). Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's disease neuroimaging initiative (ADNI). *Alzheimer's Dement*. 1, 55–66. doi: 10.1016/j.jalz.2005.06.003

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted Boltzmann machines," in *The 27th International Conference on Machine Learning (ICML 2010)*, eds J. Fürnkranz and T. Joachims (Haifa: Omnipress), 807–814.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res*. 12, 2825–2830.

Rieke, J., Eitel, F., Weygandt, M., Haynes, J.-D., and Ritter, K. (2018). Visualizing convolutional networks for MRI-based diagnosis of Alzheimer's disease. *ArXiv e-prints*. doi: 10.1007/978-3-030-02628-8_3

Sharif Razavian, A., Azizpour, H., Sullivan, J., and Carlsson, S. (2014). "CNN features off-the-shelf: an astounding baseline for recognition," in *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Columbus, OH), 806–813. doi: 10.1109/CVPRW.2014.131

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv*. *arXiv*:1409.1556

Sørensen, L., Igel, C., Pai, A., Balas, I., Anker, C., Lillholm, M., et al. (2017). Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry. *NeuroImage* 13, 470–482. doi: 10.1016/j.nicl.2016.11.025

Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. (2014). Striving for simplicity: the all convolutional net. *arXiv*. *arXiv*:1412.6806.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, MA), 1–9. doi: 10.1109/CVPR.2015.7298594

Thomann, P. A., Kaiser, E., Schönknecht, P., Pantel, J., Essig, M., and Schröder, J. (2009). Association of total tau and phosphorylated tau 181 protein levels in cerebrospinal fluid with cerebral atrophy in mild cognitive impairment and Alzheimer disease. *J. Psychiatry Neurosci*. 34, 136–142.

Thompson, P. M., Hayashi, K. M., de Zubicaray, G., Janke, A. L., Rose, S. E., Semple, J., et al. (2003). Dynamics of gray matter loss in Alzheimer's disease. *J. Neurosci*. 23, 994–1005. doi: 10.1523/JNEUROSCI.23-03-00994.2003

Tustison, N., and Avants, B. (2013). Explicit b-spline regularization in diffeomorphic image registration. *Front. Neuroinform*. 7:39. doi: 10.3389/fninf.2013.00039

van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation. *Comput. Sci. Eng*. 13, 22–30. doi: 10.1109/MCSE.2011.37

Wachinger, C., and Reuter, M. (2016). Domain adaptation for Alzheimer's disease diagnostics. *NeuroImage* 139, 470–479. doi: 10.1016/j.neuroimage.2016.05.053

Weiler, M., Agosta, F., Canu, E., Copetti, M., Magnani, G., Marcone, A., et al. (2015). Following the spreading of brain structural changes in Alzheimer's disease: a longitudinal, multimodal MRI study. *J. Alzheimer's Dis*. 47, 995–1007. doi: 10.3233/JAD-150196

Welford, B. P. (1962). Note on a method for calculating corrected sums of squares and products. *Technometrics* 4, 419–420. doi: 10.1080/00401706.1962.10490022

Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., et al. (2020). Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. *Med. Image Anal*. 2020:101694. doi: 10.1016/j.media.2020.101694

Wyman, B. T., Harvey, D. J., Crawford, K., Bernstein, M. A., Carmichael, O., Cole, P. E., et al. (2013). Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimer's Dement*. 9, 332–337. doi: 10.1016/j.jalz.2012.06.004

Zeiler, M. D., and Fergus, R. (2014). "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision (ECCV)* (Zurich: Springer), 818–833. doi: 10.1007/978-3-319-10590-1_53

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Citation: Folego G, Weiler M, Casseb RF, Pires R and Rocha A (2020) Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI. Front. Bioeng. Biotechnol. 8:534592. doi: 10.3389/fbioe.2020.534592

frontiers in Oncology

ORIGINAL RESEARCH
published: 20 April 2021
doi: 10.3389/fonc.2021.622827

A Novel Model Based on Deep Convolutional Neural Network Improves Diagnostic Accuracy of Intramucosal Gastric Cancer (With Video)

Dehua Tang[1†], Jie Zhou[2,3,4†], Lei Wang[1†], Muhan Ni[1], Min Chen[1], Shahzeb Hassan[5], Renquan Luo[2,3,4], Xi Chen[2,3,4], Xinqi He[2,3,4], Lihui Zhang[2,3,4], Xiwei Ding[1], Honggang Yu[2,3,4], Guifang Xu[1*] and Xiaoping Zou[1*]*

[1] Department of Gastroenterology, Nanjing Drum Tower Hospital, Affiliated Drum Tower Hospital, Medical School of Nanjing University, Nanjing, China, [2] Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, China, [3] Key Laboratory of Hubei Province for Digestive System Disease, Renmin Hospital of Wuhan University, Wuhan, China, [4] Hubei Provincial Clinical Research Center for Digestive Disease Minimally Invasive Incision, Renmin Hospital of Wuhan University, Wuhan, China, [5] Northwestern University Feinberg School of Medicine, Chicago, IL, United States

Background and Aims: Prediction of intramucosal gastric cancer (GC) is a big challenge. It is not clear whether artificial intelligence could assist endoscopists in the diagnosis.

Methods: A deep convolutional neural networks (DCNN) model was developed *via* retrospectively collected 3407 endoscopic images from 666 gastric cancer patients from two Endoscopy Centers (training dataset). The DCNN model's performance was tested with 228 images from 62 independent patients (testing dataset). The endoscopists evaluated the image and video testing dataset with or without the DCNN model's assistance, respectively. Endoscopists' diagnostic performance was compared with or without the DCNN model's assistance and investigated the effects of assistance using correlations and linear regression analyses.

Results: The DCNN model discriminated intramucosal GC from advanced GC with an AUC of 0.942 (95% CI, 0.915–0.970), a sensitivity of 90.5% (95% CI, 84.1%–95.4%), and a specificity of 85.3% (95% CI, 77.1%–90.9%) in the testing dataset. The diagnostic performance of novice endoscopists was comparable to those of expert endoscopists with the DCNN model's assistance (accuracy: 84.6% *vs.* 85.5%, sensitivity: 85.7% *vs.* 87.4%, specificity: 83.3% *vs.* 83.0%). The mean pairwise kappa value of endoscopists was increased significantly with the DCNN model's assistance (0.430–0.629 *vs.* 0.660–0.861). The diagnostic duration reduced considerably with the assistance of the DCNN model from 4.35s to 3.01s. The correlation between the perseverance of effort and

OPEN ACCESS

Edited by:
Francesco Rundo, STMicroelectronics, Italy

Reviewed by:
Kun Zheng, Peking Union Medical College Hospital (CAMS), China
Xiao-Bo Li, Shanghai JiaoTong University, China

***Correspondence:**
Xiaoping Zou
zouxp@nju.edu.cn
Honggang Yu
yuhonggang@whu.edu.cn
Guifang Xu
13852293376@163.com

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *29 October 2020*
Accepted: *23 March 2021*
Published: *20 April 2021*

Citation:
Tang D, Zhou J, Wang L, Ni M, Chen M, Hassan S, Luo R, Chen X, He X, Zhang L, Ding X, Yu H, Xu G and Zou X (2021) A Novel Model Based on Deep Convolutional Neural Network Improves Diagnostic Accuracy of Intramucosal Gastric Cancer (With Video). Front. Oncol. 11:622827. doi: 10.3389/fonc.2021.622827

diagnostic accuracy of endoscopists was diminished using the DCNN model (r: 0.470 *vs.* 0.076).

Conclusions: An AI-assisted system was established and found useful for novice endoscopists to achieve comparable diagnostic performance with experts.

Keywords: artificial intelligence, deep convolutional neural network, depth of invasion, gastric cancer, endoscopic resection

INTRODUCTION

Gastric cancer (GC) patients are mostly diagnosed at an advanced stage and are ineligible for curative resection, making it the third leading cause of cancer deaths worldwide (1). But if GC can be diagnosed and then curatively resected at an early stage, the 5-year survival rate of this malignancy exceeds 95% (2). Various studies have validated that endoscopic submucosal dissection (ESD) can be available to treat early gastric cancer (3–6). According to the Japanese Gastric Cancer Treatment Guidelines 2018, the absolute indications for ESD of early gastric cancer include differentiated intramucosal cancer without ulceration and differentiated intramucosal cancer with ulceration and tumor size of ≤ 3 cm (7). Previous studies demonstrated that the incidence of lymph node metastasis (LNM) of these intramucosal gastric cancer lesions is negligible (8, 9). Therefore, it is of great essence to determine whether there is deep submucosal invasion before gastric ESD. However, it remains a challenge to distinguish intramucosal gastric cancer lesions from submucosal lesions correctly.

In clinical practice, invasion depth of gastric cancer is often determined by assessing the macroscopic features using conventional white-light imaging (C-WLI) endoscopy or evaluating the linings and walls using endoscopic ultrasonography (EUS). However, various studies have demonstrated that the diagnostic performance of macroscopic features using C-WLI and linings and walls with EUS in invasion depth was comparable, with a limited accuracy of only 70–85% (10, 11). More than 15% of gastric cancer lesions have been underestimated or overestimated using both methodologies. Although enhanced imaging technologies like magnifying endoscopy (ME), narrow-band imaging (NBI), and blue laser imaging (BLI) have also been employed in the determination of intramucosal GC, the clinical value of these techniques largely depends on the experience of operators (12, 13). Moreover, the accuracy and concordance of all the methodologies were reported to vary wildly in different studies, even amongst the expertized endoscopists (10, 14, 15). Therefore, it would be very advantageous to develop efficient assistance tools to help endoscopists make robust, reproducible, and accurate diagnoses of intramucosal GC under C-WLI.

With recent technological advances, artificial intelligence (AI) has shown excellent efficacy in analyzing medical images (16). Several preclinical studies reported that AI could be used with high accuracy for detection, localization, and classification of GC (17–19). Three preliminary studies have applied AI to predict the invasion depth of GC with acceptable specificity or sensitivity (18, 20, 21). However, these studies only focused on evaluating AI's performance in predicting invasion depth instead of verifying AI's assistance in helping endoscopists make the final diagnosis. The latter is even more important than the former since endoscopists are required to make the final diagnosis due to safety and accountability.

This study aimed to develop an AI-assisted diagnostic model based on the deep convolutional neural networks (DCNN) to detect intramucosal GC from advanced lesions in real-time. We then evaluated the accuracy, concordance, and diagnostic duration of the DCNN model's assistance in helping endoscopists establish the final diagnosis.

METHODS

Study Design

This retrospective comparative study was performed at two institutions in China: Endoscopy Center of Nanjing University Medical School Affiliated Drum Tower Hospital (NJDTH) and Endoscopy Center of Renmin Hospital of Wuhan University (RHWU). We first trained our DCNN model to distinguish intramucosal gastric cancer lesions from submucosal lesions. Then, we assessed the performance of DCNN and evaluated the performance of endoscopists before (Test 1) and after referring to the DCNN-processed results (Test 2) with endoscopic images and videos. The study design was reviewed and approved by the Medical Ethics Committee at each institution (NJDTH, IRB no. 2020-026-01; RHWU, WDRY2019-K091). Informed consent was waived given the use of only retrospectively deidentified endoscopic images.

Data Preparation and Image Quality Control

A total of 870 patients who underwent endoscopic submucosal dissection (ESD) or gastrectomy with histologically proven malignancies (700 patients from NJDTH and 170 patients from RHWU) between Jan 2017 and June 2019 were retrospectively included in this study. After excluding patients with multiple synchronous lesions, gastric stump cancer, and

Abbreviations: AI, Artificial intelligence; WLI, White light imaging; NBI, Narrow band imaging; PPV, Positive predictive value; NPV, Negative predictive value; ESD, Endoscopic submucosal dissection; DCNN, Deep convolutional neural network; GC, Gastric cancer; LNM, Lymph node metastasis; EUS, Endoscopic ultrasound; ME, Magnifying endoscopy; NBI, Narrow-band imaging; BLI, Blue laser imaging; C-WLI, Conventional white-light imaging.

missing data, 3829 endoscopic images from 728 patients were obtained retrospectively from the imaging database of the two hospitals. A total of 194 endoscopic images were excluded from the study due to low quality (e.g., less insufflation of air, halation, defocus, blurs, bubble, sliding, fuzzy, and bleeding). The rest of the 3635 endoscopic images (from 728 patients) were used to develop and validate the AI model (**Table S1**). Moreover, 54 videos with single GC lesions of another 54 patients from NJDTH between Jan 2019 and June 2019, which were independent of 700 patients, were retrospectively collected in this study and used to test the AI's performance model and endoscopists. All the endoscopic images and videos were recorded with Olympus endoscopes (GIF-H260, GIF-H260Z, GIF-HQ290, GIF-H290Z, Olympus Medical Systems, Co., Ltd., Tokyo, Japan) with video processors (EVIS LUCERA CV260/CLV260SL, EVIS LUCERA ELITE CV290/CLV290SL, Olympus Medical Systems, Co., Ltd., Tokyo, Japan).

Two board-certificated pathologists determined the invasion depth of GC according to WHO Classification of Tumors 5th edition in cooperation. We defined D0 as a tumor invasion depth restricted to the mucosa and defined D1 as a tumor invasion depth deeper than mucosa. All the selected images were categorized into D0 (1924 images from 458 patients) and D1 (1711 images from 270 patients) based on the pathologic diagnosis of the resected tissues. These images were then labeled with D0 or D1 and marked with rectangular frames on the lesions by five experienced endoscopists from NJDTH (each of whom had more than 5 years of experience and had performed at least 5000 endoscopic examinations). For the D0 lesions, the whole area of the lesion was marked. But for the D1 lesions, only the region, based on pathological results that potentially invaded deeper than mucosa, was marked. The image marks were finalized only when more than four endoscopists reached a consensus to avoid individual bias. A total of 54 videos that lasted for 10s each were classified into the intramucosal category (M) and the submucosal category (SM) based on the final pathological results.

The whole dataset (3635 images from 728 patients) was divided into training and testing datasets, using random sampling based on patients. The training and testing datasets were as follows: 1) Training dataset: D0: 1798 images from 421 patients, D1: 1609 images from 245 patients, between Jan 2017 and June 2019; 2) Image testing dataset: D0: 126 images from 37 patients, D1: 102 images from 25 patients, between Jan 2017 and June 2019; and 3) Video testing dataset: M: 44 videos of intramucosal lesions from 44 patients, SM: 10 videos of submucosal lesions from 10 patients (**Figure S1**).

Development and Validation of DCNN Model

In this study, an architecture called Resnet-50 was employed to learn the features of the endoscopy images (22). For most DCNN frameworks, the network layers and the learning ability of the whole network are limited. This limitation is called the Vanishing Gradient problem of DCNN. The Shotcut connection structure enables the DCNN framework to contain more layers, thus effectively alleviating the Vanishing Gradient problem of DCNN. Resnet-50 is a classical framework and most widely employed in the Resnet family to solve complex image classification tasks (**Figure S2**). During the DCNN training process, the parameters of the neurons in the network were initially set to random values. For each input annotated image, the output was computed by the DCNN and compared with the annotation. The parameters of this mathematical function were then modified slightly to decrease the error of the output. The same process was then repeated multiple times for every image in the training set.

Evaluation of DCNN Model and Comparing With the Endoscopists

Firstly, we evaluated our DCNN model's performance to diagnose intramucosal GC in the testing datasets described above. Then, 20 endoscopists participated in the following assessment in two groups: (1) novices: 14 novice endoscopists with less than 2 years of endoscopic experience and no more than 3,000 endoscopic examinations; (2) experts: 6 experienced endoscopists with more than 10 years of endoscopic expertise and at least 8,000 endoscopic examinations (acknowledgments: YW, HMG, TY, 7NNZ; co-authors: MC, GFX). None of the endoscopists participated in the selection and labeling of the image datasets. Two-stage tests were conducted to further evaluate the DCNN model's assistance with the image and video testing datasets in our testing platform (**Figure S3**). The testing images and videos were all anonymized and randomly mixed before the assessments of endoscopists. For testing 1, each endoscopist was asked to diagnose the testing images and videos independently. A week later, these endoscopists conducted testing independently with the presentation of the DCNN-processed diagnosis. After testing 2, a Grit scale was used to assess the individual personality characteristics with 12 items. These items can be divided into two parts: consistency of interest and perseverance of effort. Each item was scored on a 5-point scale (from 1 to 5). The final score was the summed score divided by 12. Grit scale tests were conducted with a free platform (*Document Star*, https://www.wjx.cn).

Statistical Analyses

The primary outcome of this study was to evaluate the assistance of AI in improving the diagnostic performance of endoscopists. The area under the ROC curve (AUC) was calculated to assess the diagnostic ability of the DCNN model and endoscopists. The diagnostic performance of endoscopists with or without the DCNN model's assistance was evaluated and compared with the McNemar test. The diagnostic time was analyzed with Wilcoxon rank tests between groups with or without the DCNN model's assistance. The Grit scale scores were analyzed using correlations and linear regression analyses. For all the tests mentioned, a *p-value* of 0.05 was regarded as statistically significant. All statistical analysis and plotting were conducted with R software (version 4.0.2, R Foundation for Statistical Computing, Vienna, Austria) in R studio (version 1.3.959, R Studio Co., Boston, MA, USA).

RESULTS

Performance of DCNN Model in Image Testing Dataset

In the testing dataset, the DCNN model could make a diagnosis of D0 from D1 with an AUC of 0.942 [95% confidence interval (CI), 0.915–0.970], a sensitivity of 90.5% (95% CI, 84.1%–95.4%), and a specificity of 85.3% (95% CI, 77.1%–90.9%) (**Table 1** and **Figure 1**). The overall accuracy of our DCNN model was 88.2% (95% CI, 83.3%–91.7%), with a positive predictive value of 88.37% (95% CI, 81.7%–92.8%) and a negative predictive value of 87.88% (95% CI, 81.7%–92.8%) (**Table 1**).

Performance of Endoscopists Without AI Assistance (Test1) and With AI Assistance (Test2) in Image Testing Dataset

In test 1, the diagnostic performance of the DCNN model was better compared with those of endoscopists in both novice and expert groups (**Figure 2A**). All the endoscopists involved in this study exhibited a lower diagnostic accuracy than the DCNN model (69.7%–82.1% *vs.* 88.2%, P < 0.05) (**Table 2**). For diagnostic concordance, the mean pairwise kappa of the DCNN model was 0.527 (**Figure S4C**). The mean pairwise kappa value of endoscopists varied from 0.430 to 0.629 (**Table S2**).

In test 2, the performance of endoscopists was improved significantly with the DCNN model's assistance (**Figure 2B**). However, the increase of diagnostic accuracy varied between groups (**Table 2**). All the fourteen novice endoscopists showed significantly increased diagnostic accuracy with the use of the DCNN model (69.7%–77.2% *vs.* 80.7%–87.3%, P < 0.05), while two of six expert endoscopists yielded statistically improved accuracy with the assistance (P < 0.05). Notably, none of the enrolled endoscopists achieved higher accuracy than the DCNN model (**Figure 2C**). The diagnostic accuracy of novice endoscopists was significantly lower than that of expert endoscopists without the DCNN model's assistance (P < 0.01) (**Figure 2C**). Using the DCNN model, the accuracy of the novice group was comparable to that of the expert group (**Figure 2C**, P = 0.95). For sensitivity, three novices and two experts achieved significantly higher sensitivity with the DCNN model's assistance (**Table S3** and **Figure S4A**). For specificity, 10 novices and two experts showed significantly increased specificity using the DCNN model (**Table S3** and **Figure S4B**). For expert group, the diagnostic accuracy, sensitivity and specificity were increased significantly with the DCNN model's assistance (accuracy, 79.8% vs 85.5%, P < 0.001, **Figure 2C**; sensitivity, 84.3% vs 87.4%, P = 0.018; specificity, 74.2% vs 83.0%, P < 0.001; **Table S4**). For the novice group, the diagnostic accuracy, sensitivity and specificity also were elevated remarkably with the DCNN model's assistance (accuracy, 74.0%

TABLE 1 | Performance of the DCNN Model for Diagnosis of Gastric Mucosal Cancer.

	Accuracy, n (%)	Sensitivity, n (%)	Specificity, n (%)	Positive predictive value, n (%)	Negative predictive value, n (%)	Diagnostic time (s)
DCNN-model	88.16 (201/228)	90.48 (114/126)	85.29 (87/102)	88.37 (114/129)	87.88 (87/99)	0.15

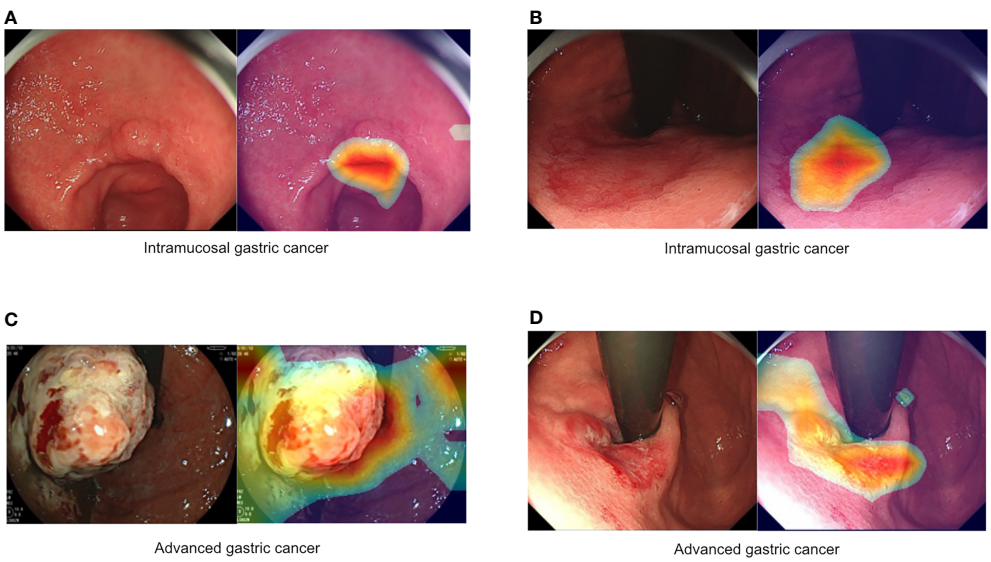


FIGURE 1 | Representative images of intramucosal and advanced gastric cancer. **(A, B)** Intramucosal gastric cancer, original c-WLI (left), and visual representation of the heatmap (right). **(C, D)** Advanced gastric cancer, original c-WLI (left), and visual representation of heatmap (right).

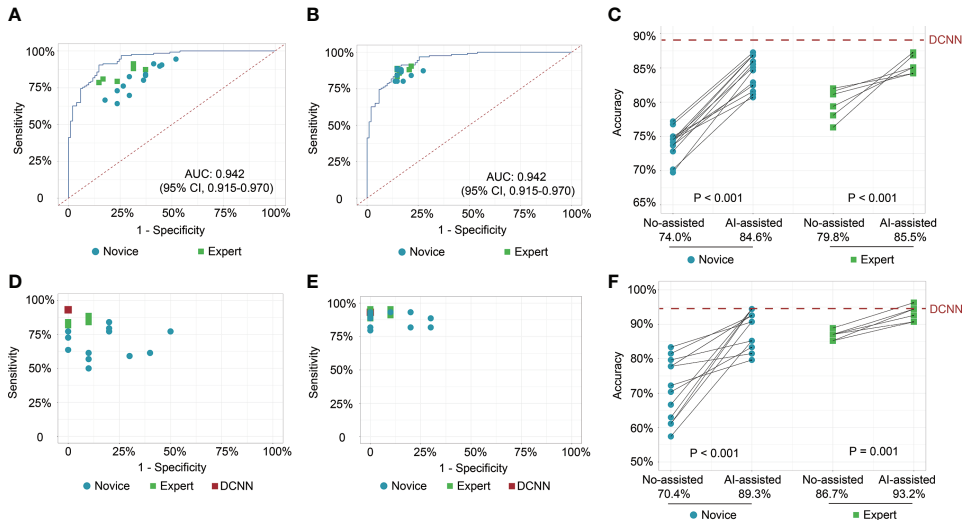


FIGURE 2 | Receiver operating characteristic curves and scatter plots illustrate the DCNN model's ability and endoscopists in discriminating intramucosal GC. **(A)** Diagnostic performance of DCNN model and endoscopists without the assistance of DCNN model in the image testing datasets; **(B)** Diagnostic performance of DCNN model and endoscopists with the assistance of DCNN model in the image testing datasets; **(C)** Diagnostic accuracy of endoscopists in the subgroup with or without the assistance of DCNN model in the image testing datasets; **(D)** Diagnostic performance of DCNN model and endoscopists without the assistance of DCNN model in the video testing datasets; **(E)** Diagnostic performance of DCNN model and endoscopists with the assistance of DCNN model in the video testing datasets; **(F)** Diagnostic accuracy of endoscopists in the subgroup with or without the assistance of DCNN model in the video testing datasets.

TABLE 2 | Diagnostic Accuracy of Endoscopists with or without the Assistance of DCNN Model.

Endoscopists	No-assistance (Test1)			AI-assistance (Test2)			Test1 *vs.* Test2
	Accuracy			Accuracy			
	n	percent	95% CI	n	percent	95% CI	
Novice (N=14)							
1	166/228	72.8	(66.9–78.7)	195/228	85.5	(81.6–89.4)	< 0.001
2	169/228	74.1	(68.2–80.0)	198/228	86.8	(82.9–90.7)	< 0.001
3	170/228	74.6	(68.7–80.5)	194/228	85.1	(81.2–89.0)	0.001
4	176/228	77.2	(71.3–83.1)	199/228	87.3	(83.4–91.2)	0.001
5	168/228	73.7	(67.8–79.6)	189/228	82.9	(79.0–86.8)	0.007
6	159/228	69.7	(63.8–75.6)	193/228	84.6	(80.8–88.6)	< 0.001
7	170/228	74.6	(68.7–80.5)	198/228	86.8	(82.9–90.7)	< 0.001
8	175/228	76.8	(70.9–82.7)	198/228	86.8	(82.9–90.7)	0.004
9	171/228	75.0	(69.1–80.9)	186/228	81.6	(75.7–87.5)	0.015
10	170/228	74.6	(68.7–80.5)	198/228	86.8	(82.9–90.7)	< 0.001
11	170/228	74.6	(68.7–80.5)	188/228	82.5	(76.6–88.4)	0.021
12	168/228	73.7	(67.8–79.6)	196/228	86.0	(82.1–89.9)	< 0.001
13	170/228	74.6	(68.7–80.5)	184/228	80.7	(74.8–86.6)	0.022
14	160/228	70.2	(64.3–76.1)	185/228	81.1	(75.2–87.0)	< 0.001
Expert endoscopists (N=6)							
1	186/228	81.6	(75.7–87.5)	194/228	85.1	(81.2–89.0)	0.153
2	181/228	79.4	(73.5–85.3)	194/228	85.1	(73.9–85.7)	0.061
3	187/228	82.0	(76.2–88.0)	192/228	84.2	(80.3–88.1)	0.473
4	174/228	76.3	(70.4–82.2)	198/228	86.8	(82.9–90.7)	< 0.001
5	185/228	81.1	(75.2–87.0)	192/228	84.2	(80.3–88.1)	0.248
6	178/228	78.1	(72.2–84.0)	199/228	87.3	(83.4–91.2)	0.005

95% CI, 95% confidence interval.

vs 84.6%, P < 0.001, **Figure 2C**; sensitivity, 81.1% vs 85.7%, P = 0.018; specificity, 65.2% vs 83.3%, P < 0.001; **Table S4**). As to concordance, the mean pairwise kappa of the DCNN model was 0.861 (**Figure S4D**). The mean pairwise kappa value of endoscopists increased significantly using the DCNN model and varied from 0.660 to 0.861 (**Table S5**).

The diagnostic time of the DCNN model was 0.15 seconds per image, which was much shorter than those of endoscopists (**Table 3**).

TABLE 3 | Diagnostic time of Endoscopists with or without the Assistance of AI.

Diagnostic time (s)	No-assistance (Test1)	AI-assistance (Test2)	P-value
DCNN model	0.15	0.15	–
Overall	4.35 ± 3.02	3.01 ± 1.66	0.03
Novice	5.09 ± 3.33	3.12 ± 1.90	0.02
Expert	2.62 ± 0.77	2.76 ± 0.99	0.64

With the DCNN model's assistance, the overall diagnostic time of endoscopists shortened significantly (4.35 *vs.* 3.01, P = 0.03). Notably, the diagnostic time of endoscopists was reduced statistically in the novice group (5.09 *vs.* 3.12, P = 0.02) with the DCNN model's assistance. However, the diagnostic time of experts was marginally increased with the DCNN model (2.62 *vs.* 2.76, P =0.64).

Performance of Endoscopists Without AI Assistance (Test1) and With AI Assistance (Test2) in Video Testing Dataset

To further explore the assistance of the DCNN model in a real-time clinical setting, we evaluated the performance of endoscopists with or without the DCNN model's assistance with 54 endoscopic videos (**Figures 2D, E**). The DCNN model showed a better performance in the video datasets with a sensitivity of 93.2%, a specificity of 100.0%, and an accuracy of 94.4% (**Table S6**). For expert endoscopists, the diagnostic accuracy and sensitivity increased significantly with the assistance of the DCNN model (accuracy, 86.7% vs 93.2%, P = 0.001, **Figure 2F**; sensitivity, 85.2% vs 92.4%, P = 0.002, **Table S6**). But the specificity showed marginal improvement (93.3% *vs.* 96.7%, P = 0.617, **Table S6**). For novice endoscopists, the diagnostic accuracy, sensitivity and specificity increased remarkably with the assistance of the DCNN model (accuracy, 70.4% vs 89.3%, P < 0.001, **Figure 2F**; sensitivity, 67.7% vs 88.6%, P < 0.001; specificity, 82.1% vs 92.1%, P = 0.008; **Table S6**).

Personality Traits and Performance of Endoscopists

Grit scale reflects the ability of individuals to maintain focus (consistency of interest) and persevering for long-term goals (perseverance of effort) (23). The correlation between the personality traits and the diagnostic accuracy was analyzed with or without the DCNN model's assistance. As is shown here, the correlation between grit score and diagnostic accuracy was marginal with or without the assistance of DCNN (r = 0.178, P = 0.452 *vs.* r = 0.145, P = 0.541, **Table 4** and **Figure S5**). The correlation between the scores for the consistency of interest and the diagnostic accuracy was also not significant with or without the assistance of DCNN (r = -0.122, P = 0.609 *vs.* r = 0.145, P = 0.541, **Table 4** and **Figure S5**). Intestinally, the results showed that a moderate correlation between the scores for perseverance of effort and the diagnostic accuracy existed when endoscopists made the diagnosis without the DCNN's assistance (r = 0.470, P = 0.037, **Table 4** and **Figure 3**). Notably, there was no significant correlation between the scores for the perseverance of effort and diagnostic accuracy when the endoscopists were assisted with the DCNN (r = 0.076, P = 0.750, **Table 4** and **Figure 3**).

DISCUSSION

In this study, we developed a DCNN model to assist endoscopists in making accurate intramucosal GC diagnoses. The DCNN model showed satisfactory diagnostic performance in discriminating intramucosal GC from advanced lesions. We investigated the assistance of the DCNN model on the improvement of diagnostic performance of endoscopists. With the DCNN model's assistance, the diagnostic accuracy of endoscopists increased statistically in both novice and expert groups. The diagnostic agreement among endoscopists also increased from a moderate level to a substantial level with the DCNN model. The diagnostic time was significantly shortened with the DCNN model's assistance, especially in the novice group.

Operational resection is the only curative therapy for GC, but this therapy can only be adopted in GC patients at an early stage. However, most patients are diagnosed at an advanced stage and are ineligible for curative resection. Previous studies reported that while the 5-year survival rate of advanced GC remained less than 25%, the 5-year survival rate of intramucosal GC exceeded 95% (4, 5). Therefore, it is exceptionally crucial to accurately

TABLE 4 | Correlation between Grit Score and Diagnostic Accuracy.

	Score		Diagnostic accuracy			
			No-assistance (Test1)		AI-assistance (Test2)	
	Mean ± sd	IQR	Correlation, r	P-value	Correlation, r	P-value
Grit score	3.546 ± 0.479	3.083–3.917	0.178	0.452	0.145	0.541
Consistency of interest	3.458 ± 0.677	3.167–3.833	-0.122	0.609	0.145	0.541
Perseverance of effort	3.633 ± 0.540	3.292–4.000	0.470	0.037	0.076	0.750

Sd, standard deviation; IQR, interquartile range.

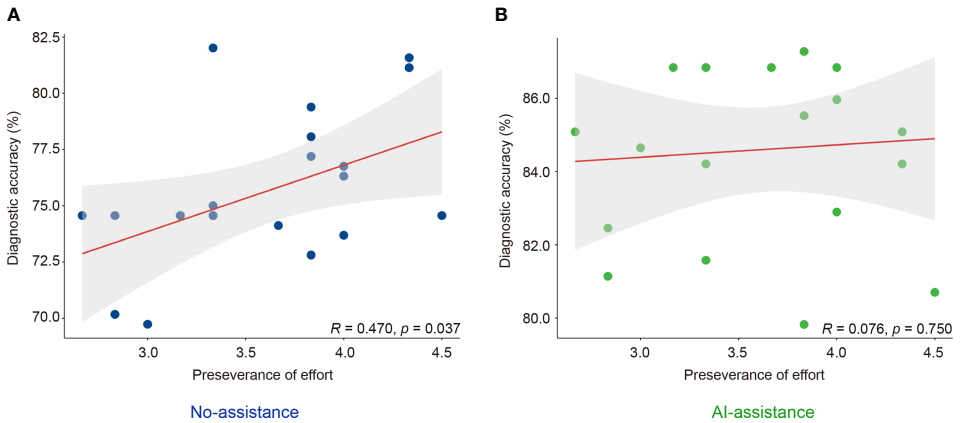


FIGURE 3 | Correlation of perseverance of effort and diagnostic accuracy of endoscopists with **(B)** or without **(A)** the DCNN model's assistance.

differentiate intramucosal GC from advanced lesions for preoperative evaluation and determining the optimal treatment (8).

C-WLI was the main-used modality to predict intramucosal GC, with its accuracy ranging from 70% to 85%. Although the diagnostic performance was comparable to other modalities like EUS or ME-NBI, 15% of cases would be underestimated or overestimated (10, 12, 14, 15). Moreover, since endoscopic examinations were relatively subjective, the interobserver agreement varied significantly amongst endoscopists with different expertise (24). AI-assisted diagnostic devices may help improve the relatively low accuracy and interobserver agreement and reduce the time and effort required to master these methodologies. Three preliminary studies have reported that AI showed a better performance in the diagnosis of intramucosal GC than C-WLI, with the accuracy ranging from 73.0% to 94.5% (18, 20, 21). Here, we developed a DCNN model with a robust performance in discriminating intramucosal GC from advanced lesions with an AUC value of 0.942. The accuracy, sensitivity, and specificity of the DCNN model were 88.2%, 90.5%, and 85.3%, respectively. Consistent with the previously reported DCNN systems, our model showed satisfactory diagnostic performance. However, these preliminary studies mainly focused on developing and validating AI models to diagnose intramucosal GC. Rare studies were conducted to evaluate the assistance of AI models in improving the diagnostic performance of endoscopists. This is extremely important since AI models cannot make the final diagnosis considering safety, accountability, and ethics despite having higher diagnostic accuracy, sensitivity, and specificity than expert endoscopists. Therefore, we further evaluated the assistant role of the DCNN model on the diagnostic performance of endoscopists.

This study showed that endoscopists could benefit significantly from AI assistance in three areas. Firstly, novices achieved considerable improvement in diagnostic performance, which was approximately the same as experts with AI assistance. This improvement significantly reduced the threshold for novices predicting intramucosal GC, which may help these novice endoscopists predict more intramucosal GC during endoscopic examinations. Technologies of visualization were used to locate the intramucosal GC lesions in the images, enabling endoscopists to understand these lesions (**Video 1**) intuitively. Moreover, consistent with previous studies, the diagnostic specificity of intramucosal GC in inexperienced endoscopists was relatively low without effective training (10, 14). Notably, low specificity can be catastrophic since it indicates too many advanced GC lesions being underestimated as intramucosal GC. Therefore, improvement in specificity is essential for optimizing the benefit for patients. Herein, we noticed a significant increase of specificity in novice endoscopists with AI assistance, which may reduce the under-diagnosis rate in clinical practice. Secondly, the interobserver agreement among endoscopists was elevated significantly with the DCNN model's assistance. Several studies have reported relatively low interobserver agreement of novice endoscopists in diagnosing gastric lesions during endoscopic examinations (25). In this study, we noticed that the interobserver agreement of novice endoscopists was comparable with that of experienced endoscopists with AI assistance. The high agreement reduced the discrepancy in diagnosis and promoted homogenization of diagnostic performance, thus alleviating the diagnostic disputes observed in China. Thirdly, diagnostic duration was statistically reduced in the novice group. This indicates that the DCNN model may help endoscopists with limited training increase their diagnostic efficiency. However, we noticed a slightly longer diagnostic time in expert endoscopists. This may be induced by time lags arising from the inconsistencies between the diagnoses made by the DCNN model and the experts. While novice endoscopists tend to accept the diagnosis of the DCNN model, the experts tend to think it over when they encounter inconsistent diagnoses made by the DCNN.

To gain competence in endoscopic procedures, endoscopists need to practice a substantial amount to reach the threshold number (26). With AI assistance, novice endoscopists achieved comparable diagnostic performance with experts without much

additional effort. Additionally, several studies have used Grit Scale tests to evaluate the perseverance and interest for long-time goals (23, 27, 28). Higher grit scores were associated with better performance in multiple settings, including medical school and residency training (27, 28). A previous study indicated that higher grit, significantly higher consistency of interest, was associated with the flexible acceptance of AI assistance (29). However, we noticed that a higher score of effort was correlated with diagnostic accuracy without AI assistance. With AI assistance, the correlation between the perseverance of effort and diagnostic accuracy was diminished. This indicates that AI assistance may reduce the threshold number of procedures required by endoscopists to gain competence. However, this also brings up the point that the novices may begin to rely too much on AI assistance, reducing their ability to make independent diagnoses. Therefore, further investigations are required to evaluate the effect of AI assistance on independent diagnosis ability in endoscopists.

This study has several limitations. Firstly, the DCNN model cannot be applied to poor-quality images, and we excluded these poor-quality images, including images with less insufflation of air, halation, defocus, blurs. We are collecting these poor-quality images and developing an AI classification model to discriminate between poor-quality and high-quality images to solve this issue. Secondly, the training and testing datasets are from one retrospective dataset, which cannot rule out selection bias. As the testing dataset was randomly selected from the retrospective dataset, the excellent performance of the DCNN model in this independent dataset partly demonstrated the potential of this DCNN model. However, the performance and generalizability remained to be evaluated in other prospective datasets. Thirdly, this is a retrospective study, and the excellent performance of the DCNN system may not reflect the clinical application in the real world. Here, we used 54 videos to assess the real-time performance of AI and evaluate the AI assistance on endoscopists to imitate the actual clinical settings. This may partly demonstrate a good result of AI assistance on the performance of endoscopists. But prospective randomized controlled trials are needed to validate the results in actual clinical settings. Fourth, we only included images with histologically proven malignancy, indicating the system could not be used to differentiate malignant lesions from non-cancer mucosa. We have established an AI system in detecting early gastric cancer from non-cancer mucosa in our previous report (30). The two systems can be used together to detect early gastric cancer lesions from non-cancer mucosa first and then differentiate intramucosal GC from advanced lesions, thus may facilitate the endoscopic treatment of GC.

CONCLUSION

In conclusion, we developed and validated an AI-assisted system that could predict intramucosal GC with high accuracy and short duration. We found that AI assistance helped novice endoscopists achieve comparable diagnostic accuracy and duration with expert endoscopists with minimal training or effort. In the future, more studies are needed to examine the effect of AI-assisted systems on the ability of novice endoscopists to establish independent diagnoses.

DATA AVAILABILITY STATEMENT

The datasets used and analyzed during the study are available from the corresponding author on reasonable request approved by the IRB of Nanjing University Medical School Affiliated Drum Tower Hospital (XP.Z. zouxp@nju.edu.cn).

ETHICS STATEMENT

The study design was reviewed and approved by the Medical Ethics Committee at each institution (NJDTH, IRB no. 2020-026-01; RHWU, WDRY2019-K091). Informed consent was waived given the use of only retrospectively deidentified endoscopic images.

AUTHOR CONTRIBUTIONS

XZ, GX, and HY conceived and designed the study. DT, JZ, LW, MN, RL, XC, XH, and LZ contributed to the acquisition of data. DT, MC, and XD contributed to analysis and interpretation of data. HY supervised the construction of deep learning algorithms. DT, JZ, GX, and XZ drafted and reviewed the manuscript. SH reviewed the manuscript and conducted language editing. XZ supported the project. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by the National Natural Science Foundation of China (Grant Nos. 81672935, 81871947), Jiangsu Clinical Medical Center of Digestive System Diseases and Gastrointestinal Cancer (Grant No. YXZXB2016002), and the Nanjing Science and technology development Foundation (Grant No. 2017sb332019). The funders were not involved in the study design, data collection, analysis, or manuscript preparation.

ACKNOWLEDGMENTS

We thank Drs. Yi Wang, Huimin Guo, Tian Yang, Nina Zhang, and Bin Yang for their contributions to this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.622827/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590
2. Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H. Gastric cancer. *Lancet* (2016) 388(10060):2654–64. doi: 10.1016/S0140-6736(16)30354-3
3. Ono H, Kondo H, Gotoda T, Shirao K, Yamaguchi H, Saito D, et al. Endoscopic mucosal resection for treatment of early gastric cancer. *Gut* (2001) 48(2):225–9. doi: 10.1136/gut.48.2.225
4. Gotoda T, Yanagisawa A, Sasako M, Ono H, Nakanishi Y, Shimoda T, et al. Incidence of lymph node metastasis from early gastric cancer: estimation with a large number of cases at two large centers. *Gastric Cancer* (2000) 3(4):219–25. doi: 10.1007/PL00011720
5. Gotoda T, Yamamoto H, Soetikno RM. Endoscopic submucosal dissection of early gastric cancer. *J Gastroenterol* (2006) 41(10):929–42. doi: 10.1007/s00535-006-1954-3
6. Ono H, Yao K, Fujishiro M, Oda I, Nimura S, Yahagi N, et al. Guidelines for endoscopic submucosal dissection and endoscopic mucosal resection for early gastric cancer. *Dig Endosc* (2016) 28(1):3–15. doi: 10.1111/den.12518
7. Japanese Gastric Cancer Association. *Japanese gastric cancer treatment guidelines 2018 (5th edition)*. Gastric Cancer (2020) 24(1):1–21. doi: 10.1007/s10120-020-01042-y
8. Isomoto H, Shikuwa S, Yamaguchi N, Fukuda E, Ikeda K, Nishiyama H, et al. Endoscopic submucosal dissection for early gastric cancer: a large-scale feasibility study. *Gut* (2009) 58(3):331–6. doi: 10.1136/gut.2008.165381
9. Hatta W, Gotoda T, Oyama T, Kawata N, Takahashi A, Yoshifuku Y, et al. A Scoring System to Stratify Curability after Endoscopic Submucosal Dissection for Early Gastric Cancer: "eCura system". *Am J Gastroenterol* (2017) 112(6):874–81. doi: 10.1038/ajg.2017.95
10. Choi J, Kim SG, Im JP, Kim JS, Jung HC, Song IS. Comparison of endoscopic ultrasonography and conventional endoscopy for prediction of depth of tumor invasion in early gastric cancer. *Endoscopy* (2010) 42(9):705–13. doi: 10.1055/s-0030-1255617
11. Choi J, Kim SG, Im JP, Kim JS, Jung HC, Song IS. Is endoscopic ultrasonography indispensable in patients with early gastric cancer prior to endoscopic resection? *Surg Endosc* (2010) 24(12):3177–85. doi: 10.1007/s00464-010-1112-0
12. Li HY, Dai J, Xue HB, Zhao YJ, Chen XY, Gao YJ, et al. Application of magnifying endoscopy with narrow-band imaging in diagnosing gastric lesions: a prospective study. *Gastrointest Endosc* (2012) 76(6):1124–32. doi: 10.1016/j.gie.2012.08.015
13. Dohi O, Yagi N, Majima A, Horii Y, Kitaichi T, Onozawa Y, et al. Diagnostic ability of magnifying endoscopy with blue laser imaging for early gastric cancer: a prospective study. *Gastric Cancer* (2017) 20(2):297–303. doi: 10.1007/s10120-016-0620-6
14. Choi J, Kim SG, Im JP, Kim JS, Jung HC, Song IS. Endoscopic prediction of tumor invasion depth in early gastric cancer. *Gastrointest Endosc* (2011) 73(5):917–27. doi: 10.1016/j.gie.2010.11.053
15. Tsujii Y, Kato M, Inoue T, Yoshii S, Nagai K, Fujinaga T, et al. Integrated diagnostic strategy for the invasion depth of early gastric cancer by conventional endoscopy and EUS. *Gastrointest Endosc* (2015) 82(3):452–9. doi: 10.1016/j.gie.2015.01.022
16. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin* (2019) 69(2):127–57. doi: 10.3322/caac.21552
17. Hirasawa T, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* (2018) 21(4):653–60. doi: 10.1007/s10120-018-0793-2
18. Yoon HJ, Kim S, Kim JH, Keum JS, Oh SI, Jo J, et al. A Lesion-Based Convolutional Neural Network Improves Endoscopic Detection and Depth Prediction of Early Gastric Cancer. *J Clin Med* (2019) 8(9):1310. doi: 10.3390/jcm8091310
19. Luo H, Xu G, Li C, He L, Luo L, Wang Z, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *Lancet Oncol* (2019) 20(12):1645–54. doi: 10.1016/S1470-2045(19)30637-0
20. Zhu Y, Wang QC, Xu MD, Zhang Z, Cheng J, Zhong YS, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. *Gastrointest Endosc* (2019) 89(4):806–15. doi: 10.1016/j.gie.2018.11.011
21. Nagao S, Tsuji Y, Sakaguchi Y, Takahashi Y, Minatsuki C, Niimi K, et al. Highly accurate artificial intelligence systems to predict the invasion depth of gastric cancer: efficacy of conventional white-light imaging, nonmagnifying narrow-band imaging, and indigo-carmine dye contrast imaging. *Gastrointest Endosc* (2020) 9(4):866–73.e1. doi: 10.1016/j.gie.2020.06.047.
22. He K, Zhang X, Ren S, Sun J. eds. Deep Residual Learning for Image Recognition, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2016) 1:770–8. doi: 10.1109/CVPR.2016.90
23. Duckworth AL, Peterson C, Matthews MD, Kelly DR. Grit: perseverance and passion for long-term goals. *J Pers Soc Psychol* (2007) 92(6):1087–101. doi: 10.1037/0022-3514.92.6.1087
24. Peery AF, Cao H, Dominik R, Shaheen NJ, Dellon ES. Variable reliability of endoscopic findings with white-light and narrow-band imaging for patients with suspected eosinophilic esophagitis. *Clin Gastroenterol Hepatol* (2011) 9(6):475–80. doi: 10.1016/j.cgh.2011.02.026
25. Miwata T, Quach DT, Hiyama T, Aoki R, Le HM, Tran PL, et al. Interobserver and intraobserver agreement for gastric mucosa atrophy. *BMC Gastroenterol* (2015) 15:95. doi: 10.1186/s12876-015-0327-x
26. Ekkelenkamp VE, Koch AD, de Man RA, Kuipers EJ. Training and competence assessment in GI endoscopy: a systematic review. *Gut* (2016) 65(4):607–15. doi: 10.1136/gutjnl-2014-307173
27. Hughes BD, Perone JA, Cummins CB, Sommerhalder C, Tyler DS, Bowen-Jallow KA, et al. Personality Testing May Identify Applicants Who Will Become Successful in General Surgery Residency. *J Surg Res* (2019) 233:240–8. doi: 10.1016/j.jss.2018.08.003
28. Sharkey CM, Bakula DM, Gamwell KL, Mullins AJ, Chaney JM, Mullins LL. The Role of Grit in College Student Health Care Management Skills and Health-Related Quality of Life. *J Pediatr Psychol* (2017) 42(9):952–61. doi: 10.1093/jpepsy/jsx073
29. Jin EH, Lee D, Bae JH, Kang HY, Kwak MS, Seo JY, et al. Improved Accuracy in Optical Diagnosis of Colorectal Polyps Using Convolutional Neural Networks with Visual Explanations. *Gastroenterology* (2020) 158(8):2169–79.e8. doi: 10.1053/j.gastro.2020.02.036
30. Tang D, Wang L, Ling T, Lv Y, Ni M, Zhan Q, et al. Development and validation of a real-time artificial intelligence-assisted system for detecting early gastric cancer: A multicentre retrospective diagnostic study. *EBioMedicine* (2020) 62:103146. doi: 10.1016/j.ebiom.2020.103146

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 04 May 2021
doi: 10.3389/fmed.2021.576422

Ultrasound Imaging and Antithrombotic Effects of PLA-Combined Fe_3O_4-GO-ASA Multifunctional Nanobubbles

Jie Zhang[1], *Zheng Liu*[1], *Cunyi Chang*[1], *Ming Hu*[2], *Yang Teng*[1], *Jinjing Li*[1], *Xiangyu Zhang*[1] *and Yanxia Chi*[3]*

[1] School of Pharmacy, Jiamusi University, Jiamusi, China, [2] School of Material Science and Engineering, Jiamusi University, Jiamusi, China, [3] School of Stomatology, Jiamusi University, Jiamusi, China

PLA-combined ferroferric oxide–graphene oxide–aspirin (Fe_3O_4-GO-ASA) multifunctional nanobubbles were prepared using the double emulsion-solvent evaporation method. The obtained composite nanobubbles had a regular spherical shape, Zeta potential of (−36.5 ± 10.0) mV, and particle size distribution range of 200–700 nm. The experimental results showed that PLA-combined Fe_3O_4-GO-ASA nanobubbles could effectively improve the antithrombin parameters of PT, TT, APTT, and INR, and significantly inhibit thrombosis when the composite nanobubbles with a concentration of 80 mg·mL^{-1} interacted with the rabbit blood. The prepared composite nanobubbles could reach a significant ultrasonic imaging effect and good magnetic targeting under the magnetic field when the nanobubbles' concentration was only 60 mg·mL^{-1}.

Keywords: graphene oxide, multifunctional nanobubbles, magnetic target, thrombosis, nanobubbles

OPEN ACCESS

Edited by:
Francesco Rundo, STMicroelectronics, Italy

Reviewed by:
Puja Panwar Hazari, Institute of Nuclear Medicine & Allied Sciences (DRDO), India
Monica Argenziano, University of Turin, Italy

***Correspondence:**
Yanxia Chi
yanxiachi123@163.com

Specialty section:
This article was submitted to Nuclear Medicine, a section of the journal Frontiers in Medicine

Received: *26 June 2020*
Accepted: *30 March 2021*
Published: *04 May 2021*

Citation:
Zhang J, Liu Z, Chang C, Hu M, Teng Y, Li J, Zhang X and Chi Y (2021) Ultrasound Imaging and Antithrombotic Effects of PLA-Combined Fe_3O_4-GO-ASA Multifunctional Nanobubbles. Front. Med. 8:576422. doi: 10.3389/fmed.2021.576422

INTRODUCTION

At present, ultrasound contrast agents with the functions of targeted diagnosis, interventional therapy, and molecular imaging have become a popular research focus (1). The multifunctional ultrasonic contrast agent nanobubbles can not only improve the contrast and clarity of ultrasound images, but can also have other auxiliary effects by combining with other biomedical nanomaterials such as magnetic nanoparticles, drugs, and genes (2–4).

Timely diagnosis and treatment of thrombotic diseases are important for the prognosis and outcome of treatment. At present, the commonly targeted thrombolytic nanoparticles can be roughly divided into three types (5): magnetic nanoparticles loaded with the thrombolytic drugs (6, 7), a polylactic acid polymer material with the thrombolytic drugs (8, 9) and liposomes with thrombolytic drugs (10, 11). Ultrasonic imaging, as a medical diagnostic technology, has advantages such as being non-invasive, having low adverse reactions, and no radiation, as well as a convenient operation and real-time imaging, which means it shows great potential in the diagnosis of thrombotic diseases (12, 13). Ultrasound imaging can not only be used to detect thrombosis symptoms, but also to dissolve blood clots (14). Studies have shown that ultrasound imaging can effectively dissolve the thrombus of the coronary arteries, cerebral aorta, and peripheral arteries, which results from the mechanical and cavitation effect during the ultrasound (15, 16). Platelet is the central link in thrombosis and plays a key role in thrombosis, particularly in arterial and microvascular thrombosis (17, 18). Aspirin (ASA) is one of the most widely used antiplatelet drugs

and is commonly used for the primary and secondary prevention and treatment of thrombocytosis or thrombosis.

Graphene oxide (GO) has good water solubility and biocompatibility (19, 20) due to the plentiful oxygen functional groups (such as -OH and -COOH) on its surface. GO can also load drugs with the benzene ring structure through the π-π conjugation effect. These properties have made GO the focus in a wide range of research, such as in biosensor, bone material, targeted deliver medication, and other biomedical fields (21). As a kind of magnetic substance, nano-iron tetroxide (Fe_3O_4) particle has various applications in magnetic targeting delivery systems. This is due to the fact that Fe_3O_4 particle is non-toxic to cells, has good biocompatibility and excellent magnetic properties, and can be excreted from the body through degradation (22). Magnetic GO is also used in many applications; GO modified with Fe_3O_4 nanoparticles was used for the nanocarrier of doxorubicin, which increased the targeted release effect of drugs (23). The magnetic GO functionalized with PEG could be used for localized photothermal ablation *in vitro* for cancer cells (24). In addition, magnetic GO is widely used for magnetic resonance imaging, molecular imaging probing, and other tasks.

In this study, GO was used as the carrier. Fe_3O_4 nanoparticles as the magnetic targeting factor grew on its surface *in-situ*, and the GO-Fe_3O_4 magnetic targeting complex was obtained. Then acetyl salicylic acid (ASA) was loaded on the surface of GO-Fe_3O_4 complex by π-π function. A PLA complexed with Fe_3O_4-GO-ASA multifunctional nanobubble was designed and prepared, and this composite nanobubble could perform magnetically targeted ultrasonic imaging on the thrombus site and simultaneously inhibit the thrombosis formation caused by platelet aggregation. This study hoped to provide a new avenue for the research and development of an ultrasonic contrast agent in the thrombosis field.

MATERIALS AND METHODS

Materials

$FeCl_2$ and $FeCl_3$ was obtained from Beichen Founder Reagent Factory, Tianjin, China. Acetylsalicylic acid was obtained from ASA, McLean Co., Ltd., Shanghai, China and polylactic acid was supplied by PLA, Kemio Chemical Reagent Co., Ltd., Tianjin, China. Polyethylene glycol (PEG), polyvinyl alcohol (PVA), methylene chloride, isopropanol, and anhydrous ethanol were obtained from Kaitong Chemical Reagent Co., Ltd., Tianjin, China, while New Zealand white rabbits and Kunming mice were provided by Jiamusi University Animal Experimental Center, Heilongjiang, China.

Characterization

The average particle size and Zeta potential of the nanobubbles were measured by laser particle size tester (DLS). The loading rate of ASA in Fe_3O_4-GO-ASA complex was determined by UV-visible spectrophotometer (UV-vis), and the chemical composition of GO-ASA complex was determined by Fourier transform infrared spectroscopy (FT-IR). Scanning electron microscope (SEM) and Transmission electron microscopy (TEM) were used to observe the microscopic morphology of PLA-combined Fe_3O_4-GO-ASA nanobubbles. The magnetic properties of Fe_3O_4 nanoparticles, Fe_3O_4-GO complex, and PLA-combined Fe_3O_4-GO-ASA nanobubbles were determined by sample vibration magnetometer (VSM).

Preparation of Fe_3O_4-GO Complex

First, GO was prepared by an improved Hummers method (25). Fe_3O_4-GO complex was synthesized by *in-situ* growth method. $FeCl_2$ of 0.23g and $FeCl_3$ of 0.47g were dissolved in the dual-distilled water and placed in a round-bottomed flask. Polyethylene glycol (PEG, M_w =1,000) was added under stirring at 60°C. GO of 0.15g was dispersed in dual-distilled water and added into the above system. Meanwhile, the pH of this system was kept at 10–11 by the drop wise addition of NaOH solution and the reaction was maintained for 20 min. Then, the system was heated to 80°C and aged for 30 min. The resulting precipitation was repeatedly washed with the dual-distilled water to neutral, and Fe_3O_4-GO complex was obtained by vacuum freeze drying.

Preparation of Fe_3O_4-GO-ASA Complex

Of the Fe_3O_4-GO complex, 0.1 g was added to an appropriate amount of anhydrous ethanol for ultrasonic dispersion. Then ASA of 0.2 g was added into the dispersion system of Fe_3O_4-GO complex and this mixture system was stirred at 25°C for 2 h. Finally, the precipitation was repeatedly washed by the anhydrous ethanol and treated by vacuum freeze drying to obtain the Fe_3O_4-GO-ASA complex.

Preparation of PLA-Combined Fe_3O_4-GO-ASA Nanobubbles

PLA-combined Fe_3O_4-GO-ASA nanobubbles were prepared by the double emulsion-solvent evaporation method. PLA of 0.6 g was dissolved in the methylene dichloride and a certain amount of the dispersion liquid of Fe_3O_4-GO-ASA complex was added, with ultrasonic emulsification under nitrogen for 5 min. The primary emulsion of W/O was obtained this way. Then primary emulsion of W/O was added into the aqueous solution of polyvinylalcohol (PVA) and ultrasonic emulsification was again performed to get the multiple emulsion of W/O/W. Multiple emulsion of W/O/W was added into aqueous solution of isopropyl alcohol. The mixed system was stirred for 6 h and washed with the centrifugal machine three times, and the obtained supernatant was retained for later use. The obtained precipitation was freeze-dried to obtain PLA-combined Fe_3O_4-GO-ASA nanobubbles.

Ultrasonic Imaging Experiment

Of the four New Zealand rabbits, one was injected with 2 mL physiological saline, and the other three were, respectively, injected with 2 mL PLA-combined Fe_3O_4-GO-ASA nanobubbles with concentrations of 20, 60, and 100 $mg{\cdot}mL^{-1}$ physiological saline solution through the auricular vein. The abdominal cavity was opened at the lower abdomen cut of the rabbit xiphoid and the abdominal aorta shell was stripped off. After the abdominal aorta was exposed, this segment of blood vessel was applied with the filter paper that was soaked in $FeCl_3$ solution. After 15 min, the filter paper was removed and the abdomen was

stitched. The composite nanobubbles suspension liquid was injected through the rabbit ear vein and the ultrasonic test (3.5 MHz) on the abdominal aorta was carried out with and without external magnetic field. This present study received protocol approval from the university's Institutional Animal Care and Use Committee (IACUC).

Anti-thrombotic Experiment *in vitro*

A certain amount of PLA-combined Fe_3O_4-GO-ASA nanobubbles were ultrasonically broken in normal saline for 2 h. The broken solution was, respectively, diluted to 20, 40, 60, 80, and 100 mg·mL^{-1}. Some blood was collected from the ear vein of rabbit. 0.1 mL broken solution with the different concentrations interacted with 2 mL of rabbit blood. Then, the obtained blood sample was added into a small test tube containing No.1 surgical line, and the test tube was kept in a water bath of 37°C for 1.5 h. Finally, the formed blood clot was taken, and the wet weight and dry weight were weighed and compared (26).

Another broken solution with different concentrations interacted with 2 mL of rabbit blood for 1 h, and PT, TT, APTT, and INR values of blood samples were, respectively, measured by semi-automatic blood coagulation analyzer (URJT-600, Uritest, Guilin, China).

Biocompatibility

Kunming mice of 23∼25 g were divided into five groups, with three mice in each group. Mice were prostrated and fixed on the experimental table. 0.2 mL of PLA-combined Fe_3O_4-GO-ASA nanobubbles (150 mg· mL^{-1}) were subcutaneously injected into the back of mice. The control group was injected with 0.2 mL of normal saline. Mice in the control group were sacrificed at 1 day and those of the composite nanobubbles group were, respectively, sacrificed at 1, 3, 9, and 21 days after injection, and the local tissue of the injection site was observed. At the same time, the muscle at the injection site was taken and fixed in 10%formalin solution for preservation. The samples were conventionally dehydrated and embedded with the paraffin. After HE staining, samples were observed with the light microscope (27).

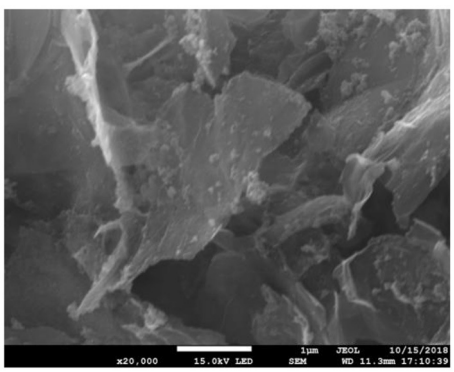


FIGURE 1 | SEM of Fe_3O_4-GO complex.

RESULTS

Figure 1 shows SEM of Fe_3O_4-GO complex. It can be seen from **Figure 1** that there were some nanoparticles on GO nanosheet, which was due to the deposition of Fe_3O_4 nanoparticles on the GO by the *in-situ* growth method. **Figure 2** shows the infrared spectra of ASA, GO, and GO-ASA complex. In **Figure 2a**, the absorption peak near 3,000 cm^{-1} was the -COOH characteristic peak, the absorption peak near 1,600 cm^{-1} was the stretching vibration peak of benzene ring skeleton, the absorption peaks at 1,224 and 1,110 cm^{-1} were caused by C-O stretching vibration, and the absorption peak at 750 cm^{-1} was attributed to the vicinal substitution of benzene ring. These peaks were the most obvious characteristics of aspirin. It can be seen from **Figure 2b** that the absorption peak near 3,430 cm^{-1} was attributed to the stretching vibration peak of –OH in –COOH, the absorption peak at 1,643 cm^{-1} was caused by the bending vibration of –COOH on the edge of GO, and the absorption peak near 1,310 cm^{-1} was the C-O stretching vibration peak. The presence of these oxygen-containing groups indicated that graphite was oxidized. From **Figure 2c**, the absorption peak at 3,430, 1,643, and 1,310 cm^{-1} belonged to the GO characteristic peak. The absorption peak at 1,110 cm^{-1} was the stretching vibration peak of C-O in ASA. The absorption peak of benzene ring appeared near 740 cm^{-1} and moved toward the low wave number, which was caused by the formed π-hydrogen bond between ASA and GO (28). As a result, ASA was loaded on GO by π-π conjugation in the study.

Figures 3–5, respectively, show the SEM, particle size distribution, and Zeta potential of PLA-combined Fe_3O_4-GO-ASA nanobubbles. It can be seen from **Figures 3**, **4** that the obtained composite nanobubbles had a smooth surface, clear boundary, and uniform size, and the particle size of the nanobubbles was 200–700 nm, which met the clinical requirements of ultrasound contrast agent. Among them, the composite nanobubbles presented the irregular spheres shown

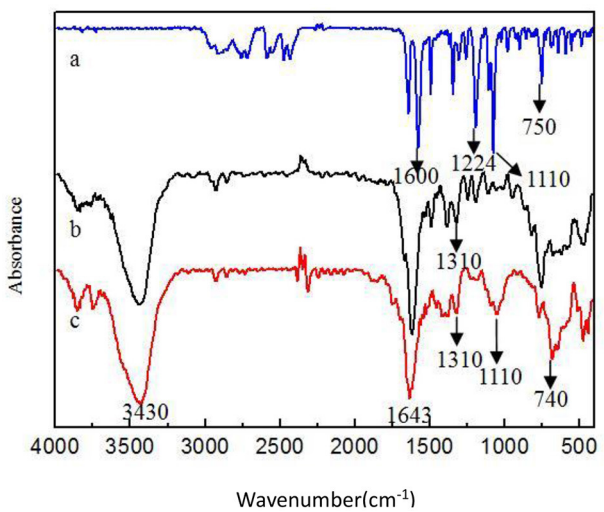


FIGURE 2 | Infrared spectra of different samples **(a)** ASA **(b)** GO **(c)** GO-ASA.

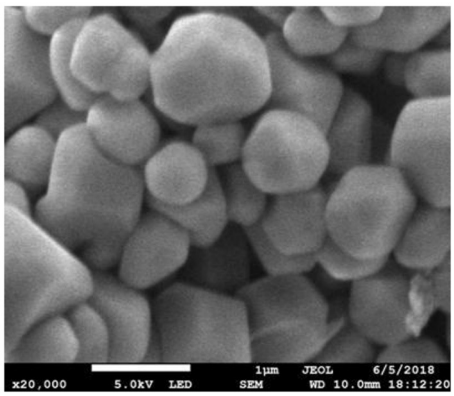


FIGURE 3 | SEM of PLA combined Fe_3O_4-GO-ASA nanobubbles.

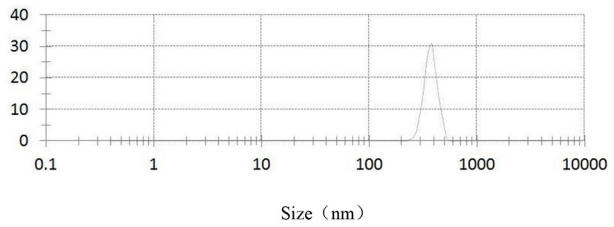


FIGURE 4 | Particle size distribution of PLA combined Fe_3O_4-GO-ASA nanobubbles.

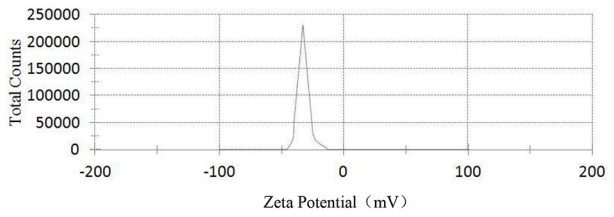


FIGURE 5 | Zeta potential of PLA combined Fe_3O_4-GO-ASA nanobubbles.

in **Figure 3**, resulting from the freeze drying process of the composite nanobubbles. According to **Figure 5**, Zeta potential of the composite nanobubbles was (-36.5 ± 10.0) mV, which proved that aqueous solution of the composite nanobubbles had good dispersion stability.

Figures 6, **7**, respectively, show the preparation process and TEM of PLA-combined Fe_3O_4-GO-ASA nanobubbles. In the preparation process, Fe_3O_4 was firstly grown on the GO surface by an *in-situ* growth method and ASA was loaded on Fe_3O_4-GO by π-π conjugation effect to obtain Fe_3O_4-GO-ASA complex. Then, Fe_3O_4-GO-ASA complex was encapsulated in PLA nanobubbles by double emulsion-solvent evaporation method. During the first ultrasonic emulsification, PLA solution was destroyed into some small droplet wrapping Fe_3O_4-GO-ASA complex in the inner water phase. Due to the continuous N_2 that was imported in the process of ultrasonic breaking, small bubbles also contained N_2. At this moment, the obtained W/O phase was transferred to the PVA solution and evenly stirred, and the second ultrasonic emulsification was performed. As a result, PVA would coat on the surface of PLA nanobubbles by hydrogen bonding to form a hard shell (29). After the vacuum freeze drying, the water in the nanobubbles disappeared, and Fe_3O_4-GO-ASA complex would gather at the inner side of the nanobubble or embed in the film of the nanobubble due to the surface tension. As shown in **Figure 7**, the film and inner side of the nanobubble presented a black shadow.

Figure 8 shows the hysteresis curves of Fe_3O_4, Fe_3O_4-GO complex, and PLA-combined Fe_3O_4-GO-ASA nanobubbles (30, 31). As can be seen from **Figure 8**, the hysteresis curves of Fe_3O_4, Fe_3O_4-GO complex, and PLA-combined Fe_3O_4-GO-ASA nanobubbles presented a similar pattern, meaning that their magnetic behaviors were similar. There was no coercive field nor residual magnetism at the zero point, which indicated that three samples had paramagnetic characteristics and could gather under external magnetic field conditions (32). In addition, the specific saturation magnetization intensity of Fe_3O_4, Fe_3O_4-GO complex and PLA-combined Fe_3O_4-GO-ASA nanobubbles decreased orderly; this was because the Fe_3O_4 content in the three samples of the same mass decreased in sequence, indicating that the magnetism of the Fe_3O_4-GO complex and PLA-combined Fe_3O_4-GO-ASA nanobubbles were derived from Fe_3O_4.

Figure 9 is the ultrasonic imaging of rabbit abdominal aorta of PLA-combined Fe_3O_4-GO-ASA nanobubbles with different concentrations. As can be seen from **Figure 9A**, when PLA-combined Fe_3O_4-GO-ASA nanobubbles were not injected, the ultrasonic imaging of rabbit abdominal aorta was dark and very fuzzy. When the concentration of PLA-combined Fe_3O_4-GO-ASA nanobubbles were 20 $mg{\cdot}mL^{-1}$ (**Figure 9B**), the ultrasonic signal was enhanced, and the contrast and identification of the image was significantly improved. When the concentration of PLA-combined Fe_3O_4-GO-ASA nanobubbles was >60 $mg{\cdot}mL^{-1}$ (**Figures 9C,D**), the effect of ultrasound imaging for the abdominal aorta was significantly improved, while the contrast of the image did not significantly change with the increase of PLA-combined Fe_3O_4-GO-ASA nanobubbles concentration. As a result, the ultrasonic effect had reached the requirements of ultrasound imaging when the concentration of PLA-combined Fe_3O_4-GO-ASA nanobubbles was 60 $mg{\cdot}mL^{-1}$.

PLA-combined Fe_3O_4-GO-ASA nanobubbles were injected into the model rabbit with abdominal aortic thrombus, and ultrasonic imaging photos of rabbit abdominal aorta with and without magnetic field are shown in **Figure 10**. From **Figure 10A**, a strong and distinct blue signal appeared in the right side of the thrombus, indicating the reflux blood presented at the thrombus obstruction site. At the same time, the composite nanobubbles were carried by blood and would flow throughout the body, and the ultrasonic signals could present in the tissues or organs where the composite nanobubbles passed through. As a result, a weak and large dark blue signal also appeared below the abdominal aorta of the rabbit. It can be seen from **Figure 10B** that when the magnetic field was added in the abdominal aorta region, the intensity of ultrasonic

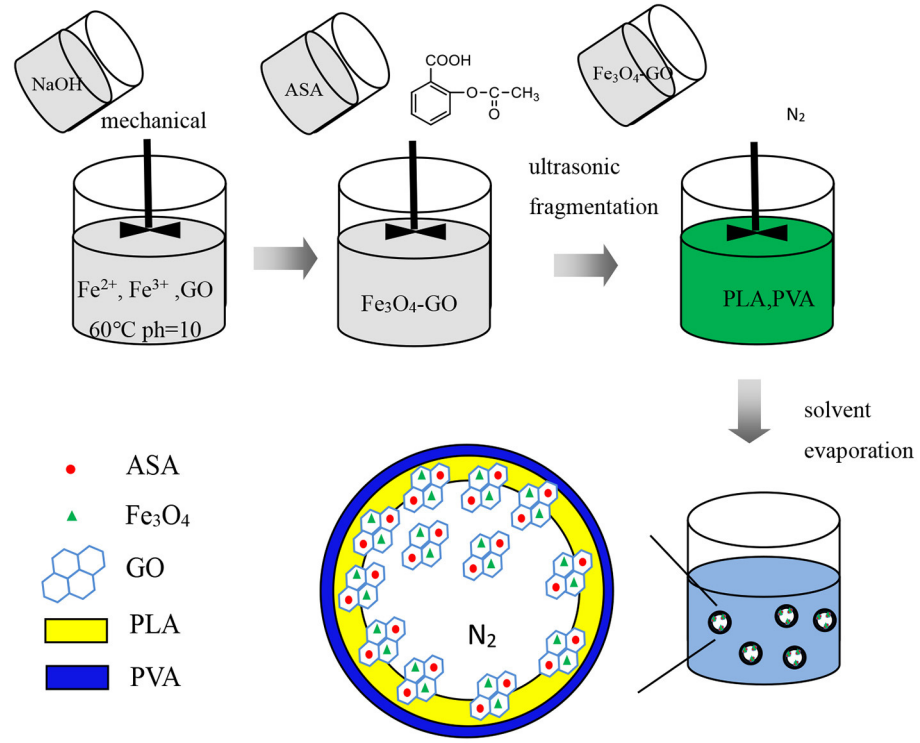


FIGURE 6 | Preparation process of PLA combined Fe_3O_4-GO-ASA nanobubbles.

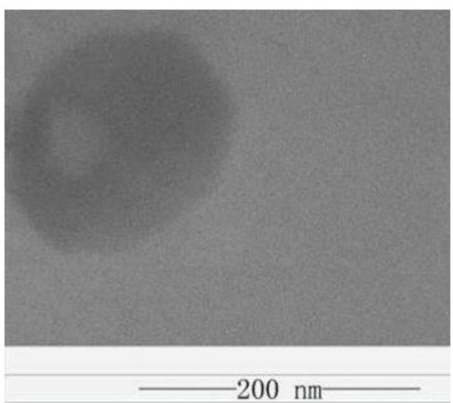


FIGURE 7 | TEM of PLA combined Fe_3O_4-GO-ASA nanobubbles.

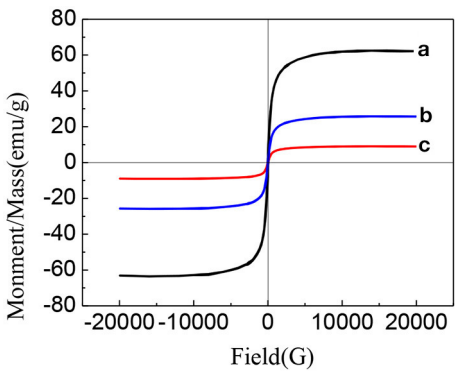


FIGURE 8 | Hysteresis curves of different samples **(a)** Fe_3O_4 **(b)** Fe_3O_4-GO complex **(c)** PLA combined Fe_3O_4-GO-ASA nanobubbles (30, 31).

signal in this region was significantly enhanced, while the signal intensity in other regions was significantly weakened, indicating that PLA-combined Fe_3O_4-GO-ASA nanobubbles could effectively enrich in the magnetic field region. Thus, the obtained PLA-combined Fe_3O_4-GO-ASA nanobubbles had a magnetic targeting effect under the action of an external magnetic field, which could significantly enhance the image clarity of the observed site without increasing the dosage of the composite nanobubbles.

Table 1 shows the wet weight and dry weight of the obtained coagulated blood clot due to the interaction between PLA-combined Fe_3O_4-GO-ASA nanobubbles with different concentrations and blood samples. It can be seen from **Table 1** that the wet weight of the normal saline control group was 95.10 ± 0.22 mg, while the wet weights for the composite nanobubbles with the concentrations of 20, 40, 60, 80, and 100 mg·mL^{-1} were, respectively, 87.27 ± 0.16 mg, 75.73 ± 0.47 mg, 56.57 ± 0.32 mg, 44.80 ± 0.16 mg, and 36.03 ± 0.22 mg, showing the

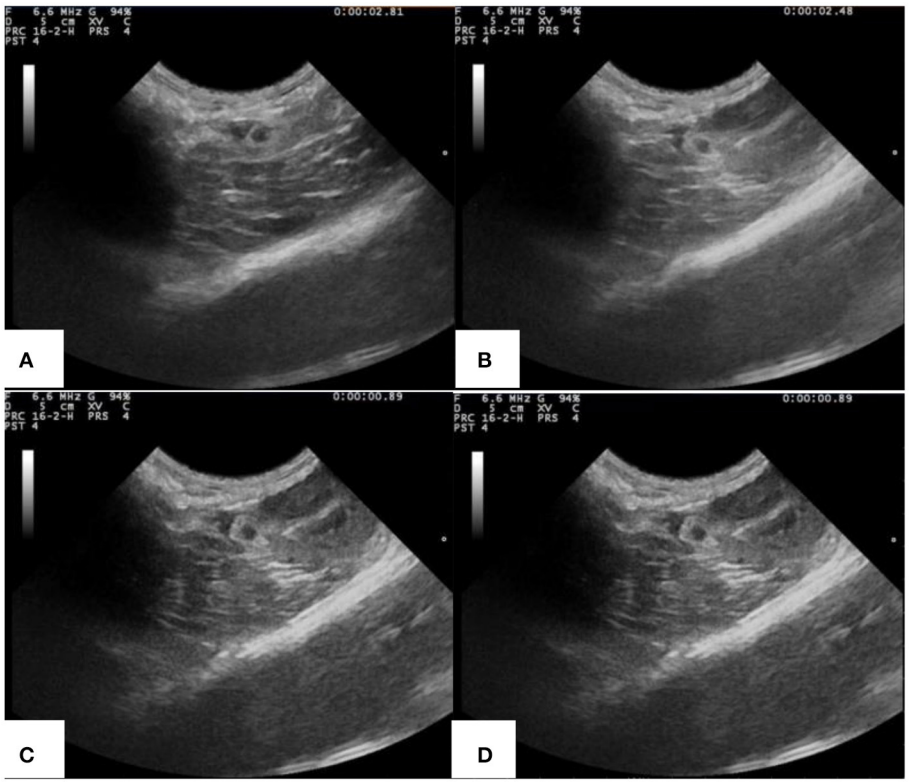


FIGURE 9 | Ultrasonic imaging of rabbit abdominal aorta of rabbits injected with PLA combined Fe_3O_4-GO-ASA nanobubbles of the different concentrations **(A)** 0 mg·mL^{-1} **(B)** 20 mg·mL^{-1} **(C)** 60 mg·mL^{-1} **(D)** 100 mg·mL^{-1}.

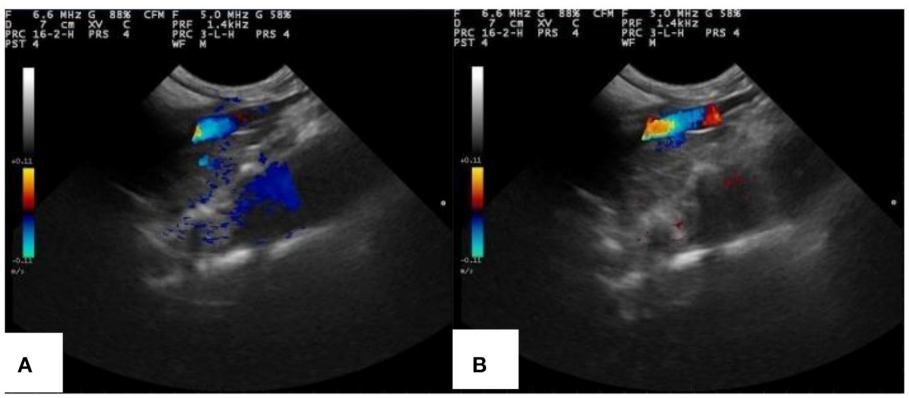


FIGURE 10 | Ultrasonic imaging of rabbit abdominal aorta of rabbits injected with PLA combined Fe_3O_4-GO-ASA nanobubbles **(A)** without the magnetic field **(B)** with the additional magnetic field in the thrombus area.

composite nanobubbles were very significantly different from the control group ($P < 0.01$). The dry weight of the normal saline control group was 52.03 ± 0.35 mg and the dry weight of the composite nanobubble with the concentration of 20 mg·mL^{-1} was 50.17 ± 0.35 mg; it was significantly different from the control group ($0.01<P<0.05$). When the composite nanobubble concentration was 40, 60, 80, and 100 mg·mL^{-1}, their dry weights were, respectively, 44.83 ± 0.36 mg, 36.67 ± 0.50 mg, 33.03 ± 0.75 mg, and 27.57 ± 0.60 mg and had very significant differences in comparison with the control group ($P < 0.01$). We could see that the wet weight and dry weight of clots significantly decreased with the concentration increase of the composite nanobubbles.

This was because the ASA content in the composite nanobubbles increased with the increase of the composite nanobubbles concentration, and ASA had an inhibitory effect on the thrombosis (33). The antithrombotic effect of ASA is mainly

realized by inhibiting the activity of the platelet cyclooxygenase and preventing the synthesis and release of prostaglandins and thromboxane (34). Moreover, we could conclude from **Table 1** that the ratio of wet weight and dry weight of clots decreased with the increase of the composite nanobubbles concentration. For the formation of thrombus, the ratio of wet weight and dry weight ratio is smaller and the dissociation of thrombus factor in the blood is stronger, thus the speed of thrombosis caused by factors becomes slow. The ratio of wet weight and dry weight is lower, indicating that the risk of thrombosis is smaller. It can be seen that the anti-thrombotic ability of PLA-combined Fe_3O_4-GO-ASA nanobubbles increased with the increase of its concentration.

TABLE 1 | Wet weight and dry weight of the coagulated blood clot.

Concentration (mg·mL^{-1})	Wet weight (mg)	Dry weight (mg)
0	95.10 ± 0.22	52.03 ± 0.35
20	87.27 ± 0.16**	50.17 ± 0.35*
40	75.73 ± 0.47**	44.83 ± 0.36**
60	56.57 ± 0.32**	36.67 ± 0.50**
80	44.80 ± 0.16**	33.03 ± 0.75**
100	36.03 ± 0.22**	27.57 ± 0.60**

*$P < 0.05$, **$P < 0.01$.

TABLE 2 | Effects of PLA combined Fe_3O_4-GO-ASA nanobubbles with the different concentrations on the coagulation blood parameters.

Concentration (mg·mL^{-1})	PT(s)	TT(s)	APTT(s)	INR(s)
0	13.2	14.0	29.6	1.09
10	13.5	14.5	29.7	1.11
20	13.7	15.7	30.1	1.16
40	15.3	16.4	30.9	1.23
80	17.2	18.0	32.8	1.45
160	18.8	20.1	33.6	1.56

Table 2 shows the changes of the coagulated blood parameters due to the interaction between PLA-combined Fe_3O_4-GO-ASA nanobubbles of different concentrations and blood samples. As can be seen from **Table 2**, the values of PT, TT, APTT, and INR in blood samples showed an increasing trend with the increase of the composite nanobubbles concentration. Specifically, APTT value significantly increased when the composite nanobubbles concentration was 20 mg·mL^{-1}, PT and TT values significantly increased at the concentration of 40 mg·mL^{-1}, and INR value significantly increased at the concentration of 80 mg·mL^{-1}. It can be seen that when the concentration of the composite nanobubbles was above 80 mg·mL^{-1}, they could effectively promote the synthesis of cAMP in platelet and interdict the platelet membrane glycoproteins receptors, which reduced the concentrations of thromboxane A and phosphodiesterase in plasma and eventually inhibited the formation process of thrombosis. So, PLA-combined Fe_3O_4-GO-ASA nanobubbles could effectively inhibit thrombosis and achieve the purpose of adjuvant therapy.

Figure 11 is the HE staining pictures of mouse muscle after the subcutaneous injection of normal saline and PLA-combined Fe_3O_4-GO-ASA nanobubbles at different times. As shown in **Figure 11A**, mouse muscle tissue had an intact structure and the histopathological change was not observed for the normal saline group. For PLA-combined Fe_3O_4-GO-ASA nanobubbles, the injection site of 1 day formed a cyst by the experimental observation, while the cyst had a thin wall and it contained the nanobubbles injection solution and interstitial fluid. As shown in **Figure 11B**, compared with the saline group, there was a slight inflammatory cell infiltration at the injection site. After PLA-combined Fe_3O_4-GO-ASA nanobubbles were injected for 3 days, the formed cyst at the

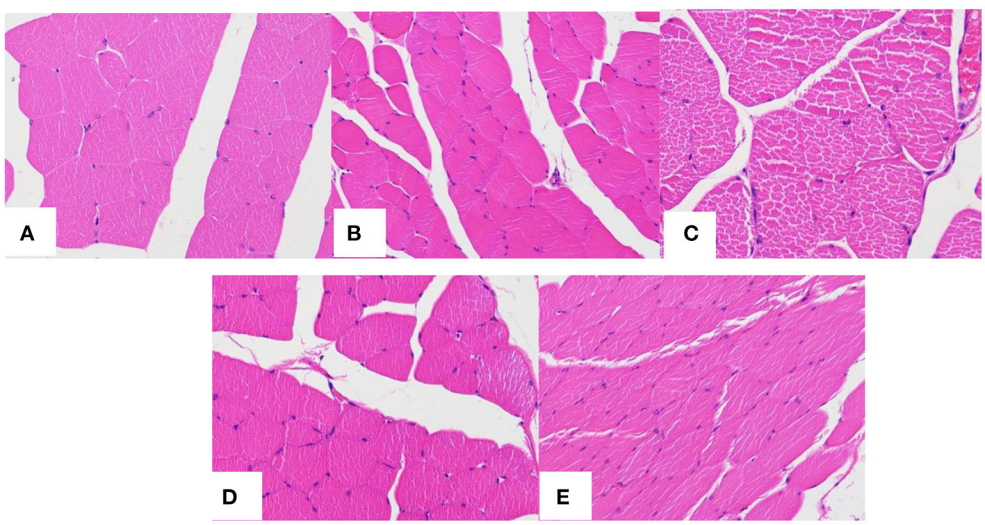


FIGURE 11 | HE staining pictures of mouse muscle after the subcutaneous injection of saline at 1 day **(A)** and the composite nanobubbles at 1 day **(B)**, 3 days **(C)**, 9 days **(D)**, and 21 days **(E)**.

injection site significantly reduced, and the inflammatory cell infiltration was weakened as shown in **Figure 11C**, however, the infiltration phenomenon did not disappear because the degradation products of the composite nanobubbles were not eliminated or absorbed in a timely manner. At 9 days after injection, the local cyst basically disappeared, and the inflammatory cell infiltration significantly reduced (**Figure 11D**). At 21 days after injection, the inflammatory cell infiltration disappeared (**Figure 11E**), and there was no obvious change in the deep muscle tissue. Therefore, the injection of PLA-combined Fe_3O_4-GO-ASA nanobubbles only caused a slight inflammatory response, and the accumulation of tissue penetrating fluid or obvious proliferation of vascular or fibers was not observed, indicating that PLA-combined Fe_3O_4-GO-ASA nanobubbles have good biocompatibility.

CONCLUSIONS

In this study, the nanobubbles were prepared in three steps. Firstly, Fe_3O_4 nanoparticles were deposited on GO using an *in-situ* growth method. Then, ASA was adsorbed on the Fe_3O_4- GO composite through π-π effect. Finally, the newly formed Fe_3O_4-GO-ASA composite was combined into the PLA nanobubbles by the double emulsion-solvent evaporation method.

The obtained PLA-combined Fe_3O_4-GO-ASA nanobubbles can inhibit thrombosis and effectively enrich the target area under the action of a magnetic field to achieve the targeted imaging effect. Moreover, if the ultrasonic power was increased, the composite nanobubbles could be broken and the controlled drug release achieved, indicating that PLA-combined Fe_3O_4-GO-ASA nanobubbles could realize the dual functions of targeted drug delivery and controlled release. However, the antithrombotic performance of PLA-combined Fe_3O_4-GO-ASA nanobubbles *in vivo* will need further study. This study lays a foundation for the development and application of magnetic targeted ultrasound contrast agents.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The animal study was reviewed and approved by Ethics Committee of Jiamusi university Jiamusi university.

AUTHOR CONTRIBUTIONS

JZ and YC are responsible for the design of experimental direction, arrangement of experimental methods, revision, and proofreading of this paper. ZL and CC were responsible for the implementation of the experiment and the preliminary writing of the paper. MH, YT, JL, and XZ were responsible for the evaluation of experimental design.

FUNDING

This work was supported by National Natural Science Foundation of China (Nos. 51671096 and 81601616), Excellent innovation team based on the basic scientific research vocational cost for the provincial undergraduate universities in Heilongjiang (No. 2018-KYYWF-0914), Heilongjiang Health Department Project (2017-419), and Heilongjiang Education Department Project (2017-KYYWF-0601).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.576422/full#supplementary-material

REFERENCES

1. Liu R, Tang J, Xu Y, Dai Z. BRET-FRET nanobubbles for bioluminescence/ultrasound dual-modal imaging of Inflammation *in vivo,* bioluminescence imaging of in?ammation *in vivo* based on bioluminescence and fluorescence resonance energy transfer using nanobubbles ultrasound contrast agent. *ACS Nano.* (2019) 13:5124–32. doi: 10.1021/acsnano.8b08359
2. de Leon A, Perera R, Nittayacharn P, Cooley M, Jung O, Exner AA. Ultrasound contrast agents and delivery systems in cancer detection and therapy. *Adv Cancer Res.* (2018) 139:57–84. doi: 10.1016/bs.acr.2018.04.002
3. Koczera P, Appold L, Shi Y, Liu M, Dasgupta A, Pathak V, et al. PBCA-based polymeric microbubbles for molecular imaging and drug delivery. *J Control Release.* (2017) 259:128–35. doi: 10.1016/j.jconrel.2017.03.006
4. Martin KH, Dayton PA. Current status and prospects for microbubbles in ultrasound theranostics. *Wiley Interdiscip Rev Nanomed Nanobiotechnol.* (2013) 5:329–45. doi: 10.1002/wnan.1219
5. Zhou J, Guo DJ. Research progress of targeted thrombolysis in drug delivery system. *Chongqing Med.* (2014) 43:617–20. doi: 10.3969/j.issn.1671-8348.2014.05.042
6. Chen JP, Yang PC, Ma YH, Wu T. Characterization of chitosan magnetic nanoparticles for in situ delivery of tissue plasminogen activator. *Carbohydr Polym.* (2011) 84:364–72. doi: 10.1016/j.carbpol.2010.11.052
7. Ma YH, Wu SY, Wu T, Chang YJ, Hua MY, Chen JP. Magnetically targeted thrombolysis with recombinant tissue plasminogen activator bound to polyacrylic acid-coated nanoparticles. *Biomaterials.* (2009) 30:3343–51. doi: 10.1016/j.biomaterials.2009.02.034
8. Zhou J. *Construction and Characterization of Multifunctional MR Molecular Probe Carrying rtPA Targeting Thrombus [D].* Chongqing: Chongqing Medical University (2014). doi: 10.7666/d.Y2690646
9. Zhong YX. *Phase Variant Multimodal Nanoparticle Thrombosis Imaging and Thrombolysis of Targeted Fibrin [D].* Chongqing: Chongqing Medical University (2019).
10. Hua X. *Preparation of Ultrasonic Contrast Agent Containing tPA and RGDs and Mechanism of Drug Release and Solubilization [D].* Chongqing: Third Military Medical University (2007). doi: 10.7666/d.y1179627
11. Laing ST, Moody MR, Kim H, Smulevitz B, Huang SL, Holland CK, et al. Thrombolytic efficaccy of tissue plasminogen activitor-loaded echogenic liposomes in a rabbit thrombus. *Thromb Res.* (2012) 130:629–35. doi: 10.1016/j.thromres.2011.11.010

12. Ebelt H, Offhaus A, Wiora M, Roehl P, Schwenzky A, Weida A, et al. Impact of ultrasound contrast agent on the detection of thrombi during transoesophageal echocardiography. *Open Heart*. (2019) 6:e00124. doi: 10.1136/openhrt-2019-001024
13. Plachá D, Jampilek J. Graphenic materials for biomedical applications. *Nanomaterials*. (2019) 9:1–37. doi: 10.3390/nano9121758
14. Zhang B, Wang H. Progress in clinical application of ultrasonic thrombolysis. *Chin J Med Ultrasound*. (2019) 16:785–9. doi: 10.3877/cma.j.issn.1672-6448.2019.10.013
15. Chernysh IN, Everbach CE, Purohit PK, Weisel JW. Molecular mechanisms of the effect of ultrasound on the fibrinolysis of clots. *J Thromb Haemost*. (2015) 13:601–9. doi: 10.1111/jth.12857
16. Lee TH, Yeh JC, Tsai CH, Yang JT, Lou SL, Seak CJ, et al. Improved thrombolytic effect with focused ultrasound and neuroprotective agent against acute carotid artery thrombosis in rat. *Sci Rep*. (2017) 7:1638. doi: 10.1038/s41598-017-01769-2
17. Huang Y, Chen KJ, Yin HJ. Thrombus formation and prevention and treatment of traditional Chinese medicine. *Med. Rev*. (2010) 16:1410–3. doi: 10.3969/j.issn.1006-2084.2010.09.045s
18. Feng ZC. Chronic inflammation of the atherosclerotic vessel wall. *Chin. J. Cardiol*. (2005) 33:393–4. doi: 10.3760/j:issn:0253-3758.2005.05.001
19. Qiu ZC. *Graphene Oxide-Based Nanocomposite for Targeting Colorectal Cancer Therapy*. Wuxi: Jiangnan University (2020). doi: 10.27169/d.cnki.gwqgu.2020.000296
20. Hussien NA, Işiklan N, Türk M. Pectin-conjugated magnetic graphene oxide nanohybrid as a novel drug carrier for paclitaxel delivery. *Artif Cells Nanomed Biotechnol*. (2018) 46:264–73. doi: 10.1080/21691401.2017.1421211
21. Liu J, Xu F, Huang J, Xu J, Liu Y, Yao Y, et al. Low-intensity focused ultrasound (LIFU)-activated nanodroplets as a theranostic agent for noninvasive cancer molecular imaging and drug delivery. *Biomater Sci*. (2018) 6:2838–49. doi: 10.1039/C8BM00726H
22. Zhang F. *Preparation of Magnetic Go Based Non-covalent Modified Nanocomposites and Their Application in Drug Delivery*. Zhenjiang: Jiangsu University (2019). doi: 10.27170/d.cnki.gjsuu.2019.000049
23. Yang X, Zhang X, Ma Y, Huang Y, Wanga Y, Chen Y. Superparamagnetic graphene oxide-Fe_3O_4 nanoparticles hybrid for controlled targeted drug carriers. *J Materials Chem*. (2009) 19:2710–4. doi: 10.1039/b821416f
24. Shi S, Yang K, Hong H, Valdovinosd HF, Nayakc TR, Zhang Y, et al. Tumor vasculature targeting and imaging in living mice with reduced graphene oxide. *Biomaterials*. (2013) 34:3002–9. doi: 10.1016/j.biomaterials.2013.01.047
25. Ren L, Pan S, Li H, Li Y, He L, Zhang S, et al. Effects of aspirin-loaded graphene oxide coating of a titanium surface on proliferation and osteogenic differentiation of MC3T3-E1 cells. *Sci Rep*. (2018) 8:15143–56. doi: 10.1038/s41598-018-33353-7
26. Lei H, Hu Y, Li G. Magnetic poly(phenylene ethynylene) conjugated microporous polymer microspheres for bactericides enrichment and analysis by ultra-high performance liquid chromatography-tandem mass spectrometry. *Chromatogr A*. (2018) 1580:22–29. doi: 10.1016/j.chroma.2018.10.043
27. Otani K, Nishimura H, Kamiya A, Harada-Shiba M. Simplified preparation of αvβ3 integrin-targeted nanobubbles based on a clinically available ultrasound contrast agent: validation in a tumor-bearing mouse model. *Ultrasound Med Biol*. (2018) 44:1063–73. doi: 10.1016/j.ultrasmedbio.2018.01.017
28. Xu Y. *The Regulation of π-Conjugate Oligomer Aggregation Behavior by Hydrogen Bond and π-π Synergy [D]*. Changchun: Jilin University (2009).
29. Abdal-Hay A, Hussein KH, Casettari L, Khalil KA, Hamdy AS. Fabrication of novel high performance ductile poly (lactic acid) nanofiber scaffold coated with poly(vinyl alcohol) for tissue engineering applications. *Mater Sci Eng C Mater Biol Appl*. (2016) 60:143–50. doi: 10.1016/j.msec.2015.11.024
30. Zhang J, Yang J, Zhang H, Hu M, Lia J, Zhanga X. New Span-PEG-composited Fe3O4-CNT as a multifunctional ultrasound contrast agent for inflammation and thrombotic niduses. *RSC Adv*. (2020) 10:38592–600. doi: 10.1039/D0RA05401A
31. Zhang J, Liu Z, Zhou S, Teng Y, Zhang X, Li J. Novel span-PEG multifunctional ultrasound contrast agent based on CNTs as a magnetic targeting factor and a drug carrier. *ACS Omega*. (2020) 5:31525–34. doi: 10.1021/acsomega.0c03325
32. Tang H, Guo Y, Peng L, Fang H, Wang Z, Zheng Y, et al. *In vivo* targeted, responsive, and synergistic cancer nanotheranostics by magnetic resonance imaging-guided synergistic high-intensity focused ultrasound ablation and chemotherapy. *ACS Appl Mater Interfaces*. (2018) 10:15428–41. doi: 10.1021/acsami.8b01967
33. Ornelas A, Zacharias-Millward N, Menter DG, Davis JS, Lichtenberger L, Hawke D, et al. Beyond COX-1: the effects of aspirin on platelet biology and potential mechanisms of chemoprevention. *Cancer Metastasis Rev*. (2017) 36:289–303. doi: 10.1007/s10555-017-9675-z
34. Liu TS. Study on clopidogrel combined with aspirin in the treatment of acute myocardial infarction. *Med Inform*. (2020) 33:59–61.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 13 May 2021
doi: 10.3389/fcomp.2021.613981

FusionNet: A Deep Fully Residual Convolutional Neural Network for Image Segmentation in Connectomics

Tran Minh Quan[1], *David Grant Colburn Hildebrand*[2] *and Won-Ki Jeong*[3]*

[1]College of Engineering and Computer Science, VinUniversity, Ha Noi, Vietnam, [2]Laboratory of Neural Systems, The Rockefeller University, New York, NY, United States, [3]Department of Computer Science and Engineering, Korea University, Seoul, South Korea

Cellular-resolution connectomics is an ambitious research direction with the goal of generating comprehensive brain connectivity maps using high-throughput, nano-scale electron microscopy. One of the main challenges in connectomics research is developing scalable image analysis algorithms that require minimal user intervention. Deep learning has provided exceptional performance in image classification tasks in computer vision, leading to a recent explosion in popularity. Similarly, its application to connectomic analyses holds great promise. Here, we introduce a deep neural network architecture, FusionNet, with a focus on its application to accomplish automatic segmentation of neuronal structures in connectomics data. FusionNet combines recent advances in machine learning, such as semantic segmentation and residual neural networks, with summation-based skip connections. This results in a much deeper network architecture and improves segmentation accuracy. We demonstrate the performance of the proposed method by comparing it with several other popular electron microscopy segmentation methods. We further illustrate its flexibility through segmentation results for two different tasks: cell membrane segmentation and cell nucleus segmentation.

Keywords: connectomic analysis, image segementation, deep learning, refinement, skip connection

OPEN ACCESS

Edited by:
Francesco Rundo, STMicroelectronics, Italy

Reviewed by:
Francesco Guarnera, University of Catania, Italy
Francesca Trenta, University of Catania, Italy

***Correspondence:**
Won-Ki Jeong
wkjeong@korea.ac.kr

Specialty section:
This article was submitted to Computer Vision, a section of the journal Frontiers in Computer Science

Received: *04 October 2020*
Accepted: *12 April 2021*
Published: *13 May 2021*

Citation:
Quan TM, Hildebrand DGC and Jeong W-K (2021) FusionNet: A Deep Fully Residual Convolutional Neural Network for Image Segmentation in Connectomics. Front. Comput. Sci. 3:613981. doi: 10.3389/fcomp.2021.613981

1 INTRODUCTION

The brain is considered the most complex organ in the human body. Despite decades of intense research, our understanding of how its structure relates to its function remains limited (Lichtman and Denk, 2011). Connectomics research seeks to disentangle the complicated neuronal circuits embedded within the brain. This field has gained substantial attention recently thanks to the advent of new serial-section electron microscopy (EM) technologies (Briggman and Bock, 2012; Hayworth et al., 2014; Eberle and Zeidler, 2018; Zheng et al., 2018; Graham et al., 2019). The resolution afforded by EM is sufficient for resolving tiny but important neuronal structures that are often densely packed together, such as dendritic spine necks and synaptic vesicles. These structures are often only tens of nanometers in diameter (Helmstaedter, 2013). **Figure 1** shows an example of such an EM image and its cell membrane segmentation. Such high-resolution imaging results in enormous datasets, approaching one petabyte for only the relatively small tissue volume of one cubic millimeter. Therefore, handling and analyzing EM datasets is one of the most challenging problems in connectomics.

Early connectomics research focused on the sparse reconstruction of neuronal circuits (Bock et al., 2011; Briggman et al., 2011), meaning they focused reconstruction efforts on a subset of neurons in the data using manual or semi-automatic tools (Jeong et al., 2010; Sommer et al., 2011; Cardona et al.,

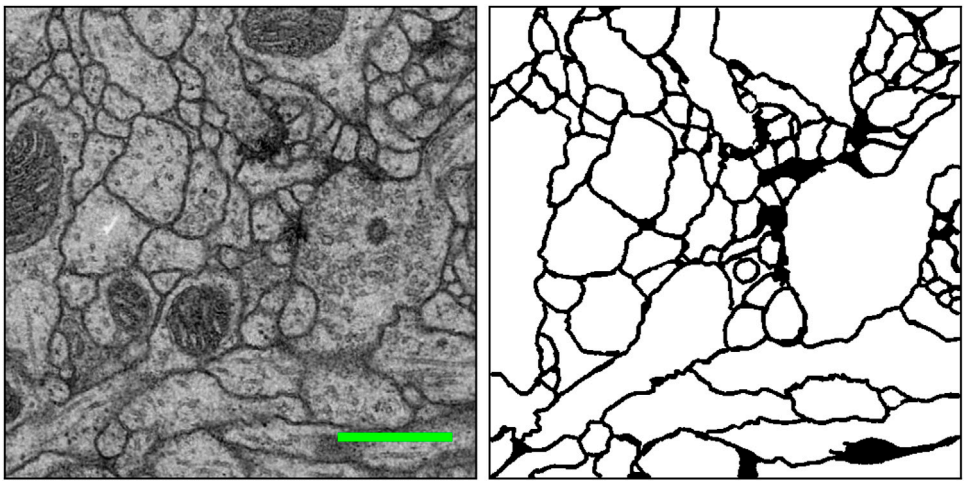


FIGURE 1 | An example EM image **(left)** and its manually extracted cellular membrane segmentation result **(right)** from the ISBI 2012 EM segmentation challenge (Arganda-Carreras et al., 2015). Scale bar (green): 500 nm.

2012). Unfortunately, this approach requires too much human interaction to scale well over the vast amount of EM data that can be collected with new technologies. Therefore, developing scalable and automatic image analysis algorithms is an important and active research direction in the field of connectomics.

Although some EM image processing pipelines use conventional, light-weight pixel classifiers [e.g., RhoANA (Kaynig et al., 2015)], the majority of automatic image segmentation algorithms for connectomics rely on deep learning. Earlier automatic segmentation work using deep learning mainly focused on patch-based pixel-wise classification based on a convolutional neural network (CNN) for affinity map generation (Turaga et al., 2010) and cell membrane probability estimation (Ciresan et al., 2012). However, one limitation of applying a conventional CNN to EM image segmentation is that per-pixel network deployment scaling becomes prohibitively expensive considering the tera-scale to peta-scale EM data size. For this reason, more efficient and scalable deep neural networks are important for image segmentation of the large datasets that can now be produced. One approach is to extend a fully convolutional neural network (FCN) (Long et al., 2015), which uses encoding and decoding phases similar to an autoencoder for the end-to-end semantic segmentation problem (Ronneberger et al., 2015; Chen et al., 2016a).

The motivation of the proposed work stems from our recent research effort to develop a deeper neural network for end-to-end cell segmentation with higher accuracy. We observed that, like conventional CNNs, a popular deep neural network for end-to-end segmentation known as U-net (Ronneberger et al., 2015) is limited by gradient vanishing with increasing network depth. To address this problem, we propose two extensions of U-net: using residual layers in each level of the network and introducing summation-based skip connections to make the entire network much deeper. Our segmentation method produces an accurate result that is competitive with similar EM segmentation methods. The main contribution of this study can be summarized as follows:

- We introduce an end-to-end automatic EM image segmentation method using deep learning. The proposed method combines a variant of U-net and residual CNN with novel summation-based skip connections to make the proposed architecture, a *fully residual deep CNN*. This new architecture directly employs residual properties within and across levels, thus providing a *deeper* network with higher accuracy.
- We demonstrate the performance of the proposed deep learning architecture by comparing it with several EM segmentation methods listed in the leader board of the ISBI 2012 EM segmentation challenge (Arganda-Carreras et al., 2015). Our method outperformed many of the top-ranked methods in terms of segmentation accuracy.
- We introduce a *data enrichment* method specifically built for EM data by collecting all the orientation variants of the input images (eight in the 2D case, including all combinations of flipping and rotation). We used the same augmentation process for deployment: the final output is a combination of eight different probability values, which increases the accuracy of the method.
- We demonstrate the flexibility of the proposed method on two different EM segmentation tasks. The first involves cell membrane segmentation on a fruit fly (*Drosophila*) EM dataset (Arganda-Carreras et al., 2015). The second involves cell nucleus feature segmentation on a whole-brain larval zebrafish EM dataset (Hildebrand et al., 2017).

2 RELATED WORK

Deep neural networks (LeCun et al., 2015) have surpassed human performance in solving many complex visual recognition

problems. Systems using this method can flexibly learn to recognize patterns such as handwritten digits (Krizhevsky et al., 2012) in images with increasing layers hierarchically corresponding to increasing feature complexity (Zeiler and Fergus, 2014). A major drawback of using deep neural networks is that they often require a huge amount of training data. In order to overcome this issue, researchers have started to collect large databases containing millions of images that span hundreds of categories (Russakovsky et al., 2015). Largely thanks to such training datasets, many advanced architectures have been introduced, including VGG (Simonyan and Zisserman, 2014) and GoogleNet (Szegedy et al., 2015). With these architectures, computers are now able to perform even more complex tasks, such as transferring artistic styles from a source image to an unrelated target (Gatys et al., 2016). To leverage these new capabilities, researchers are actively working to extend deep learning methods for analyzing biomedical image data (Cicek et al., 2016). Developing such methods for automatic classification and segmentation of different biomedical image modalities, such as CT (Zheng et al., 2015) and MRI (Isin et al., 2016), is leading to faster and more accurate decision-making processes in laboratory and clinical settings.

Similarly, deep learning has been quickly adopted by connectomics researchers to enhance automatic EM image segmentation. One of the earliest applications to EM segmentation involved the straightforward application of a convolutional neural network (CNN) for pixel-wise membrane probability estimation (Ciresan et al., 2012), an approach that won the ISBI 2012 EM segmentation challenge (Arganda-Carreras et al., 2015). As more deep learning methods are introduced, automatic EM segmentation techniques evolve and new groups overtake the title of state-of-the-art performance in such challenges. One notable recent advancement was the introduction of a fully convolutional neural network (FCN) (Long et al., 2015) for end-to-end semantic segmentation. Inspired by this work, several modified FCNs have been proposed for EM image segmentation. One variant combined multi-level upscaling layers to produce a final segmentation (Chen et al., 2016a). Additional post-processing steps such as lifted multi-cut (Beier et al., 2016; Pape et al., 2019) further refined this segmentation result.

Another approach added skip connections for concatenating feature maps into a "U-net" architecture (Ronneberger et al., 2015). While U-net and its variants can learn multi-contextual information from input data, they are limited in the depth of the network they can construct because of the vanishing gradient problem. On the other hand, the addition of shortcut connections and direction summations (He et al., 2016) allows gradients to flow across multiple layers during the training phase. This creates a fully residual CNN where the architecture is a fusion of the U-net design and networks with summation-based skip connections, similar to Fully Convolutional Residual Networks (FC-ResNets) (Drozdzal et al., 2016) and Residual Deconvolutional Networks (RDN) (Fakhry et al., 2017). These related studies inspired us to propose a fully residual CNN for analyzing connectomics data.

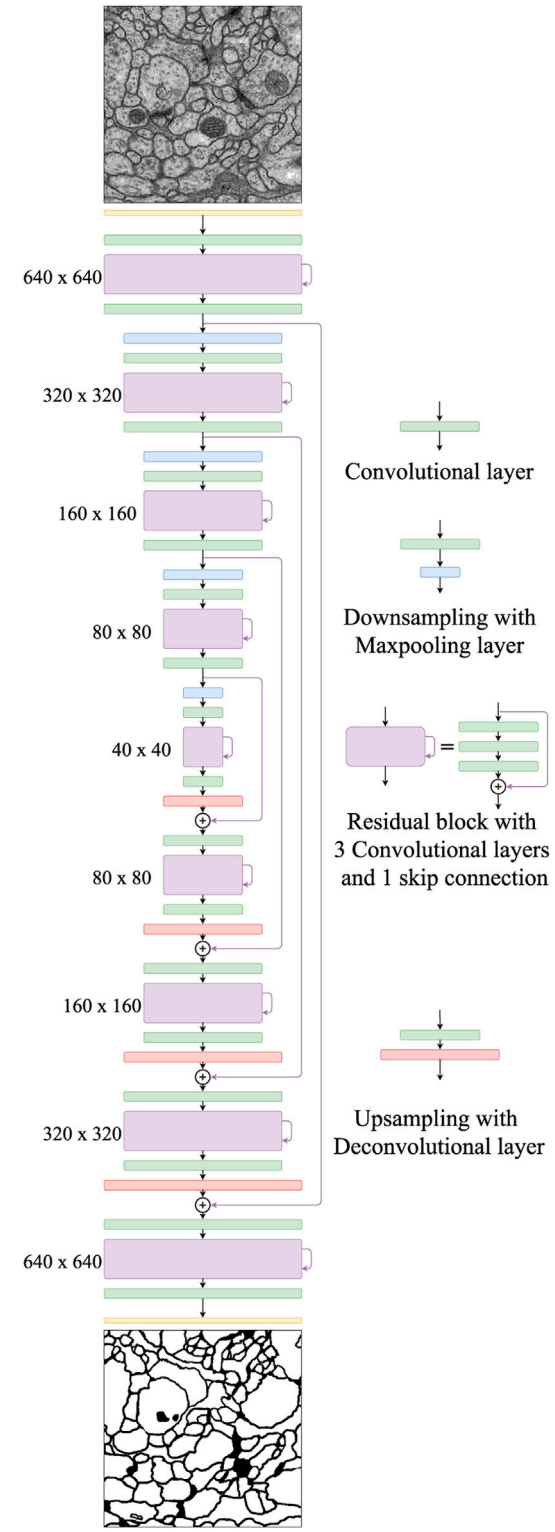


FIGURE 2 | The proposed FusionNet architecture. An illustration of the encoding path **(top to middle)** and the decoding path **(middle to bottom)**. Each intermediate residual block contains a residual skip connection within the same path, while the nested residual skip connections connect two different paths.

TABLE 1 | Architecture of the proposed network.

Block type	Ingredients	Size of feature maps
input		640 × 640 × 1
down 1	conv + res + conv	640 × 640 × 64
	+ maxpooling	320 × 320 × 64
down 2	conv + res + conv	320 × 320 × 128
	+ maxpooling	160 × 160 × 128
down 3	conv + res + conv	160 × 160 × 256
	+ maxpooling	80 × 80 × 256
down 4	conv + res + conv	80 × 80 × 512
	+ maxpooling	40 × 40 × 512
bridge	conv + res + conv	40 × 40 × 1024
up 4	deconv + merge +	80 × 80 × 512
	conv + res + conv	80 × 80 × 512
up 3	deconv + merge +	160 × 160 × 256
	conv + res + conv	160 × 160 × 256
up 2	deconv + merge +	320 × 320 × 128
	conv + res + conv	320 × 320 × 128
up 1	deconv + merge +	640 × 640 × 64
	conv + res + conv	640 × 640 × 64
output	conv	640 × 640 × 1

Work that leverages recurrent neural network (RNN) architectures can also accomplish this segmentation task (Stollenga et al., 2015). Instead of simultaneously considering all surrounding pixels and computing responses for the feature maps, RNN-based networks treat the pixels as a list or sequence with various routing rules and recurrently update each feature pixel. In fact, RNN-based membrane segmentation approaches are crucial for connected component labeling steps that can resolve false splits and merges during the post-processing of probability maps (Ensafi et al., 2014; Parag et al., 2015).

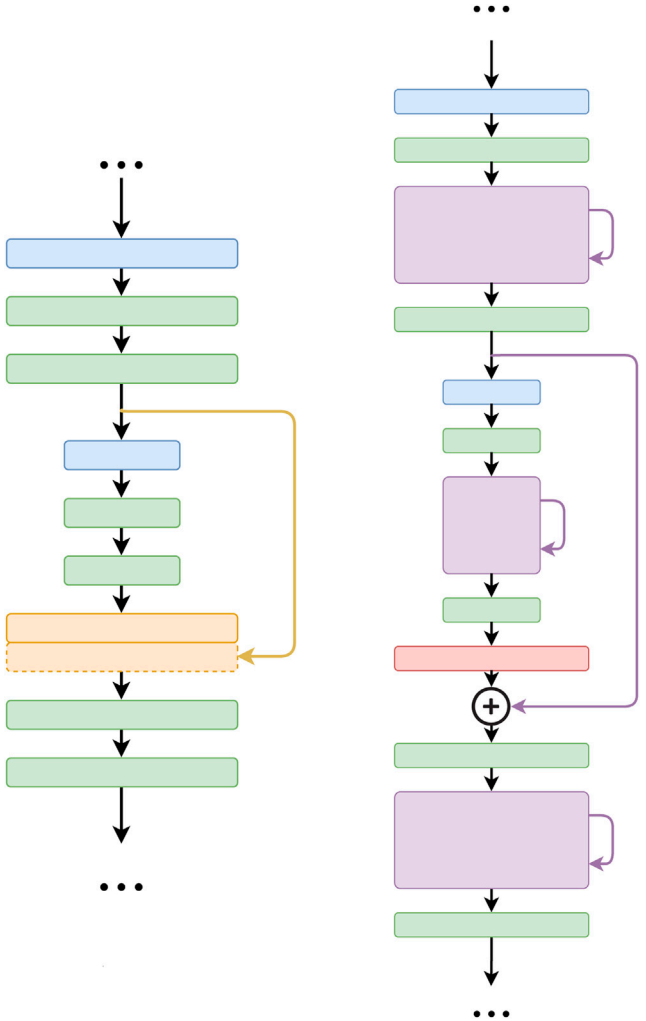


FIGURE 3 | Difference between the core connections of U-net (Ronneberger et al., 2015) **(left)** and FusionNet **(right)**. Note that FusionNet is a fully residual network due to the summation-based skip connections and is a much deeper network.

3 METHODS

3.1 Network Architecture

Our proposed network, FusionNet, is based on the architecture of a convolutional autoencoder and is illustrated in **Figure 2**. It consists of an encoding path (upper half, from 640 × 640 to 40 × 40) that retrieves features of interest and a symmetric decoding path (lower half, from 40 × 40 to 640 × 640) that accumulates the feature maps from different scales to form the segmentation. Both the encoding and decoding paths consist of multiple levels (i.e., resolutions). Four basic building blocks are used to construct the proposed network. Each *green block* is a regular convolutional layer followed by rectified linear unit activation and batch normalization (omitted from the figure for simplicity). Each *violet block* is a residual layer that consists of three convolutional blocks and a residual skip connection. Each *blue block* is a maxpooling layer located between levels only in the encoding path to perform downsampling for feature compression. Each *red block* is a deconvolutional layer located between levels only in the decoding path to upsample the input data using learnable interpolations. A detailed specification of FusionNet, including the number of feature maps and their sizes, is provided in **Table 1**.

One major difference between the FusionNet and U-net architectures is the way in which skip connections are used (**Figure 3**). In FusionNet, each level in the decoding path begins with a deconvolutional block (red) that un-pools the feature map from a coarser level (i.e., resolution), then merges it by pixel-wise *addition* with the feature map from the corresponding level in the encoding path by a *long* skip connection. There is also a *short* skip connection contained in each residual block (violet) that serves as a direct connection from the previous layer within the same encoding or decoding path. In contrast, U-net concatenates feature maps using only long skip connections. Additionally, by replacing concatenation with addition, FusionNet becomes a *fully* residual network, which resolves some common issues in deep networks

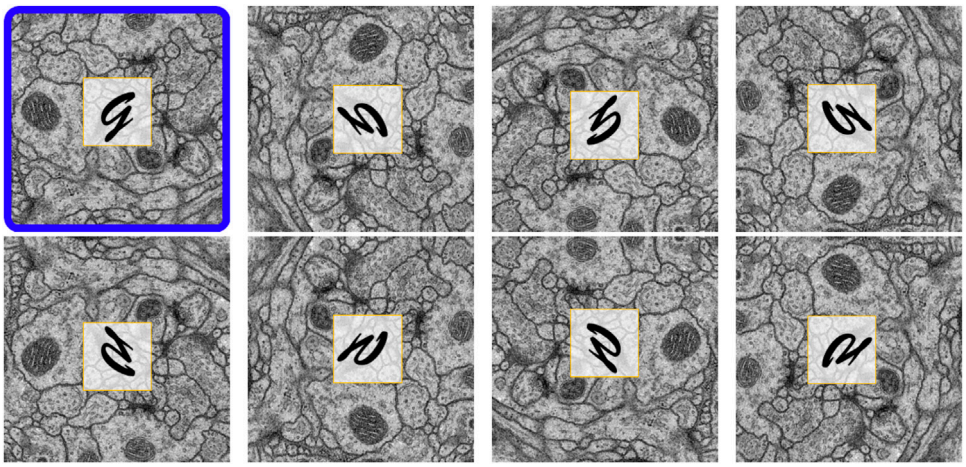


FIGURE 4 | Eight reoriented versions of the same EM image. The original image is outlined in blue. By adding these reoriented images, the input data size is increased by eight times.

(i.e., gradient vanishing). Furthermore, the nested short and long skip connections in FusionNet permit information flow within and across levels.

In the FusionNet encoding path, the number of feature maps doubles whenever downsampling is performed. After passing through the encoding path, the bridge level (i.e., 40 × 40 layer) residual block starts to expand feature maps into the following decoding path. In the decoding path, the number of feature maps is halved at every level, which maintains network symmetry. Note that there are convolutional layers both before and after each residual block. These convolutional layers serve as portal gateways that effectively adjust the amount of feature maps before and after residual blocks to the appropriate numbers. The placement of these convolutional layers on either side of the residual block leads the entire network to be perfectly symmetric (see **Figure 2**).

FusionNet performs end-to-end segmentation from the input EM data to the output segmentation label prediction. We train the network with pairs of EM images and their corresponding manually segmented label images as input. The training process involves comparing the output prediction with the input target labels using a mean-absolute-error (MAE) loss function to back-propagate adjustments to the connection weights. We considered the network sufficiently trained when its loss function values plateaued over several hundred epochs.

3.2 Data Augmentation

Our system involves data augmentation in multiple stages during both the training and deployment phases.

For training:

- The order of the image and label pairs are shuffled and organized with three-fold cross-validation to improve the generalization of our method.
- Offline, all training images and labels are reoriented to first produce an enriched dataset.
- Online, elastic field deformation is applied to both images and corresponding labels, followed by noise addition to only the images.

For prediction:

- Offline, input images are reoriented as for training.
- Inference is performed on all reoriented images separately, then each intermediate result is reverted to the original orientation, and all intermediate results are averaged to produce the final prediction.

Boundary extension is performed for all input images and labels. We describe each augmentation step in more detail in the following subsections.

Reorienation enrichment: Different EM images typically share similar orientation-independent textures in structures such as mitochondria, axons, and synapses. We reasoned that it should therefore be possible to enrich our input data with seven additional image and label pairs by reorienting the EM images, and in the case of training, their corresponding labels. **Figure 4** shows all eight orientations resulting from a single EM image after performing this data enrichment, with an overlaid letter "g" in each panel to provide a simpler view of the generated orientation. To generate these permutations, we rotated each EM image (and corresponding label) by 90°, 180°, and 270°. We then vertically reflected the original and rotated images. For training, each orientation was added as a new image and label pair. For prediction, inference was performed on each of these data orientations separately, then each prediction result was reverted to the original orientation before averaging to produce the final accumulation. Our intuition here is that, based on the equivariance of isotropic data, each orientation will contribute equally toward the final prediction result. Note that because the image and label pairs are enriched eight times by this process, other on-the-fly linear data augmentation techniques such as random rotation, flipping, or transposition are unnecessary.

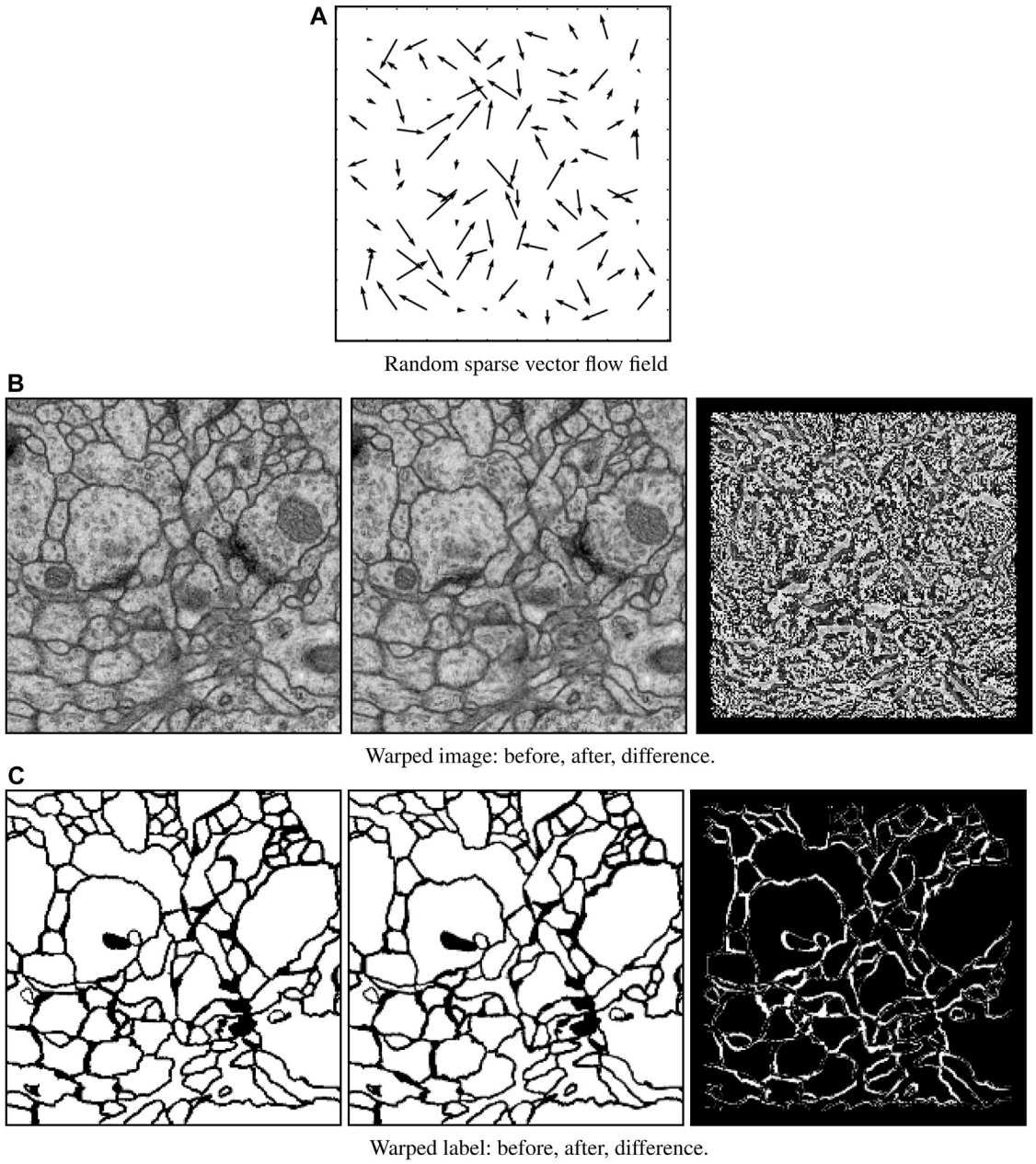


FIGURE 5 | Elastic field deformation example. A randomly sparse vector field **(A)** is generated for each training image and label pair. This sparse vector field is then used to warp both the original image data (**B**, left) and its corresponding label (**C**, left) to form an augmentation pair consisting of warped image data (**B**, middle) and warped label (**C**, middle). The difference between the original and warped images (**B**, right) and labels (**C**, right) show the effect of deformation.

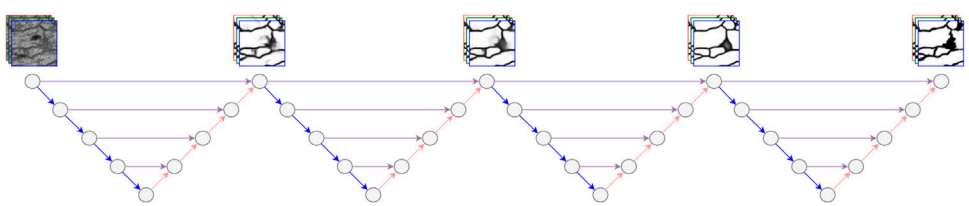


FIGURE 6 | FusionNetW^4, a chain of four concatenated FusionNet units.

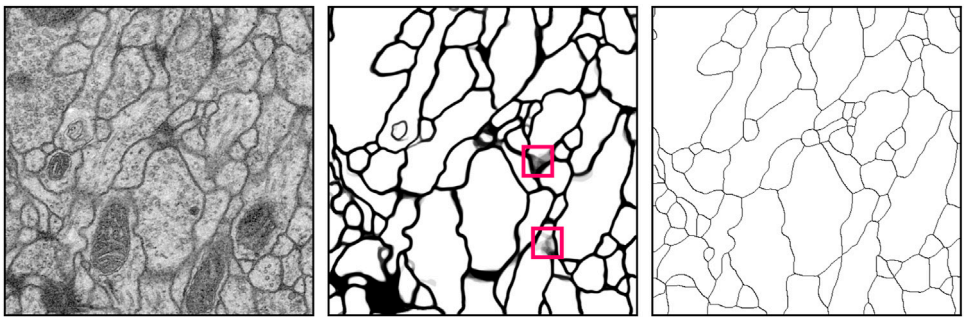


FIGURE 7 | Example results of cellular membrane segmentation on test data from the ISBI 2012 EM segmentation challenge (slice 22/30) illustrating an input EM image **(left)**, the probability prediction from FusionNetW$^2_{64}$ **(middle)**, and the thinned probability prediction after applying LMC (Beier et al., 2017) post-processing **(right)**. Pink boxes highlight uncertain regions that are ambiguous because of membrane smearing, likely due to anisotropy in the data.

TABLE 2 | Accuracy of various segmentation methods on the *Drosophila* EM dataset (ISBI 2012 EM segmentation challenge leaderboard, June 2020). Bold values correspond to the method presented here.

Methods	V_{rand}	V_{info}
Human values	0.997847778	0.998997659
PatchPerPix Hirsch et al. (2020)	0.988290649	0.991641507
IAL MutexWS Wolf et al. (2019)	0.987922250	0.991833594
CASIA MIRA Xiao et al. (2018)	0.987877739	0.990920188
IAL - SFCNN Weiler et al. (2017)	0.986800916	0.991438892
ACE-net Zhu et al. (2019)	0.985032746	0.989490497
M2FCN-MFA Shen et al. (2017)	0.983651122	0.991303595
FusionNetW$^2_{64}$ LMC	**0.983651122**	**0.991303595**
IAL MC/LMC Beier et al. (2017)	0.982616131	0.989461939
IAL LMC Beier et al. (2016)	0.982240005	0.988448278
FusionNetW$^2_{64}$	**0.981586186**	**0.990099898**
PolyMtl Drozdzal et al. (2016)	0.980582825	0.988163049
KUnet Chen et al. (2016b)	0.980222514	0.988967601
FusionNetW$^1_{64}$	**0.978042575**	**0.989945379**
IAL IC Lin et al. (2014)	0.977345721	0.989240736
Masters Wiehman and Villiers (2016)	0.977141154	0.987534429
CUMedVision Chen et al. (2016a)	0.976824580	0.988645822
ICNN Wu (2015)	0.976546913	0.988341665
DIVE-SCI Fakhry et al. (2016)	0.976229111	0.987392123
LSTM Stollenga et al. (2015)	0.975366444	0.987425430
U-net Ronneberger et al. (2015)	0.972760748	0.986616590

Elastic field deformation: To avoid overfitting (i.e., network remembering the training data), elastic deformation was performed on the entire enriched image dataset for every training epoch. This strategy is common in machine learning, especially for deep networks, to overcome limitations associated with small training dataset sizes. This procedure is illustrated in **Figure 5**. We first initialized a random sparse 12 × 12 vector field whose amplitudes at the image border boundaries vanish to zero. This field was then interpolated to the input size and used to warp both EM images and corresponding labels. The flow map was randomly generated for each epoch. No elastic field deformation was performed during deployment.

Random noise addition: During only the training phase, we randomly added Gaussian noise (mean $\mu = 0$, variance $\sigma = 0.1$) to each EM input image but not its corresponding label.

Boundary extension: FusionNet accepts an input image size of 512 × 512. Each input image, and in the case of training its corresponding label, was automatically padded with the mirror reflections of itself across the image border boundary (radius = 64 px) to maintain similar statistics for pixels that are near the edges. This padding is the reason why FusionNet starts with a 640 × 640 image, which is 128 px larger along each edge than the original input. However, we performed convolution with 3 × 3 kernel size and "SAME" mode, which leads the final segmentation to have the same padded size. To account for this, the final output prediction was cropped to eliminate the padded regions.

3.3 Experimental Setup

FusionNet was implemented using the Keras open-source deep learning library (Chollet, 2015). This library provides an easy-to-use, high-level programming API written in Python, with Theano or TensorFlow as a back-end engine. The model was trained with the Adam optimizer with a decaying learning rate of $2e^{-4}$ for over 50,000 epochs to harness the benefits of heavy elastic deformation on the small annotated datasets. FusionNet has also been translated to PyTorch and pure TensorFlow for other applications, such as Image-to-Image translation (Lee et al., 2018) and MRI reconstruction (Quan et al., 2018). All training and deployment presented here was conducted on a system with an Intel i7 CPU, 32 GB RAM, and a NVIDIA GTX GeForce 1080 GPU.

3.4 Network Chaining

FusionNet by itself performs end-to-end segmentation from the EM data input to the final prediction output. In typical real world applications of end-to-end segmentation approaches, however, manual proofreading by human experts is usually performed in an attempt to "correct" any mistakes in the output labels. We therefore reasoned that concatenating a chain of several FusionNet units could serve as a form of built-in refinement similar to proofreading that could resolve ambiguities in the initial predictions. **Figure 6** shows an example case with four chained FusionNet units (FusionNetW^4). To impose a target-driven approach across the chained network during training, we calculate the loss between the output of each separate unit and the training labels. As a result, chained FusionNet architectures have a single input and multiple outputs, where the end of each

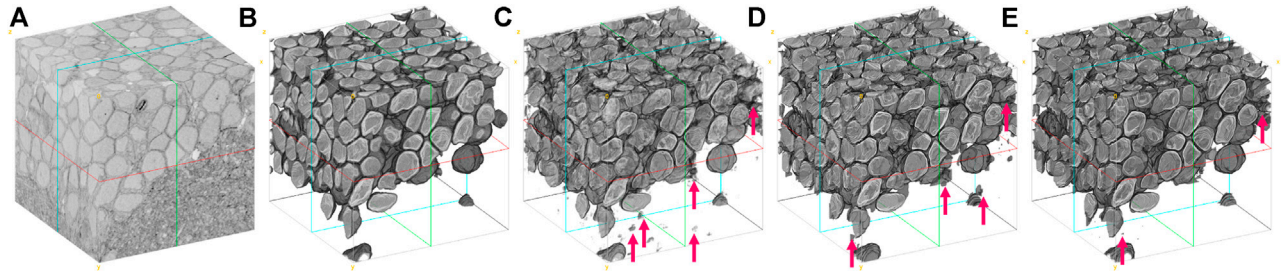


FIGURE 8 | Visual comparison of the larval zebrafish EM volume segmentation. **(A)** Input serial-section EM volume. **(B)** Manual segmentation (ground truth). **(C)** U-net (Ronneberger et al., 2015) result. **(D)** RDN (Fakhry et al., 2017) result. **(E)** FusionNetW$^4_{16}$ result. Red arrows indicate errors.

TABLE 3 | Segmentation accuracy on a test volume from the zebrafish EM dataset. Bold values correspond to the method presented here.

Methods	FusionNetW$^2_{64}$	RDN Fakhry et al. (2017)	U-net Ronneberger et al. (2015)
V_{rand}	**0.998648782**	0.991844302	0.987366177
V_{info}	**0.996929124**	0.994208722	0.992482059
V_{dice}	**0.963047248**	0.946099985	0.908491647

decoding path serves as a checkpoint between units attempting to produce better and better segmentation results.

Since the architecture of each individual unit is the same, the chained FusionNet model can be thought of as similar to an unfolded Recurrent Neural Network (RNN) with each FusionNet unit akin to a single feedback cycle but with weights that are *not* shared across cycles. Each FusionNet can be considered as a V-cycle in the multigrid method (Shapira, 2008) commonly used in numerical analysis, where the contraction in the encoding path is similar to *restriction* from a fine to a coarse grid, the expansion in the decoding is similar to the *prolongation* toward the final segmentation, and the skip connections play a role similar to *relaxation*. The simplest chain of two V-cycle units forms a W shape, so we refer to FusionNet chains using a "FusionNetW" terminology. To differentiate various configurations, we use the superscript to indicate how many FusionNet units are chained and the subscript to show the initial number of feature maps in the original resolution. For example, FusionNetW$^4_{64}$ would signify a network that chains four FusionNet units, each of them with the base number of convolution kernels (in Keras, nb_filters parameter) set to 64. We chose this specific 4-chain example case as the maximum chain length used here ad hoc to roughly match the memory available on our GPU. We also used 64 for the base number of convolution kernels in every case to match the backbone architecture of U-net. During training, the weights of each FusionNet unit ($\theta^k[i]$) are updated independently, as opposed to the RNN strategy of averaging the gradients from shared weights. For the example FusionNetW^4 case, we trained with the input images S and corresponding manual labels L. Each FusionNet unit in FusionNetW^4, which can be indexed as FusionNetW$^4[i]$ where i = 1, 2, 3 or 4, generates the prediction P[i] by minimizing the MAE loss between its prediction values and the target labels L. For each epoch, we incrementally train FusionNetW$^4[i]$ and fix its weights before training FusionNetW$^4[i+1]$. This procedure can be summarized as follows:

$$\begin{aligned}
&\min_{\theta^4[1]} \mathrm{MAE}(P[1], L) \ s.t. \quad P[1] = FusionNet^4[1](S)\\
&\min_{\theta^4[2]} \mathrm{MAE}(P[2], L) \ s.t. \quad P[2] = FusionNet^4[2](P[1])\\
&\min_{\theta^4[3]} \mathrm{MAE}(P[3], L) \ s.t. \quad P[3] = FusionNet^4[3](P[2])\\
&\min_{\theta^4[4]} \mathrm{MAE}(P[4], L) \ s.t. \quad P[4] = FusionNet^4[4](P[3])
\end{aligned} \tag{1}$$

The loss training curves decrease as i increases, eventually converging as the number of training epochs increases.

4 RESULTS

4.1 Fruit Fly Data

The fruit fly (*Drosophila*) ventral nerve cord EM data used here was captured from a first instar larva (Cardona et al., 2010). Training and test datasets were provided as part of the ISBI 2012 EM segmentation challenge[1] (Arganda-Carreras et al., 2015). Each dataset consisted of a 512 × 512 × 30 volume acquired at anisotropic $4 \times 4 \times \sim 50\,\mathrm{nm}^3\,\mathrm{vx}^{-1}$ resolution with transmission EM. These datasets were originally chosen in part because they contained noise and small image alignment errors that frequently occur in serial-section EM. For training, the provided dataset included EM image data and publicly available manual segmentation labels. The first 20 of 30 slices of the training volume were used for training and the last 10 slices were used for validation. For testing, the provided dataset included only EM image data, while segmentation labels were kept private for the assessment of segmentation accuracy (Arganda-Carreras et al., 2015). Test segmentations were produced for all 30 slices of the test volume and were then uploaded for comparison to the hidden ISBI Challenge segmentation labels.

Figure 7 illustrates the FusionNetW$^2_{64}$ probability map extraction results from test data without any post-processing steps (middle) and with lifted multi-cut (LMC) algorithm post-processing (right) (Beier et al., 2017), which resulted in thinning

[1]http://brainiac2.mit.edu/isbi_challenge/

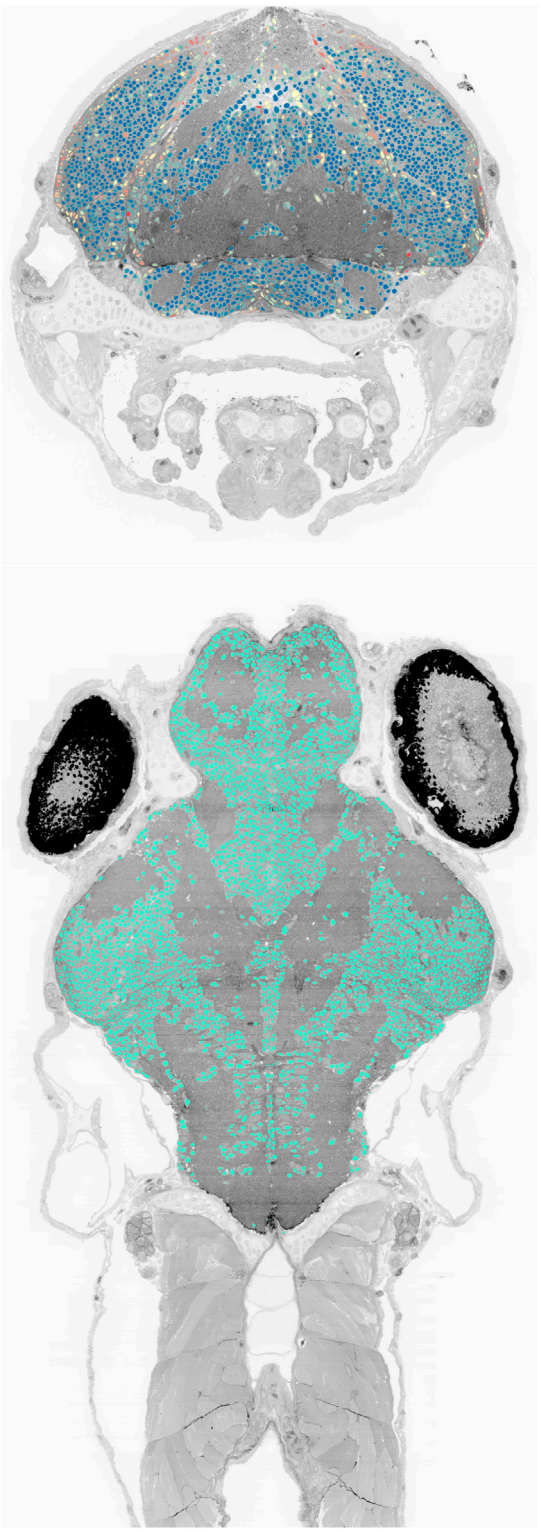


FIGURE 9 | Cell nucleus segmentation results overlaid onto zebrafish EM volume cross-sections through the transverse (**top**, blue to red color map varies with cell sphericity) and horizontal **(bottom)** planes.

of the probability map. As this shows, our chained FusionNet method is able to remove extraneous structures belonging to mitochondria (appearing as dark shaded textures) and vesicles (appearing as small circles). Uncertain regions in the prediction results without post-processing appear as blurry gray smears (highlighted by pink boxes). In cases like this, FusionNetW$^2_{64}$ must decide whether or not the highlighted pixels should be segmented as membrane, but the region is ambiguous because of membrane smearing, likely due to anisotropy in the data.

FusionNet approaches outperformed several other methods in segmenting the ISBI 2012 EM challenge data by several standard metrics. These metrics include foreground-restricted Rand scoring after border thinning (V_{rand}) and foreground-restricted information-theoretic scoring after border thinning (V_{info}) (Arganda-Carreras et al., 2015). Quantitative comparisons with other methods are summarized in **Table 2**. Even using a single FusionNet unit (FusionNetW$^1_{64}$), we achieved better results compared to many well-known methods, such as U-net (Ronneberger et al., 2015), network-in-network (Lin et al., 2014), fused-architecture (Chen et al., 2016a), and long short-term memory (LSTM) (Stollenga et al., 2015) approaches. Using a chained FusionNet with two modules (FusionNetW$^2_{64}$) performed even better, surpassing the performance of many previous state-of-the-art deep learning methods (Chen et al., 2016b; Drozdzal et al., 2016). These results confirm that chaining a deeper architecture with a residual bottleneck helps to increase the accuracy of the EM segmentation task. Both with and without LMC post-processing, FusionNetW$^2_{64}$ ranks among the top 10 in the ISBI 2012 EM segmentation challenge leaderboard (as of June 2020).

4.2 Zebrafish Data

The zebrafish EM data used here was taken from a publicly available database[2]. It was captured from a 5.5 days post-fertilization larval specimen. This specimen was cut into ~18,000 serial sections and collected onto a tape substrate with an automated tape-collecting ultramicrotome (ATUM) (Hayworth et al., 2014). A series of images spanning the anterior quarter of the larval zebrafish was acquired at 56.4 × 56.4 × ~ 60 nm^3 vx^{-1} resolution from 16,000 sections using scanning EM (Hildebrand, 2015; Hildebrand et al., 2017). All 2D images were then co-registered into a 3D volume using an FFT signal whitening approach (Wetzel et al., 2016). For training, two small sub-volume crops were extracted from a near-final iteration of the full volume alignment in order to avoid deploying later segmentation runs on training data. Two training volumes that contained different tissue features were chosen. One volume was 512 × 512 × 512 and the other was 512 × 512 × 256. The blob-like features of interest—neuronal nuclei—were manually segmented as area-lists in each training volume using the Fiji (Schindelin et al., 2012) TrakEM2 plug-in (Cardona et al., 2012). From each of these two training volumes, three

[2]http://zebrafish.link/hildebrand16/

quarters were used for training and one quarter was used for validation. These area-lists were exported as binary masks for use in the training procedure. For accuracy assessments, an additional non-overlapping 512 × 512 × 512 testing sub-volume and corresponding manual segmentation labels were used.

To assess the performance of FusionNetW$^4_{16}$ on this segmentation task, we first deployed it on 512 × 512 × 512 test volume alongside the U-net (Ronneberger et al., 2015) and RDN (Fakhry et al., 2017) methods. **Figure 8** displays volume renderings of the zebrafish test set EM data, its manual cell nucleus segmentation, and segmentation results from U-net, RDN, and FusionNetW$^4_{16}$. As this shows, FusionNetW$^4_{16}$ introduced less false predictions compared to U-net and RDN. **Table 3** compares U-net, RDN, and FusionNetW$^4_{16}$ using three quality metrics: foreground-restricted Rand scoring after border thinning (V_{rand}), foreground-restricted information theoretic scoring after border thinning (V_{info}), and the Dice coefficient (V_{dice}). By all of these metrics, FusionNetW$^4_{16}$ produced more accurate segmentation results.

We also deployed the trained network to the complete set of 16,000 sections of the larval zebrafish brain imaged at 56.4 × 56.4 × ~ 60 nm^3 vx^{-1} resolution, which is about 1.2 terabytes in data size. **Figure 9** shows EM dataset cross-sections in the transverse (top, *x-y*) and horizontal (bottom, *x-z*) planes of the larval zebrafish overlaid with the cell nucleus segmentation results. The transverse view overlay also shows the sphericity of each segmented cell nucleus in a blue to red color map, which can help to visually identify the location of false positives.

5 CONCLUSIONS

In this paper, we introduced a deep neural network architecture for image segmentation with a focus on connectomics EM image analysis. The proposed architecture, FusionNet, extends the U-net and residual CNN architectures to develop a deeper network for a more accurate end-to-end segmentation. We demonstrated the flexibility and performance of FusionNet in membrane- and blob-type EM segmentation tasks.

Several other approaches share similarities with FusionNet, particularly in concatenated chain forms. Chen et al. proposed concatenating multiple FCNs to build a RNN that extracts inter-slice contexts (Chen et al., 2016b). Unlike FusionNet, this approach takes as input multiple different resolutions of the raw image to produce a single segmentation output and uses a single loss function. Wu proposed iteratively applying a pixel-wise CNN (ICNN) to refine membrane detection probability maps (MDPM) (Wu, 2015). In this method, a regular CNN for generating MDPM from the raw input images and an iterative CNN for refining MDPM are trained independently. In contrast, FusionNet is trained as a single chained network. Additionally, FusionNet can refine errors in MDPM more completely using a chained network (i.e., by correcting errors in the error-corrected results) and scales better to larger image sizes due to the end-to-end nature of the network. More in-depth analyses into why chaining approaches are beneficial to improve the prediction accuracy of such deep networks will be an important goal for future work.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

TMQ developed the methods and performed the experiments with input from W-KJ and DGCH. Project supervision and funding were provided by W-KJ. All authors wrote the paper.

FUNDING

This work was supported by a NAVER Fellowship to TMQ and by a Leon Levy Foundation Fellowship in Neuroscience to DGCH.

ACKNOWLEDGMENTS

We thank Woohyuk Choi and Jungmin Moon for assistance in creating the larval zebrafish EM volume renderings. This manuscript has been released as a pre-print at https://arxiv.org/abs/1612.05360 (Quan et al., 2016).

REFERENCES

Arganda-Carreras, I., Turaga, S. C., Berger, D. R., Cireşan, D., Giusti, A., Gambardella, L. M., et al. (2015). Crowdsourcing the Creation of Image Segmentation Algorithms for Connectomics. *Front. Neuroanat.* 910, 142. doi:10.3389/fnana.2015.00142

Beier, T., Andres, B., Köthe, U., and Hamprecht, F. A. (2016). "An Efficient Fusion Move Algorithm for the Minimum Cost Lifted Multicut Problem," in Proceedings of ECCV 2016, Amsterdam, Netherlands, October 8–16, 2016 (Springer, Cham), 715–730.

Beier, T., Pape, C., Rahaman, N., Prange, T., Berg, S., Bock, D. D., et al. (2017). Multicut Brings Automated Neurite Segmentation Closer to Human Performance. *Nat. Methods* 14, 101–102. doi:10.1038/nmeth.4151

Bock, D. D., Lee, W.-C. A., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., et al. (2011). Network Anatomy and *In Vivo* Physiology of Visual Cortical Neurons. *Nature* 471, 177–182. doi:10.1038/nature09802

Briggman, K. L., and Bock, D. D. (2012). Volume Electron Microscopy for Neuronal Circuit Reconstruction. *Curr. Opin. Neurobiol.* 22, 154–161. doi:10.1016/j.conb.2011.10.022

Briggman, K. L., Helmstaedter, M., and Denk, W. (2011). Wiring Specificity in the Direction-Selectivity Circuit of the Retina. *Nature* 471, 183–188. doi:10.1038/nature09818

Cardona, A., Saalfeld, S., Preibisch, S., Schmid, B., Cheng, A., Pulokas, J., et al. (2010). An Integrated Micro- and Macroarchitectural Analysis of the Drosophila Brain by Computer-Assisted Serial Section Electron Microscopy. *PLoS Biol.* 8 (10), e1000502. doi:10.1371/journal.pbio.1000502

Cardona, A., Saalfeld, S., Schindelin, J., Arganda-Carreras, I., Preibisch, S., Longair, M., et al. (2012). TrakEM2 Software for Neural Circuit Reconstruction. *PLoS ONE* 7 (6), e38011. doi:10.1371/journal.pone.0038011

Chen, H., Qi, X., Cheng, J., and Heng, P. (2016). "Deep Contextual Networks for Neuronal Structure Segmentation," in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, February 12–17, 2016 (AAAI Press), 1167–1173.

Chen, J., Yang, L., Zhang, Y., Alber, M., and Chen, D. Z. (2016). "Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation," in Proceedings of NIPS 2016, Barcelona, Spain, September 5, 2016 (Curran Associates, Inc.), 3036–3044.

Chollet, F. (2015). Keras: Deep Learning Library for Theano and Tensorflow. Available at: http://keras.io/ (Accessed May 20, 2017).

Cicek, O., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in Proceedings of MICCAI 2016, Athens, Greece, October 17–21, 2016 (Springer, Cham), 424–432.

Ciresan, D., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2012). "Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images," in Proceedings of NIPS 2012, Stateline, NV, December 3–8, 2012 (Curran Associates Inc.), 2843–2851.

Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., and Pal, C. (2016). "The Importance of Skip Connections in Biomedical Image Segmentation," in Proceedings of DLMIA 2016, Athens, Greece, October 21, 2016 (Springer, Cham), 179–187.

Eberle, A. L., and Zeidler, D. (2018). Multi-Beam Scanning Electron Microscopy for High-Throughput Imaging in Connectomics Research. *Front. Neuroanat.* 12, 112. doi:10.3389/fnana.2018.00112

Ensafi, S., Lu, S., Kassim, A. A., and Tan, C. L. (2014). "3D Reconstruction of Neurons in Electron Microscopy Images," in Proceedings of IEEE EMBS 2014, Chicago, Illinois, August 27–31, 2014 (IEEE), 6732–6735.

Fakhry, A., Peng, H., and Ji, S. (2016). Deep Models for Brain EM Image Segmentation: Novel Insights and Improved Performance. *Bioinformatics* 32, 2352–2358. doi:10.1093/bioinformatics/btw165

Fakhry, A., Zeng, T., and Ji, S. (2017). Residual Deconvolutional Networks for Brain Electron Microscopy Image Segmentation. *IEEE Trans. Med. Imaging* 36, 447–456. doi:10.1109/tmi.2016.2613019

Gatys, L. A., Ecker, A. S., and Bethge, M. (2016). "Image Style Transfer Using Convolutional Neural Networks," in Proceedings of IEEE CVPR 2016, Las Vegas, NV, June 27–30, 2016 (IEEE), 2414–2423.

Graham, B. J., Hildebrand, D. G. C., Kuan, A. T., Maniates-Selvin, J. T., Thomas, L. A., Shanny, B. L., et al. (2019). High-throughput Transmission Electron Microscopy With Automated Serial Sectioning. doi:10.1101/657346 Preprint. Available at: https://doi.org/10.1101/657346 (Accessed June 2, 2019).

Hayworth, K. J., Morgan, J. L., Schalek, R., Berger, D. R., Hildebrand, D. G. C., and Lichtman, J. W. (2014). Imaging ATUM Ultrathin Section Libraries With WaferMapper: A Multi-Scale Approach to EM Reconstruction of Neural Circuits. *Front. Neural Circuits* 8, 68. doi:10.3389/fncir.2014.00068

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep Residual Learning for Image Recognition," in Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, June 27–30, 2016 (IEEE), 770–778.

Helmstaedter, M. (2013). Cellular-Resolution Connectomics: Challenges of Dense Neural Circuit Reconstruction. *Nat. Methods* 10, 501–507. doi:10.1038/nmeth.2476

Hildebrand, D. G. C. (2015). Whole-Brain Functional and Structural Examination in Larval Zebrafish. PhD thesis. Cambridge (MA): Harvard University, Graduate School of Arts and Sciences.

Hildebrand, D. G. C., Cicconet, M., Torres, R. M., Choi, W., Quan, T. M., Moon, J., et al. (2017). Whole-Brain Serial-Section Electron Microscopy in Larval Zebrafish. *Nature* 545, 345–349. doi:10.1038/nature22356

Hirsch, P., Mais, L., and Kainmueller, D. (2020). "Patchperpix for Instance Segmentation," in Proceedings of ECCV 2020, Glasgow, United Kingdom, August 23–28, 2020 (Springer, Cham), 288–304.

Isin, A., Direkoglu, C., and Sah, M. (2016). Review of MRI-Based Brain Tumor Image Segmentation Using Deep Learning Methods. *Procedia Comput. Sci.* 102, 317–324. doi:10.1016/j.procs.2016.09.407

Jeong, W.-K., Beyer, J., Hadwiger, M., Blue, R., Law, C., Vázquez-Reina, A., et al. (2010). Secrett and NeuroTrace: Interactive Visualization and Analysis Tools for Large-Scale Neuroscience Data Sets. *IEEE Comput. Graphics Appl.* 30, 58–70. doi:10.1109/MCG.2010.56

Kaynig, V., Vazquez-Reina, A., Knowles-Barley, S., Roberts, M., Jones, T. R., Kasthuri, N., et al. (2015). Large-scale Automatic Reconstruction of Neuronal Processes From Electron Microscopy Images. *Med. Image Anal.* 22, 77–88. doi:10.1016/j.media.2015.02.001

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "ImageNet Classification With Deep Convolutional Neural Networks, in Proceedings of NIPS 2012, Stateline, NV, December 3–8, 2012 (Curran Associates Inc.), 1097–1105.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature* 521, 436–444. doi:10.1038/nature14539

Lee, G., Oh, J.-W., Kang, M.-S., Her, N.-G., Kim, M.-H., and Jeong, W.-K. (2018). "DeepHCS: Bright-Field to Fluorescence Microscopy Image Conversion Using Deep Learning for Label-Free High-Content Screening, " in Medical Image Computing and Computer Assisted Intervention–MICCAI 2018, Granada, Spain, September 16–20, 2018 (Springer International Publishing), 335–343.

Lichtman, J. W., and Denk, W. (2011). The Big and the Small: Challenges of Imaging the Brain's Circuits. *Science* 334, 618–623. doi:10.1126/science.1209168

Lin, M., Chen, Q., and Yan, S. (2014). "Network in Network, " in Proceedings of ICLR 2014. arXiv:1312.4400v3.

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully Convolutional Networks for Semantic Segmentation," in Proceedings of IEEE CVPR 2015, Boston, MA, June 7–12, 2015 (IEEE), 3431–3440.

Pape, C., Matskevych, A., Wolny, A., Hennies, J., Mizzon, G., Louveaux, M., et al. (2019). Leveraging Domain Knowledge to Improve Microscopy Image Segmentation with Lifted Multicuts. *Front. Comput. Sci.* 1, 657. doi:10.3389/fcomp.2019.00006

Parag, T., Ciresan, D. C., and Giusti, A. (2015). "Efficient Classifier Training to Minimize False Merges in Electron Microscopy Segmentation," in Proceedings of IEEE ICCV 2015, Santiago, Chile, December 7–13, 2015 (IEEE), 657–665.

Quan, T. M., Hildebrand, D. G. C., and Jeong, W.-K. (2016). FusionNet: A Deep Fully Residual Convolutional Neural Network for Image Segmentation in Connectomics. arXiv preprint arXiv:1612.05360.

Quan, T. M., Nguyen-Duc, T., and Jeong, W.-K. (2018). Compressed Sensing MRI Reconstruction Using a Generative Adversarial Network With a Cyclic Loss. *IEEE Trans. Med. Imaging* 37, 1488–1497. doi:10.1109/tmi.2018.2820120

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation. "in Proceedings of MICCAI 2015, Munich, Germany, October 5–9, 2015 (Springer, Cham), 234–241.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vision* 115, 211–252. doi:10.1007/s11263-015-0816-y

Y. Shapira (Editor) (2008). *Matrix-Based Multigrid*. New York, NY: Springer US.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an Open-Source Platform for Biological-Image Analysis. *Nat. Methods* 9, 676–682. doi:10.1038/nmeth.2019

Shen, W., Wang, B., Jiang, Y., Wang, Y., and Yuille, A. (2017). "Multi-stage Multi-Recursive-Input Fully Convolutional Networks for Neuronal Boundary Detection," in IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October 22–29, 2017 (IEEE), 2391–2400.

Simonyan, K., and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556

Sommer, C., Strähle, C., Köthe, U., and Hamprecht, F. A. (2011). "Ilastik: Interactive Learning and Segmentation Toolkit," in Proceedings of IEEE ISBI 2011, Chicago, IL, March 30–April 2, 2011 (IEEE), 230–233.

Stollenga, M. F., Byeon, W., Liwicki, M., and Schmidhuber, J. (2015). "Parallel Multi-Dimensional LSTM, with Application to Fast Biomedical Volumetric Image Segmentation," in Proceedings of NIPS 2015, June 24, 2015, Montreal, QC (Curran Associates, Inc.), 2998–3006.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going Deeper with Convolutions," in Proceedings of IEEE CVPR, Boston, MA, June 7–12, 2015 (IEEE) 1–9.

Turaga, S. C., Murray, J. F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., et al. (2010). Convolutional Networks can Learn to Generate Affinity Graphs for Image Segmentation. *Neural Comput.* 22, 511–538. doi:10.1162/neco.2009.10-08-881

Weiler, M., Hamprecht, F. A., and Storath, M. (2017). "Learning Steerable Filters for Rotation Equivariant cnns," in Computer Vision and Pattern Recognition, Honolulu, HI, July 21–26, 2017 (IEEE), 849–858.

Wetzel, A. W., Bakal, J., Dittrich, M., Hildebrand, D. G. C., Morgan, J. L., and Lichtman, J. W. (2016). "Registering Large Volume Serial-Section Electron Microscopy Image Sets for Neural Circuit Reconstruction Using FFT Signal Whitening," in Proceedings of AIPR Workshop 2016, Washington, D.C., United States, October 18–20, 2016 (IEEE), 1–10.

Wiehman, S., and Villiers, H. D. (2016). "Semantic Segmentation of Bioimages Using Convolutional Neural Networks," in Proceedings of IJCNN 2016, Vancouver, BC, July 24–29, 2016 (IEEE), 624–631.

Wolf, S., Bailoni, A., Pape, C., Rahaman, N., Kreshuk, A., Köthe, U., et al. (2019). The Mutex Watershed and its Objective: Efficient, Parameter-Free Image Partitioning. *IEEE Trans. Pattern Anal. Mach. Intell.* doi:10.1109/TPAMI.2020.2980827

Wu, X. (2015). An Iterative Convolutional Neural Network Algorithm Improves Electron Microscopy Image Segmentation. arXiv preprint arXiv:150605849.

Xiao, C., Liu, J., Chen, X., Han, H., Shu, C., and Xie, Q. (2018). "Deep Contextual Residual Network for Electron Microscopy Image Segmentation in Connectomics," in IEEE 15th International Symposium On Biomedical Imaging (ISBI 2018), Washington, D.C., United States, April 4–7, 2018 (IEEE), 378–381.

Zeiler, M. D., and Fergus, R. (2014). "Visualizing and Understanding Convolutional Networks," in Proceedings of ECCV 2014, Zurich, Switzerland, September 6–12, 2014 (Springer, Cham), 818–833.

Zheng, Y., Liu, D., Georgescu, B., Nguyen, H., and Comaniciu, D. (2015). "3D Deep Learning for Efficient and Robust Landmark Detection in Volumetric Data," in Proceedings of MICCAI 2015, Munich, Germany, October 5–9, 2015 (Springer, Cham), 565–572.

Zheng, Z., Lauritzen, J. S., Perlman, E., Robinson, C. G., Nichols, M., Milkie, D., et al. (2018). A Complete Electron Microscopy Volume of the Brain of Adult drosophila Melanogaster. *Cell* 174, 730–743. doi:10.1016/j.cell.2018.06.019e22

Zhu, Y., Torrens, Y., Chen, Z., Zhao, S., Xie, H., Guo, W., and Zhang, Y. (2019). "Ace-net: Biomedical Image Segmentation with Augmented Contracting and Expansive Paths,"in International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, October 13–17, 2019 (Springer, Cham).712–720.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 26 May 2021
doi: 10.3389/fonc.2021.583921

Statistical Evaluation of Different Mathematical Models for Diffusion Weighted Imaging of Prostate Cancer Xenografts in Mice

*Harri Merisaari[1,2], Hanne Laakso[3], Heidi Liljenbäck[4], Helena Virtanen[4,5], Hannu J. Aronen[1,6], Heikki Minn[4,7], Matti Poutanen[5], Anne Roivainen[4,5], Timo Liimatainen[3,8,9,10] and Ivan Jambor[1,6]**

[1] Department of Radiology, University of Turku, Turku, Finland, [2] Turku Brain and Mind Center, University of Turku, Turku, Finland, [3] Department of Biotechnology and Molecular Medicine, A.I. Virtanen Institute for Molecular Sciences, Kuopio, Finland, [4] Turku PET Centre, University of Turku and Turku University Hospital, Turku, Finland, [5] Turku Center for Disease Modeling, University of Turku, Turku, Finland, [6] Medical Imaging Centre of Southwest Finland, Turku University Hospital, Turku, Finland, [7] Department of Oncology and Radiotherapy, Turku University Hospital, Turku, Finland, [8] Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Oulu, Finland, [9] Department of Clinical Radiology, Oulu University Hospital, Oulu, Finland, [10] Department of Radiology, University of Oulu, Oulu, Finland

OPEN ACCESS

Edited by:
Concetto Spampinato, University of Catania, Italy

Reviewed by:
Chris Albanese, Georgetown University, United States
Carmelo Pino, University of Catania, Italy

***Correspondence:**
Ivan Jambor
ivjamb@utu.fi

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 22 July 2020
Accepted: 23 March 2021
Published: 26 May 2021

Citation:
Merisaari H, Laakso H, Liljenbäck H, Virtanen H, Aronen HJ, Minn H, Poutanen M, Roivainen A, Liimatainen T and Jambor I (2021) Statistical Evaluation of Different Mathematical Models for Diffusion Weighted Imaging of Prostate Cancer Xenografts in Mice. Front. Oncol. 11:583921. doi: 10.3389/fonc.2021.583921

Purpose: To evaluate fitting quality and repeatability of four mathematical models for diffusion weighted imaging (DWI) during tumor progression in mouse xenograft model of prostate cancer.

Methods: Human prostate cancer cells (PC-3) were implanted subcutaneously in right hind limbs of 11 immunodeficient mice. Tumor growth was followed by weekly DWI examinations using a 7T MR scanner. Additional DWI examination was performed after repositioning following the fourth DWI examination to evaluate short term repeatability. DWI was performed using 15 and 12 b-values in the ranges of 0-500 and 0-2000 s/mm^2, respectively. Corrected Akaike information criteria and F-ratio were used to evaluate fitting quality of each model (mono-exponential, stretched exponential, kurtosis, and bi-exponential).

Results: Significant changes were observed in DWI data during the tumor growth, indicated by ADC_m, ADC_s, and ADC_k. Similar results were obtained using low as well as high b-values. No marked changes in model preference were present between the weeks 1–4. The parameters of the mono-exponential, stretched exponential, and kurtosis models had smaller confidence interval and coefficient of repeatability values than the parameters of the bi-exponential model.

Conclusion: Stretched exponential and kurtosis models showed better fit to DWI data than the mono-exponential model and presented with good repeatability.

Keywords: diffusion weighted imaging, PC-3 xenograft prostate tumors, prostate cancer mouse model, repeatability, Akaike information criteria (AIC), F-ratio

INTRODUCTION

Diffusion weighed imaging (DWI) has extensively been used for cancer characterization in both pre-clinical (1, 2) and clinical settings (3) during the last decade. Furthermore, DWI is increasingly being used for monitoring cancer therapy responses (4). In biological tissue, DWI contrast is predominantly affected by microscopic motion of water molecules and water interactions with surroundings. The most recognized DWI imaging acquisition method is the Stejskal–Tanner pulsed field gradient method. With this method, motion caused by self diffusion of a proton is acquired by applying a pair of motion-encoding gradients. The first gradient dephases and second one rephrases stationary protons, while moving water protons stays dephased resulting to decreased signal intensity. The signal attenuation depends on water diffusion coefficient (D [mm^2/s]) as well as direction of the self diffusion of water (5).

Several different mathematical models have been proposed to describe the DWI signal decay. The mono-exponential model is the simplest and widely used, in which one parameter D (or often the apparent diffusion coefficient, ADC) describes the diffusion. This model fits well to DWI data measured from pure water without any restrictions. At low b-values DWI signal decay deviates from the mono-exponential function due to presence of intra-voxel incoherent motion (IVIM), as originally proposed by Le Bihan and co-workers (6, 7). The search for new non-invasive imaging markers has led to increased interests in IVIM-derived parameters, which have demonstrated correlation with microvessel density in colorectal cancer (HT29) model (8). Nevertheless, IVIM-derived parameters are not directly related to tissue perfusion (6, 7), but perfusion and IVIM-derived parameters are rather related to the capillary structure (9, 10). Similarly to low b-values, DWI signal decay at high b-values deviates from the mono-exponential function, and it is better described by non-Gaussian mathematical models (11–14).

In general, a single mono-exponential decay provides oversimplified description of the complicated water motion in the tissue. However, the modeling with several free parameters could lead to "over-fitting" of data, and poor repeatability of the fitted parameters. Optimal model would have the highest information content and provides independent parameters, which are related to physical quantities (e.g. cell and/or vessel density) while still retain high repeatability/reliability of fitted parameters.

The Akaike information criteria (AIC) has been widely used for model selection in previous studies (15–17). A model with smaller AIC value would be a preferred model due to less information loss as compared with a model presenting with higher AIC. Similarly to AIC, F-ratio is commonly being applied for model selection (18). Model selection based on F-ratio tends to prefer a more simplified model (19) in contrast to AIC.

In the current study, we evaluated four different mathematical models for DWI within a study applying PC-3 prostate cancer cells grown in immunodeficient mice using both low (0-500 s/mm^2) and high b-values (0-2000 s/mm^2). The tumor growth was followed for four weeks with repeated MR examinations performed once a week. Corrected Akaike information criteria (AIC_c) and F-ratio were used to evaluate information content of the models. Non-Gaussian DWI models provided better fit to DWI data obtained using both low and high b-values. However, non-Gaussian DWI models were not clearly preferred over the mono-exponential model for DWI data obtained using low b-values in contrast to DWI data obtained using high b-values. Furthermore, DWI fitted parameters changed significantly during tumor progression.

MATERIAL AND METHODS

Animal Tumor Model

One million PC-3 (Anticancer Inc., USA) human prostate cancer cells were inoculated subcutaneously in immunodeficient mice (n=11, HSD: Athymic Nude Foxn 1nu, Harlan Laboratories, Indianapolis, IN, USA). The cells also expressed red florescent protein, while this property was not applied in the present study. Mice were housed in individually ventilated cages under controlled conditions of light (12h light/12h dark), temperature (21 ± 3°C), and humidity (55% ± 15%) in specific pathogen-free conditions at the Central Animal Laboratory, University of Turku for the first 5 days, and thereafter in similar conditions at the University of Eastern Finland Kuopio campus. Mice were provided with irradiated soy-free natural-ingredient feed (RM3 (E), Special Diets Services, Essex, UK) and autoclaved tap water *ad libitum*, and were housed complying with international guidelines on the care and use of laboratory animals. All animal handling was conducted in accordance with the Finnish Committee for the use and care of laboratory animals and the institutional animal care policies, which fully meet the requirements as defined in the U.S. National Institutes of Health guidelines on animal experimentation.

MR Imaging

The first MR examination was performed 8 days after cell implantation. Tumor growth was followed for four weeks with repeated MR examinations once a week. Immediately following the fourth MR examination, six and seven mice had repeated DWI scan performed using low and high b-values, respectively. The second repetition was performed following animal and coil repositioning approximately 60 minutes after the first set of DWI. The repeated DWI examinations were used to evaluate short term repeatability of the measured parameters. The anesthetized mice (1.5% isoflurane in $70\%N_2/30\%O_2$) were imaged using a 7T animal MR scanner (7T Pharmascan, Bruker GmbH, Ettlingen, Germany) with 72 mm volume transmitter (Bruker GmbH) and 10 mm surface receiver coil (Bruker GmbH). Multislice T_2-weighted anatomical images covering the whole tumor area were obtained (TR/TE 2500 ms/33 ms, field of view (FOV) = 30 × 30 mm^2, matrix size 256 × 256, 15 slices) to localize a slice with maximum tumor diameter for DWI measurements. Diffusion weighted single shot spin-echo echo planar imaging was applied with the parameters: TR/TE 3750/25.3 (low b-value set) 3000/30 ms (high b-value set), FOV 3 × 1.5 cm^2, matrix 128 ×64, slice thickness 1 mm, three orthogonal diffusion directions, and

two different sets of b-values: low b-value set (15 b-values in total): 0, 2, 4, 6, 9, 12, 14, 18, 23, 25, 28, 50, 100, 300, 500 s/mm^2, and high b-value set (12 b-values in total): 0, 100, 300, 500, 700, 900, 1100, 1300, 1500, 1700, 1900, 2000 s/mm^2. For further analysis, the mean value of the signal from three directions was calculated.

Data Modeling

The following four mathematical models were applied to the DWI signal obtained using low and high b-values:

1. Mono-exponential model:

$$S(b) = S(0)e^{-bADC_m} \quad [1]$$

where b is the b-value, S(0) is the signal intensity at b-value of 0 s/mm^2, and ADC$_m$ is the apparent diffusion coefficient calculated using the mono-exponential model.

2. Stretched exponential model also known as Kohlrausch-Williams-Watts model (20):

$$S(b) = S(0)e^{-(bADC_s)^{\alpha}} \quad [2]$$

where ADC$_s$ is the apparent diffusion coefficient calculated using the stretched exponential model, and α is the heterogeneity index. The dimensionless α parameter varies from 0 to 1. During the fitting procedure, α parameter was constrained to be in the range of 0 to 1.

3. Kurtosis model:

$$S(b) = S(0)e^{(-bADC_k+\frac{1}{6}b^2ADC_k^2K)} \quad [3]$$

where ADC$_k$ is the apparent diffusion coefficient calculated using the kurtosis model, and K is the kurtosis. Jensen at al. (21) originally developed the kurtosis model to fit deviation of diffusion tensor signal from the mono-exponential function. The dimensionless positive K parameter characterizes the deviation from the mono-exponential signal decay.

4a. Bi-exponential model for low b-values:

$$S(b) = S(0)(1-f_p)e^{-bD_f} + f_p e^{-bD_p} \quad [4]$$

where f$_p$ is the "pseudodiffusion" fraction, D$_f$ is the fast diffusion coefficient, and D$_p$ is the "pseudodiffusion" coefficient. The intravoxel incoherent motion (IVIM) theory is an advanced method to separate diffusion and perfusion effects using DWI (6) at low b-values. According to the IVIM theory, the blood flow in the capillaries causes a dephasing of the magnetization when motion-encoding gradients are applied. This means that the motion of water molecules due to microcirculation of blood in the capillaries has a similar effect on the resulting DWI signal as their motion due to molecular diffusion.

4b. Bi-exponential model for high b-values:

$$S(b) = S(0)(1-f_f)e^{-bD_s} + f_f e^{-bD_f} \quad [5]$$

where f$_f$ is the fraction of fast diffusion, D$_f$ is the fast diffusion coefficient, and D$_s$ is the slow diffusion coefficient.

The DWI signal decay of each individual voxel has been fitted using four mathematical models, as described above, to generate parametric maps of the parameters. The fitting procedure has been performed using in-house written C++ code utilizing Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (22) in dlib library (23).

Following multiple initializations values were used to prevent local minima in the fitting procedure in order to avoid local minima in the fitting procedure (initializations values for high b-values data are in brackets):

1. Mono-exponential:

ADC_m – from 0.001 (0.001) to 0.003 (0.003) with the step size of 0.0005 (0.0005)

2. Stretched exponential:

ADC_s – from 0.001 (0.001) to 0.003 (0.003) with the step size of 0.0005 (0.0005)

α – from 0.1 (0.1) to 1.0 (1.0) with the step size of 0.15 (0.15)

3. Kurtosis:

ADC_k – from 0.001 to 0.003 with the step size of 0.0003

K – from 0.0001 (0.0001) to 4.0 (2.0) with the step size of 0.05 (0.02)

4a (b). Biexponential:

$f_p(f_f)$ – from 0.0 (0.5) to 1.0 (1.0) with the step size of 0.1 (0.1)

$D_p(D_f)$ – from 0.0001 (0.0001) to 0.04 (0.003) with the step size of 0.0005 (0.0003)

$D_f(D_s)$ – from 0.0001 (0.00002) *to* 0.003 (0.001) with the step size of 0.0003 (0.00005)

Image Analysis

The tumor area was manually delineated on T2-weighted anatomical images and the regions of interest (ROIs) were defined to the corresponding parametric images. Voxels with ADC$_m$ values higher than 8.0^10^{-3} s^2/mm were discarded as those voxels were considered to represent necrosis. Median values of the fitted parameters of each ROI between repeated scans were compared using one-way analysis of variance with Bonferroni test ($p<0.05$ statistically significant).

EVALUATION OF FITTING QUALITY

Corrected Akaike information criteria difference (ΔAIC_c) (15) was used to evaluate model fit to DWI data of each individual voxel:

$$\Delta AIC_c = N\left(\ln\left(\frac{SS_B}{N}\right) - \ln\left(\frac{SS_A}{N}\right)\right) + 2(P_B - P_A) + 2\left(\frac{P_B(P_B+1)}{N-P_B-1} - \frac{P_A(P_A+1)}{N-P_A-1}\right) \quad [6]$$

where N is the sample size, P is the number of parameters, SS is the sum of squares between data points and fitted curve,

A subscript represents the simpler model, B subscripts represents the more complex model.

In addition to ΔAIC_c, F-ratio (F) with 1% level of significance was used to evaluate model fit to data:

$$F = \frac{(SS_A - SS_B)/SS_B}{(DF_A - DF_B)/DF_B} \quad [7]$$

where DF (= number of data points − number of parameters) is the degree of freedom, A subscript represents the simpler model, B subscripts represents the more complex model.

REPEATABILITY OF THE FITTED PARAMETERS

Repeatability of the fitted parameters was evaluated using the same methodology as in a previous human DWI drug intervention study (24). The difference (d) in median values per ROI between two repeated scans performed 4 weeks after the initial scan was calculated for a subset of mice (six mice for low b-values and seven for high b-values). Mean squared difference (msd) was calculated as follows:

$$msd = \sqrt{\sum_{i=1}^{n} d^2 \times (n-1)^{-1}} \quad [8]$$

where d is the difference between two repeated scans, n is the number of mice with repeated scan. Subsequently, 95% confidence interval (CI) for changes in the study cohort was calculated:

$$CI = \pm 1.96 \times msd/\sqrt{n} \quad [9]$$

where msd is the mean squared difference, n is the number of mice with repeated scan.Finally, coefficient of repeatability (r) was calculated as follows:

$$r = 1.96 \times msd \quad [10]$$

RESULTS

PC-3 cancer cell growth was followed for 4 weeks with repeated MR examinations performed once a week. The changes in diffusion parameters are visualized in parametric maps on top of T2-weighted images for a representative case (**Figures 1** and **2**) while the rest of data are shown in Supplementary material (**Figures S1**–**S20**). Median signal intensity of tumor ROI and the correcting fitted curves of week 1 and week 4 are shown in **Figure 3**. The data is shown for the same representative tumor as shown in **Figure 1**.

Median values of the fitted parameters ADC_m, ADC_s, ADC_k, and D_s between week 1 and those measured at weeks 2, 3 and 4 differed significantly while the differences between weeks 2, 3 and 4 were not significant using the low b-value data. Similarly, median values of K parameter increased significantly between week 1 and week 3 and 4. In contrast, no significant differences were present between median values of different weeks for α, D_p, and f_f parameters (**Figure 4**). Using high b-values, significant changes were present in median values of ADC_m, ADC_s, α, and ADC_k between week 1 and weeks 2, 3 and 4, while differences between week 2, 3 and 4 were not significant. The changes in K parameter were significant only between week 1 and week 4, while differences between week 1 and weeks 3 and 4 were significant for D_f and D_s parameters (**Figure 5**).

Model Selection

In general, DWI data obtained using low b-values fitted better by the stretched exponential model as compared with mono-exponential, kurtosis and bi-exponential models based on AIC_c and F-ratio. In more than 50% of voxels the kurtosis model had lower AIC_c values than the mono-exponential model. However, the kurtosis model did not provide significantly better fit to data than the mono-exponential model in more than 50% of voxels (averaged medians of 11 mice) based on F-test. Similarly, the bi-exponential model was not preferred over mono-exponential in more than 50% of voxels (averaged medians of 11 mice). No dramatic changes in model preference were present between different time points (**Tables 1** and **2**).

In contrast to low b-values, in vast majority of voxels stretched exponential, kurtosis and bi-exponential models fitted DWI data obtained using high b-values better than the mono-exponential model based on AIC_c and F-test. The kurtosis model was preferred over the stretched exponential model in average in ~75% of voxels based on AIC_c. The bi-exponential models still provided significantly better fit to data than the stretched exponential and kurtosis models based on F-test.

Repeatability of the Fitted Parameters

The parameters of mono-exponential, stretched exponential and kurtosis models had confidence interval values smaller than 25% of the corresponding averaged median values (**Table 3**) for DWI data obtained using both low as well as high b-values. Similarly, coefficients of repeatability were smaller than 45% of the corresponding averaged median values (**Table 3**), with the exception of K parameter for low b-value DWI data (r% 59.7%). In contrast, the parameters of the bi-exponential model had much larger confidence interval and coefficient of repeatability values, especially for low b-value DWI data. Large confidence interval and coefficient of repeatability values for the parameters of bi-exponential model implicate poor measurement repeatability. Confidence interval and coefficient of repeatability values for mono-exponential, stretched exponential and kurtosis models we similar for DWI data acquired using low and high b-values. However, K parameter of kurtosis models had approximately 2-times higher relative coefficient of repeatability values than the ADC_m, ADC_s, ADC_k parameters for low as well as high b-values.

Signal intensities at the second repeated DWI examination differed systematically from those measured at the first DWI examination, and the median values of ADC_m, ADC_s, and ADC_k parameters were lower in the repeated DWI in all mice.

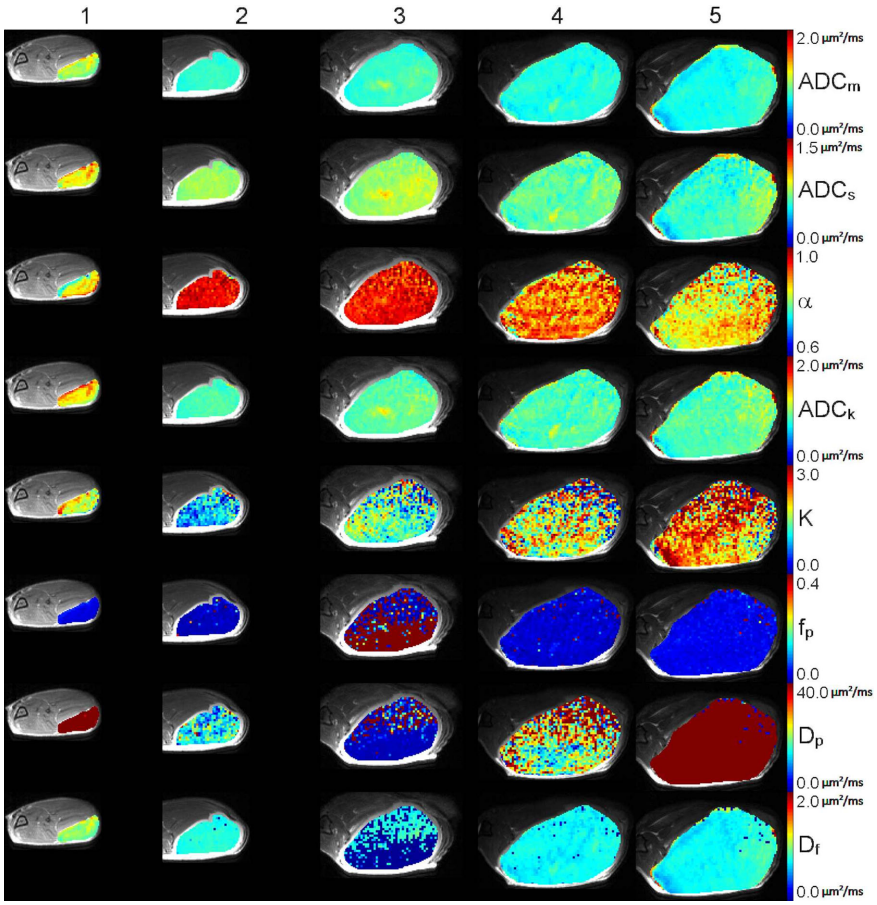


FIGURE 1 | Low b-value DWI data of a representative tumor: T2-weighted image fused with parametric maps for ADC_m, ADC_s, α, ADC_k, K, f_p, D_p, and D_f parameters are shown, and represent different degree of tumor homogeneity between week 1 (column 1), week 2 (column 2), week 3 (column 4) and week 4 (column 1). Furthermore, the second repeated imaging on week 4 is shown (column 5). The parametric maps are scaled as follows: ADC_m, min–max: 0–2.0 µm²/ms; ADC_s, min–max: 0–1.5 µm²/ms; α, min–max: 0.6–1.0; ADC_k, min–max: 0–2.0 µm²/ms; K, min–max: 0–3.0; f_p, min–max: 0–0.4; D_p, min–max: 0–40.0 µm²/ms; D_f, min–max: 0–2.0 µm²/ms.

DISCUSSION

The use of DWI for cancer detection, characterization and cancer therapy response monitoring continues to increase in both pre-clinical and clinical settings. Despite wide use of DWI, accurate and robust modeling of DWI signal decay remains a challenge. In the current study, we have evaluated four different mathematical models for DWI data (low and high b-values) of PC-3-cell derived human prostate cancer xenografts in mice, in terms of fitting quality and repeatability. Significant changes were observed in median values of ADC parameters detected between week 1 and those measured 1-3 weeks later, while the difference between the values obtained at weeks 2-4 were not significant. In previous studies cell density was shown to correlate with ADC_m parameter (25, 26). In the light of these previous studies, our findings indicated lower cell density of the tumors at the first time point measured.

According to our finding, non-Gaussian DWI models provided better fit to DWI data than the most commonly applied mono-exponential model, which is in line with previous findings (6, 16, 21, 27). The use of high b-values and non-Gaussian DWI models for early therapy response evaluation has demonstrated promising results in human brain tumor (28), colon cancer mouse model (29), and glioma mouse model (30). In a study by Hoff and co-workers (30), fast diffusion component of the bi-exponential model had the largest percent change from baseline in glioma mouse model, suggesting a role of non-Gaussian DWI models for prediction of early therapy response.

Several recent pre-clinical (31–33) and clinical (34–36) studies demonstrated promising results for IVIM derived parameters especially in organs with high perfusion, such as liver or kidney. In these organs the intravoxel incoherent motion present with relatively larger contribution to signal decay. The perfusion fraction was shown to be in the range of > 20-30%, being substantially more than that observed in brain, for example (37). In the current study, the averaged median perfusion fraction value was 9%, thus, being similar to human brain. Due to relatively low contribution of the intravoxel incoherent motion

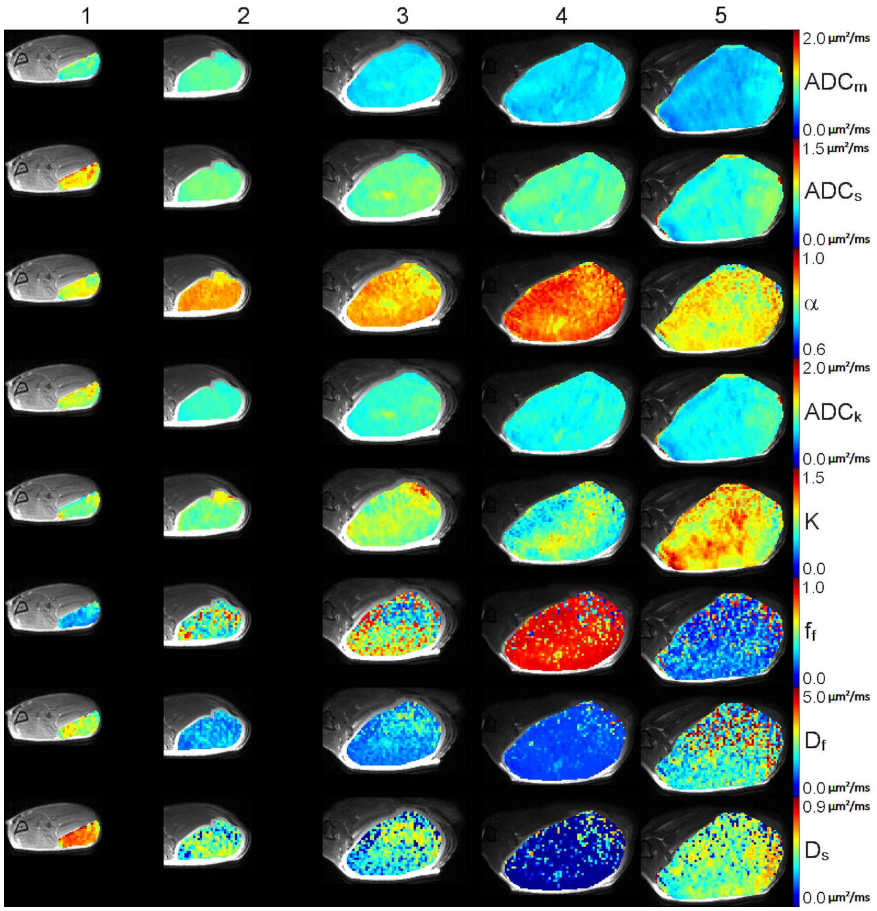


FIGURE 2 | High b-value DWI data of a representative tumor: T2-weighted image fused with parametric maps for ADC_m, ADC_s, α, ADC_k, K, f_f, D_f, and D_s parameters are shown, and represent different degree of tumor homogeneity between week 1 (column 1), week 2 (column 2), week 3 (column 3), week 4 (column 4). Furthermore, the second repetition done on week 4 is shown (column 5). The parametric maps are scaled as follows: ADC_m (A1–5) min–max: 0–2.0 µm²/ms, ADC_s (B1–5) min–max: 0–1.5 µm²/ms, α (C1-5) min–max: 0.6–1.0, ADC_k (D1–5) min–max: 0–2.0 µm²/ms, K (E1–5), min–max: 0–1.5, f_f (F1–5), min–max: 0–1.0, D_f (G1–5) min–max: 0–5.0 µm²/ms, D_s (H1–5) min–max: 0–0.9 µm²/ms.

fraction to the measured signal, it is questionable how accurately a least square fitting procedure performed independently for each voxel (in a presence of measurement and physiological noise) can evaluated such small exponential component of the bi-exponential model. Despite preventing local minima in the fitting procedure, the repeatability of IVIM derived parameters (bi-exponential model for low b-values) was low. Coefficient of repeatability (expressed in % of averaged median values) for f_p, D_p and D_f were 491.8%, 399.1% and 163.2%, respectively. Similar to our study, IVIM derived parameters using least square fitting have shown low reproducibility in human liver (38). The small contribution of "pseudodiffusion" component (f_p, D_p) to the final fitting residuals during least square fitting procedure, calls into question the validity of IVIM parameters that are estimated using least square fitting procedure for organs with small "pseudodiffusion" component (39–41). Efforts to increase fitting robustness of the bi-exponential model resulted into wider use of "segmented analysis" (32, 42) where each exponential component is being fitted individually in subsequent fashion. However, it should be noted that the resulting fitting residual is likely to be higher for "segmented analysis" than for "simultaneous" least square fitting of the bi-exponential model. Information content and bias of "segmented analysis" in comparison with other mathematical models remains to be established. Orton and the co-workers (43) have proposed the use of a Bayesian approach for improved estimation accuracy of IVIM parameters. Bayesian approach shrinks the distribution of parameters and "moves" outliers closer to the central distribution. Despite very promising results (36), this approach might not be applicable for cases with limited number of fitting voxels with large physiological voxel heterogeneity. Bayesian approach is a balance between improving quality of parametric maps and suppressing heterogeneity. Other possible approach is the use of neighborhood information during the fitting procedure (44).

Deviation of the DWI signal from mono-exponential decay at low b-values is, according to IVIM theory (6), due to intravoxel incoherent motion associated with capillary perfusion. "Perfusion fraction" (f_p in eq. 4) was proposed to reflex

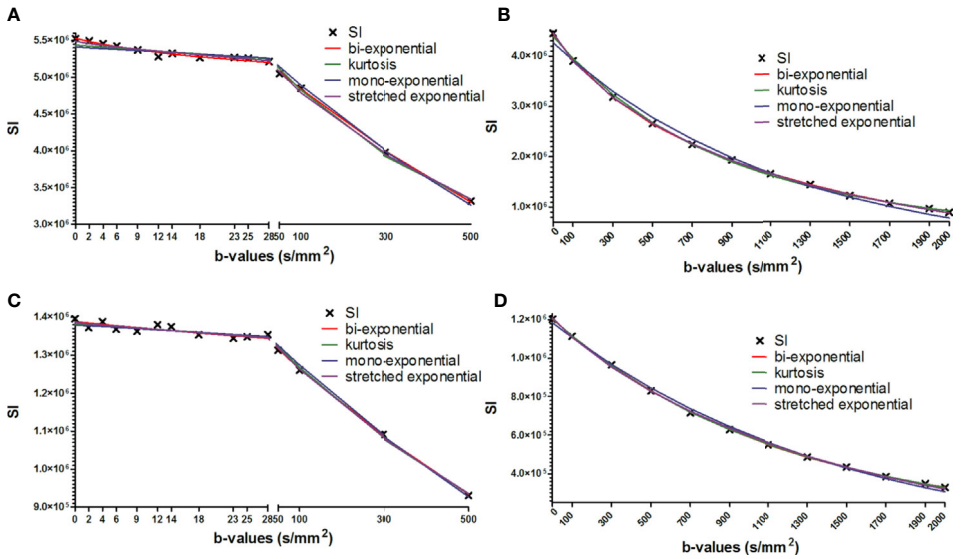


FIGURE 3 | Mean signal intensity as a function of b-values (x-axis) fitted using all four models. The data is shown for the same representative tumor as shown in Figures 1 and 2. **(A)**; week 1, low b-value DWI data. **(B)**; week 1, high b-value DWI data. **(C)**; week 4, low b-value DWI data. **(D)**; week 4 high b-value DWI data. Bi-exponential, kurtosis and stretched exponential models provide better fit to the DWI decay curve than the mono-exponential model especially at high b-value DWI data.

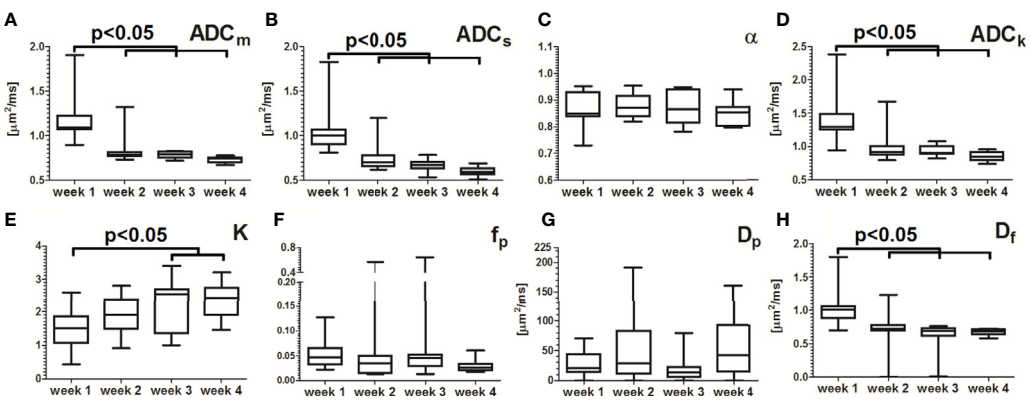


FIGURE 4 | Median values of regions of interest (n=11) derived from DWI data obtained using low b-values. Significant changes ($p<0.05$) were present in ADC_m (part **A**), ADC_s (part **B**), ADC_k (part **D**), and D_f (part **H**) values between week 1 and those from the weeks 2, 3 and 4. K (part **E**) parameter differed significantly ($p<0.05$) between week 1 and weeks 3 and 4. The differences between weeks 2, 3 and 4 were not significant. The remaining differences in the fitted values (alpha, part **C**; Fp, part **F**; Dp part **G**) between weeks did not reach the level of statistical significance. The box extends from the 25th to 75th percentiles while the error bars extend from minimal to maximal values.

fractional volume (%) of capillary blood flow while "pseudodiffusion" (D_p in eq. 4) probably relates to blood velocity. Biological reasons for non-Gaussian DWI signal decay at high b-values remains open despite several proposed theories (7, 27, 45, 46). In the current study, the highest used b-value was 2000 s/mm^2, which could have resulted in a less accurate estimation of slow diffusion component. Nevertheless, in vast majority of voxels non-Gaussian models provided better fit to DWI obtained using b-values up to 2000 s/mm^2. As shown in prior studies (47, 48) of DWI models functions for PCa, that b-value distribution and DWI acquisition parameters may contribute to the fitting performance, and exploration of these factors is left for future studies.

Our study is limited by relatively small sample size. Furthermore, no attempts to histologically validate our findings with cell density have been made. Thus, further studies are needed to better investigate correlation of parameters derived from non-Gaussian DWI models with histopathological markers. Signal intensities differed systematically in the second DWI examination of the same tumor performed approximately 60

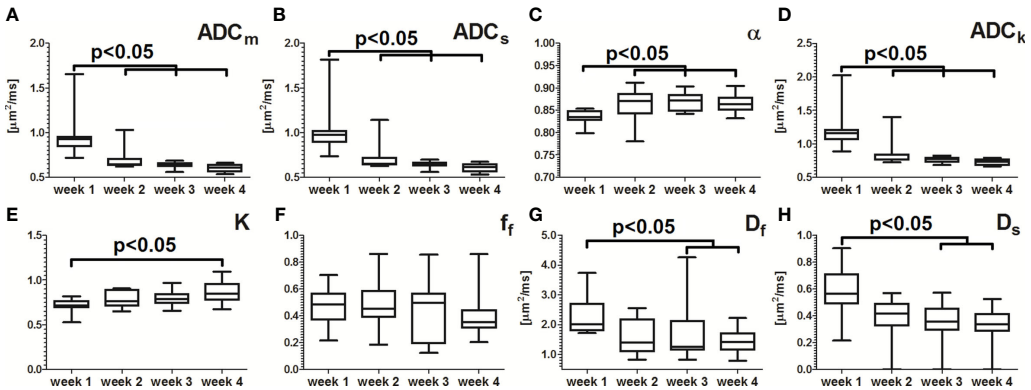


FIGURE 5 | Median values of regions of interest (n=11) derived from DWI data obtained using high b-values. Significant changes ($p<0.05$) were present in ADC_m (part **A**), ADC_s (part **B**), alpha (part **C**), and ADC_k (part **D**) values between week 1 and weeks 2, 3 and 4. K (part **E**) parameter differed significantly ($p<0.05$) between week 1 and week 4. The differences between weeks 2, 3 and 4 were not significant. Values of D_f (part **G**) and D_s (part **H**) parameters differed significantly between week 1 and weeks 3 and 4. The remaining differences in the fitted values between weeks did not reach the level of statistical significance (Ff, part **F**). The box extends from the 25th to 75th percentiles while the error bars extend from minimal to maximal values.

minutes after the first examination. This systematic bias in repeated DWI examinations has an effect on the estimation of repeatability. The confidence interval and coefficient of repeatability values between the repeated DWI examinations performed on week 4 are likely worse than those between weeks 1–4. Thus, the presented repeatability values should be regarded as the worst estimates due to systematic bias caused by DWI signal differences between the repeated DWI examinations. A potential

TABLE 1 | Mean ± standard deviation of median percentage values per mouse described better by the first model of the comparison is shown in the table for DWI data obtained using low b-values.

Low b-values		week 1*	week 2	week 3	week 4
AIC_c	stretched *vs.* mono-ex	81 ± 29%	69 ± 32%	71 ± 30%	78 ± 28%
	kurtosis *vs.* mono-ex	79 ± 30%	59 ± 38%	62 ± 36%	51 ± 33%
	bi-ex *vs.* mono-ex	75 ± 33%	56 ± 41%	53 ± 38%	69 ± 31%
	kurtosis *vs.* stretched	12 ± 16%	15 ± 28%	19 ± 24%	9 ± 15%
	bi-ex *vs.* stretched	23 ± 31%	15 ± 18%	10 ± 23%	29 ± 35%
	bi-ex *vs.* kurtosis	61 ± 33%	49 ± 37%	41 ± 38%	63 ± 35%
F-ratio	stretched *vs.* mono-ex	70 ± 34%	45 ± 44%	38 ± 39%	36 ± 30%
	kurtosis *vs.* mono-ex	45 ± 35%	28 ± 40%	19 ± 26%	8 ± 19%
	bi-ex *vs.* mono-ex	64 ± 34%	39 ± 40%	35 ± 35%	16 ± 16%
	biex *vs.* stretched	6 ± 12%	2 ± 5%	0 ± 0%	12 ± 30%
	bi-ex *vs.* kurtosis	42 ± 32%	23 ± 28%	13 ± 17%	26 ± 35%

*AIC_c, Corrected Akaike information criteria; mono-ex, mono-exponential; bi-ex, bi-exponential; *tumor growth after subcutaneous inoculation of PC-3 prostate cancer cells in mice.*

TABLE 2 | Mean ± standard deviation of median percentage values per mouse described better by the first model of the comparison is shown in the table for DWI data obtained using high b-values.

High b-values		week 1*	week 2	week 3	week 4
AIC_c	stretched *vs.* mono-ex	99 ± 4%	100 ± 1%	100 ± 1%	99 ± 2%
	kurtosis *vs.* mono-ex	99 ± 4%	100 ± 1%	100 ± 1%	100 ± 1%
	bi-ex *vs.* mono-ex	98 ± 8%	100 ± 2%	100 ± 1%	99 ± 2%
	kurtosis *vs.* stretched	69 ± 36%	81 ± 31%	71 ± 39%	78 ± 24%
	bi-ex *vs.* stretched	66 ± 38%	76 ± 27%	59 ± 39%	62 ± 31%
	bi-ex *vs.* kurtosis	36 ± 27%	33 ± 34%	18 ± 23%	16 ± 19%
F-ratio	stretched *vs.* mono-ex	96 ± 14%	95 ± 16%	94 ± 19%	94 ± 12%
	kurtosis *vs.* mono-ex	97 ± 10%	99 ± 3%	99 ± 2%	97 ± 4%
	bi-ex *vs.* mono-ex	100 ± 0%	100 ± 0%	100 ± 0%	100 ± 0%
	bi-ex *vs.* stretched	82 ± 32%	95 ± 8%	85 ± 27%	89 ± 18%
	bi-ex *vs.* kurtosis	92 ± 10%	92 ± 9%	93 ± 9%	94 ± 6%

*AIC_c, Corrected Akaike information criteria; mono-ex, mono-exponential; bi-ex, bi-exponential: *tumor growth after subcutaneous inoculation of PC-3 prostate cancer cells in mice.*

TABLE 3 | Repeatability of the fitted parameters for DWI data obtained using low and high b-values.

	Low b-values			High b-values		
	avg median	95% CI (%)	r% (%)	avg median	95% CI (%)	r% (%)
ADC_m	0.709^10^{-3} mm^2/s	7.2	19.9	0.581^10^{-3} mm^2/s	10.8	28.6
ADC_s	0.586^10^{-3} mm^2/s	11.7	28.8	0.583^10^{-3} mm^2/s	11.8	31.4
α	0.86	4.9	11.9	0.87	3.9	10.4
ADC_k	0.837^10^{-3} mm^2/s	9.2	22.6	0.704^10^{-3} mm^2/s	10.4	27.6
K	2.48	24.4	59.7	0.91	16.1	42.7
f_p	0.09	200	491.8	NA	NA	NA
f_f	NA	NA	NA	0.47	74.6	197.4
D_p	74.819^10^{-3} mm^2/s	162.9	399.1	NA	NA	NA
D_f	0.544^10^{-3} mm^2/s	58.9	163.2	1.564^10^{-3} mm^2/s	62.2	164.5
D_s	NA	NA	NA	0.271^10^{-3} mm^2/s	96	254.8

avg median, averaged median values per mouse; 95% CI (%), 95% confidence interval expressed as percentage of the averaged median values per mouse (avg median); r% (%), coefficient of repeatability as percentage of the averaged median value per mouse (avg median); NA, not applicable.

cause of the bias is a temperature drop of the tumor (not mouse core temperature) in the second repeated DWI examination. Nevertheless, it is beyond the scope of the current study to fully explore an effect of temperature and anesthesia (26) on DWI decay curve derived parameters.

In conclusion, we have evaluated four different mathematical models for DWI of PC-3 prostate cancer cell derived xenografts in mice. Significant changes in the fitted parameters were present during tumor progression potentially due to increased cell density in later stages. The "pseudodiffusion" component in the analyzed tumors was shown to be less than 10% of the bi-exponential model. Due to low repeatability of the bi-exponential model parameters derived from low and high b-values DWI data using independent least square fitting on a voxel level, a degree of caution should be applied if these parameters are used for cancer characterization and therapy response monitoring. On the other hand, mono-exponential, stretched exponential, and kurtosis models shown high information content and robust repeatability.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by Turku Center for Disease Modeling, University of Turku, Turku, Finland.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication. Specifically, Conceptualization: HaM, AR, TL, IJ. Data curation: All authors. Funding acquisition: HA, TL, IJ. Investigation: HaL, HeL, HV, MP, TL. Methodology: All authors. Project administration: AR, TL, IJ. Resources: AR, MP, TL, IJ. Software and statistical analyses: HaM, IJ. Supervision: TL, IJ. Validation: HaM, IJ, Visualization: HaM, IJ. Writing – original draft: HaM, IJ. Writing – review & editing: HaM, AR, TL, IJ.

FUNDING

This study was financially supported by grants Sigrid Jusélius Foundation (HaM, HA, and IJ), Finnish Cultural Foundation (HM and IJ), Finnish Cancer Society (IJ), Orion Research Foundation (HaM), Instrumentarium Research Foundation (IJ), Academy of Finland (AR and TL), Turku University Hospital (HA and IJ), TYKS-SAPA research fund (HA and IJ), Turku University Foundation (HaM and IJ), and University of Eastern Finland strategic funding (TL and HL).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.583921/full#supplementary-material

REFERENCES

1. Moffat BA, Hall DE, Stojanovska J, McConville PJ, Moody JB, Chenevert TL, et al. Diffusion Imaging for Evaluation of Tumor Therapies in Preclinical Animal Models. *MAGMA* (2004) 17:249–59. doi: 10.1007/s10334-004-0079-z
2. Liimatainen T, Hakumaki JM, Kauppinen RA, Ala-Korpela M. Monitoring of Gliomas in Vivo by Diffusion MRI and (1)H MRS During Gene Therapy-Induced Apoptosis: Interrelationships Between Water Diffusion and Mobile Lipids. *NMR Biomed* (2009) 22:272–9. doi: 10.1002/nbm.1320
3. Padhani AR, Liu G, Koh DM, Chenevert TL, Thoeny HC, Takahara TT, et al. Diffusion-Weighted Magnetic Resonance Imaging as a Cancer Biomarker: Consensus and Recommendations. *Neoplasia* (2009) 11:102–25. doi: 10.1593/neo.81328
4. Padhani AR, Makris A, Gall P, Collins DJ, Tunariu N, de Bono JS. Therapy Monitoring of Skeletal Metastases With Whole-Body Diffusion MRI. *J Magn Reson Imaging* (2014) 39:1049–78. doi: 10.1002/jmri.24548
5. Bammer R. Basic Principles of Diffusion-Weighted Imaging. *Eur J Radiol* (2003) 45:169–84. doi: 10.1016/S0720-048X(02)00303-0

6. LeBihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of Diffusion and Perfusion in Intravoxel Incoherent Motion MR Imaging. *Radiology* (1988) 168:497–505. doi: 10.1148/radiology.168.2.3393671
7. LeBihan D. The 'Wet Mind': Water and Functional Neuroimaging. *Phys Med Biol* (2007) 52:R57–90. doi: 10.1088/0031-9155/52/7/R02
8. Lee HJ, Rha SY, Chung YE, Shim HS, Kim YJ, Hur J, et al. Tumor Perfusion-Rrelated Parameter of Diffusion-Weighted Magnetic Resonance Imaging: Correlation With Histological Microvessel Density. *Magn Reson Med* (2014) 71:1554–8. doi: 10.1002/mrm.24810
9. LeBihan D, Turner R. The Capillary Network: A Link Between IVIM and Classical Perfusion. *Magn Reson Med* (1992) 27:171–8. doi: 10.1002/mrm.1910270116
10. LeBihan D. Intravoxel Incoherent Motion Perfusion MR Imaging: A Wake-Up Call. *Radiology* (2008) 249:748–52. doi: 10.1148/radiol.2493081301
11. Niendorf T, Dijkhuizen RM, Norris DG, van Lookeren CM, Nicolay K. Biexponential Diffusion Attenuation in Various States of Brain Tissue: Implications for Diffusion-Weighted Imaging. *Magn Reson Med* (1996) 36:847–57. doi: 10.1002/mrm.1910360607
12. Bennett KM, Schmainda KM, Bennett RT, Rowe DB, Lu H, Hyde JS. Characterization of Continuously Distributed Cortical Water Diffusion Rates with a Stretched-Exponential Model. *Magn Reson Med* (2003) 50:727–34. doi: 10.1002/mrm.10581
13. Clark CA, Le BD. Water Diffusion Compartmentation and Anisotropy at High B Values in the Human Brain. *Magn Reson Med* (2000) 44:852–9. doi: 10.1002/1522-2594(200012)44:6<852::AID-MRM5>3.0.CO;2-A
14. Yablonskiy DA, Bretthorst GL, Ackerman JJ. Statistical Model for Diffusion Attenuated MR Signal. *Magn Reson Med* (2003) 50:664–9. doi: 10.1002/mrm.10578
15. Schuster DM, Votaw JR, Nieh PT, Weiping W, Jonathon JA, Viraj VM, et al. Initial Experience with the Radiotracer Anti-1-Amino-3-18F-Fluorocyclobutane-1-Carboxylic Acid with PET/CT in Prostate Carcinoma. *J Nucl Med* (2007) 48:56–63.
16. Bourne RM, Panagiotaki E, Bongers A, Sved P, Watson G, Alexander DC. Information Theoretic Ranking of Four Models of Diffusion Attenuation in Fresh and Fixed Prostate Tissue Ex Vivo. *Magn Reson Med* (2014) 72:1418–26. doi: 10.1002/mrm.25032
17. Wittsack HJ, Lanzman RS, Mathys C, Janssen H, Modder U, Blondin D. Statistical Evaluation of Diffusion-Weighted Imaging of the Human Kidney. *Magn Reson Med* (2010) 64:616–22. doi: 10.1002/mrm.22436
18. Boxenbaum HG, Riegelman S, Elashoff RM. Statistical Estimations in Pharmacokinetics. *J Pharmacokinet Biopharm* (1974) 2:123–48. doi: 10.1007/BF01061504
19. Ludden TM, Beal SL, Sheiner LB. Comparison of the Akaike Information Criterion, the Schwarz Criterion and the F Test as Guides to Model Selection. *J Pharmacokinet Biopharm* (1994) 22:431–45. doi: 10.1007/BF02353864
20. Kopf M, Corinth C, Haferkamp O, Nonnenmacher TF. Anomalous Diffusion of Water in Biological Tissues. *Biophys J* (1996) 70:2950–8. doi: 10.1016/S0006-3495(96)79865-X
21. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional Kurtosis Imaging: The Quantification of Non-Gaussian Water Diffusion by Means of Magnetic Resonance Imaging. *Magn Reson Med* (2005) 53:1432–40. doi: 10.1002/mrm.20508
22. Jensen RT, Battey JF, Spindel ER, Benya RV. International Union of Pharmacology. LXVIII. Mammalian Bombesin Receptors: Nomenclature, Distribution, Pharmacology, Signaling, and Functions in Normal and Disease States. *Pharmacol Rev* (2008) 60:1–42. doi: 10.1124/pr.107.07108
23. Cescato R, Maina T, Nock B, Nikolopoulou A, Charalambidis D, Piccand V, et al. Bombesin Receptor Antagonists may be Preferable to Agonists for Tumor targeting. *J Nucl Med* (2008) 49:318–26. doi: 10.2967/jnumed.107.045054
24. Koh DM, Blackledge M, Collins DJ, Padhani AR, Wallace T, Wilton B, et al. Reproducibility and Changes in the Apparent Diffusion Coefficients of Solid Tumours Treated With Combretastatin A4 Phosphate and Bevacizumab in a Two-Centre Phase I Clinical Trial. *Eur Radiol* (2009) 19:2728–38. doi: 10.1007/s00330-009-1469-4
25. Zelhof B, Pickles M, Liney G, Gibbs P, Rodrigues G, Kraus S, et al. Correlation of Diffusion-Weighted Magnetic Resonance Data with Cellularity in Prostate Cancer. *BJU Int* (2009) 103:883–8. doi: 10.1111/j.1464-410X.2008.08130.x
26. Valette J, Guillermier M, Besret L, Hantraye P, Bloch G, Lebon V. Isoflurane Strongly Affects The Diffusion of Intracellular Metabolites, as Shown by 1H Nuclear Magnetic Resonance Spectroscopy of the Monkey Brain. *J Cereb Blood Flow Metab* (2007) 27:588–96. doi: 10.1148/radiol.10091343
27. Mulkern RV, Gudbjartsson H, Westin CF, Zengingonul HP, Gartner W, Guttmann CR, et al. Multi-Component Apparent Diffusion Coefficients in Human Brain. *NMR Biomed* (1999) 12:51–62. doi: 10.1002/(SICI)1099-1492(199902)12:1<51::AID-NBM546>3.0.CO;2-E
28. Mardor Y, Pfeffer R, Spiegelmann R, Roth Y, Maier SE, Nissim O, et al. Early Detection of Response to Radiation Therapy in Patients with Brain Malignancies Using Conventional and High B-Value Diffusion-Weighted Magnetic Resonance imaging. *J Clin Oncol* (2003) 21:1094–100. doi: 10.1200/JCO.2003.05.069
29. Roth Y, Tichler T, Kostenich G, Ruiz-Cabello J, Maier SE, Cohen JS, et al. High-B-Value Diffusion-Weighted MR Imaging for Pretreatment Prediction and Early Monitoring of Tumor Response to Therapy in Mice. *Radiology* (2004) 232:685–92. doi: 10.1148/radiol.2322030778
30. Hoff BA, Chenevert TL, Bhojani MS, Kwee TC, Rehemtulla A, Le Bihan D, et al. Assessment of Multiexponential Diffusion Features as MRI Cancer Therapy Response Metrics. *Magn Reson Med* (2010) 64:1499–509. doi: 10.1002/mrm.22507
31. Zhang Y, Jin N, Deng J, Guo Y, White SB, Yang GY, et al. Intra-Voxel Incoherent Motion MRI in Rodent Model of Diethylnitrosamine-Induced Liver Fibrosis. *Magn Reson Imaging* (2013) 31:1017–21. doi: 10.1016/j.mri.2013.03.007
32. Kim S, Decarlo L, Cho GY, Jensen JH, Sodickson DK, Moy L, et al. Interstitial Fluid Pressure Correlates with Intravoxel Incoherent Motion Imaging Metrics in a Mouse Mammary Carcinoma Model. *NMR Biomed* (2012) 25:787–94. doi: 10.1002/nbm.1793
33. Chow AM, Gao DS, Fan SJ, Qiao Z, Lee FY, Yang J, et al. Liver Fibrosis: An Intravoxel Incoherent Motion (IVIM) Study. *J Magn Reson Imaging* (2012) 36:159–67. doi: 10.1002/jmri.23607
34. Penner AH, Sprinkart AM, Kukuk GM, Gütgemann I, Gieseke J, Schild HH, et al. Intravoxel Incoherent Motion Model-Based Liver Lesion Characterisation From Three B-Value Diffusion-Weighted MRI. *Eur Radiol* (2013) 23:2773–83. doi: 10.1007/s00330-013-2869-z
35. Jerome NP, Orton MR, d'Arcy JA, Collins DJ, Koh DM, Leach MO. Comparison of Free-Breathing with Navigator-Controlled Acquisition Regimes in Abdominal Diffusion-Weighted Magnetic Resonance Images: Effect on ADC and IVIM Statistics. *J Magn Reson Imaging* (2014) 39:235–40. doi: 10.1002/jmri.24140
36. Dyvorne HA, Galea N, Nevers T, Fiel MI, Carpenter D, Wong E, et al. Diffusion-Weighted Imaging of the Liver with Multiple B Values: Effect of Diffusion Gradient Polarity and Breathing Acquisition on Image Quality and Intravoxel Incoherent Motion Parameters–A Pilot Study. *Radiology* (2013) 266:920–9. doi: 10.1148/radiol.12120686
37. Pekar J, Moonen CT, van Zijl PC. On the Precision of Diffusion/Perfusion Imaging by Gradient Sensitization. *Magn Reson Med* (1992) 23:122–9. doi: 10.1002/mrm.1910230113
38. Andreou A, Koh DM, Collins DJ, Blackledge M, Wallace T, Leach MO, et al. Measurement Reproducibility of Perfusion Fraction and Pseudodiffusion Coefficient Derived by Intravoxel Incoherent Motion Diffusion-Weighted MR Imaging in Normal Liver and Metastases. *Eur Radiol* (2013) 23:428–34. doi: 10.1007/s00330-012-2604-1
39. Federau C, Maeder P, O'Brien K, Browaeys P, Meuli R, Hagmann P. Quantitative Measurement of Brain Perfusion with Intravoxel Incoherent Motion MR Imaging. *Radiology* (2012) 265:874–81. doi: 10.1148/radiol.12120584
40. Federau C, Meuli R, O'Brien K, Maeder P, Hagmann P. Perfusion Measurement in Brain Gliomas with Intravoxel Incoherent Motion MRI. *AJNR Am J Neuroradiol* (2014) 35:256–62. doi: 10.3174/ajnr.A3686
41. Bisdas S, Koh TS, Roder C, Braun C, Schittenhelm J, Ernemann U, et al. Intravoxel Incoherent Motion Diffusion-Weighted MR Imaging of Gliomas: Feasibility of the Method and Initial Results. *Neuroradiology* (2013) 55:1189–96. doi: 10.1007/s00234-013-1229-7
42. Luciani A, Vignaud A, Cavet M, Van Nhieu JT, Mallat A, Ruel L, et al. Liver Cirrhosis: Intravoxel Incoherent Motion MR Imaging–Pilot Study. *Radiology* (2008) 249:891–9. doi: 10.1148/radiol.2493080080
43. Orton MR, Collins DJ, Koh DM, Leach MO. Improved intravoxel incoherent Motion Analysis of Diffusion Weighted Imaging by Data Driven Bayesian Modeling. *Magn Reson Med* (2014) 71:411–20. doi: 10.1002/mrm.24649

44. Freiman M, Perez-Rossello JM, Callahan MJ, Voss SD, Ecklund K, Mulkern R, et al. Reliable Estimation of Incoherent Motion Parametric Maps from Diffusion-Weighted MRI Using Fusion Bootstrap Moves. *Med Image Anal* (2013) 17:325–36. doi: 10.1016/j.media.2012.12.001
45. Schwarcz A, Bogner P, Meric P, Correze JL, Berente Z, Pál J, et al. The Existence of Biexponential Signal Decay in Magnetic Resonance Diffusion-Weighted Imaging Appears to be Independent of Compartmentalization. *Magn Reson Med* (2004) 51:278–85. doi: 10.1002/mrm.10702
46. Ababneh Z, Beloeil H, Berde CB, Gambarota G, Maier SE, Mulkern RV. Biexponential Parameterization of Diffusion and T2 Relaxation Decay Curves in a Rat Muscle Edema Model: Decay Curve Components and Water Compartments. *Magn Reson Med* (2005) 54:524–31. doi: 10.1002/mrm.20610
47. Meisaari H, Toivonen J, Pesola M, Taimen P, Boström PJ, Pahikkala T, et al. Diffusion-Weighted Imaging of Prostate Cancer: Effect of B-Value Distribution on Repeatability and Cancer Characterization. *Magn Reson Imaging* (2015) 33:1212–8. doi: 10.1016/j.mri.2015.07.004
48. Jambor I, Merisaari H, Taimen P, Boström P, Aronen H, Pesola M, et al. Evaluation of Different Mathematical Models for Diffusion-Weighted Imaging of Normal Prostate and Prostate Cancer Using High B-Values: A Repeatability Study. *Magn Reson Med* (2005) 54:524–31.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neurorobotics

ORIGINAL RESEARCH
published: 14 June 2021
doi: 10.3389/fnbot.2021.563682

Reaction–Diffusion Model-Based Research on Formation Mechanism of Neuron Dendritic Spine Patterns

Yiqing Jia[1†], *Qili Zhao*[1†], *Hongqiang Yin*[2†], *Shan Guo*[1], *Mingzhu Sun*[1], *Zhuo Yang*[2*] and *Xin Zhao*[1*]

[1] Institute of Robotics and Automatic Information Systems, College of Artificial Intelligence, Nankai University, Tianjin, China,
[2] State Key Laboratory of Medicinal Chemical Biology, School of Medicine, Nankai University, Tianjin, China

The pattern abnormalities of dendritic spine, tiny protrusions on neuron dendrites, have been found related to multiple nervous system diseases, such as Parkinson's disease and schizophrenia. The determination of the factors affecting spine patterns is of vital importance to explore the pathogenesis of these diseases, and further, search the treatment method for them. Although the study of dendritic spines is a hot topic in neuroscience in recent years, there is still a lack of systematic study on the formation mechanism of its pattern. This paper provided a reinterpretation of reaction-diffusion model to simulate the formation process of dendritic spine, and further, study the factors affecting spine patterns. First, all four classic shapes of spines, mushroom-type, stubby-type, thin-type, and branched-type were reproduced using the model. We found that the consumption rate of substrates by the cytoskeleton is a key factor to regulate spine shape. Moreover, we found that the density of spines can be regulated by the amount of an exogenous activator and inhibitor, which is in accordance with the anatomical results found in hippocampal CA1 in SD rats with glioma. Further, we analyzed the inner mechanism of the above model parameters regulating the dendritic spine pattern through Turing instability analysis and drew a conclusion that an exogenous inhibitor and activator changes Turing wavelength through which to regulate spine densities. Finally, we discussed the deep regulation mechanisms of several reported regulators of dendritic spine shape and densities based on our simulation results. Our work might evoke attention to the mathematic model-based pathogenesis research for neuron diseases which are related to the dendritic spine pattern abnormalities and spark inspiration in the treatment research for these diseases.

Keywords: dendritic spine, Turing instability, reaction-diffusion model, branching morphogenesis, glioma

OPEN ACCESS

Edited by:
Francesco Rundo, STMicroelectronics, Italy

Reviewed by:
Takeshi Kano, Tohoku University, Japan
Shigeo Okabe, The University of Tokyo, Japan

***Correspondence:**
Xin Zhao
zhaoxin@nankai.edu.cn
Zhuo Yang
zhuoyang@nankai.edu.cn

[†]These authors have contributed equally to this work

Received: *19 May 2020*
Accepted: *17 May 2021*
Published: *14 June 2021*

Citation:
Jia Y, Zhao Q, Yin H, Guo S, Sun M, Yang Z and Zhao X (2021) Reaction–Diffusion Model-Based Research on Formation Mechanism of Neuron Dendritic Spine Patterns. Front. Neurorobot. 15:563682. doi: 10.3389/fnbot.2021.563682

INTRODUCTION

Dendritic spines are tiny protrusions on neuron dendrites which widely exist in the dendrites of higher animals and play an important role in the formation of most excitatory axodendritic synapses (Harris and Kater, 1994). The function of a spine is related to its shape (Kasai et al., 2003; Bourne and Harris, 2007). Traditionally, there are four basic shapes for dendritic spines: thin-type, stubby-type, mushroom-type, and branched-type (González-Tapia et al., 2016; Luczynski et al., 2016). Among them, thin dendritic spines show high plasticity and are related to learning,

while mushroom dendritic spines show weak plasticity and are related to memory function. In addition, the density of spines directly influences the density of synapses. Researchers have found that pattern abnormalities of dendritic spine, especially the abnormal proportion of various types of dendritic spines and density variation of dendritic spines, were related to multiple nervous system diseases. For example, Pyronneau et al. reported an overabundance of thin-type spines, a kind of immature dendritic spines, in the somatosensory cortex of Fragile X syndrome model mice (Pyronneau et al., 2017). It has been reported that there are striatal dendrites with few dendritic spines in Parkinson's disease (McNeill et al., 1988). It was also been found that reduced dendritic spine density in individuals with schizophrenia (Glantz and Lewis, 2000; Sweet et al., 2008) and Huntington's disease (Richards et al., 2011). Also, it is recognized that dendritic spine loss is an early feature of Alzheimer's disease (Kommaddi et al., 2018; O'Neal et al., 2018). Thus, the exploration of shape and density factors of dendritic spines is of vital importance to understand the pathogenesis of these diseases, and further, search the treatment method for them.

The current research on dendritic spines pattern is mainly performed by statically observing the cerebral cortex in animals (Kommaddi et al., 2018; Ratliff et al., 2019). It has been confirmed that the pattern of dendritic spines is influenced by neuron activity (Portera-Cailliau et al., 2003; González-Tapia et al., 2016) and some substances, such as drebrin (Hayashi et al., 1996), Rho GTPase Rac1 (Pyronneau et al., 2017) and F-actin (Kommaddi et al., 2018). The above researches usually only proposed one factor of dendritic spine patterns once while the pattern formation of dendritic spines is a dynamic process involving a variety of chemical reactions that are regulated by multiple factors. In summary, there is still a lack of systematic study on the mechanism of pattern formation showing influences of multiple factors on the formed pattern of dendritic spines.

Mathematic modeling on dendritic spines development has become an important tool to study the structure and plasticity of dendritic spines in recent years. For example, Kasai et al. used the volume of dendritic spines as an index to measure the structure of dendritic spines and applied the Brownian motion model to simulate the volume of dendritic spines, exploring the close relationship between spine structure and function (Kasai et al., 2010). The Brownian motion model describes a random phenomenon, but the pattern formation of dendritic spines is a process regulated by gene and environment instead of a random process, making that model unsuitable for simulating the pattern formation. Besides, Miermans et al. simulated dendritic spine membranes during shape alternation using the Canham-Helfrich energy functional, which is used to describe the relationship between the bending rigidity of the membrane and the force generated by the cytoskeleton (Miermans et al., 2017). Their results demonstrate that the cytoskeleton is a key factor in determining the shape of dendritic spines, but this model lacks an explanation for the change in cytoskeletons, and their hypothesis about the approximate rotational symmetry of dendritic spines seems inapplicable to branched-type dendritic spines. Varner et al. explained the process of epithelial cell formation patterns using four mechanisms: cell division, cell insertion, cell deformation, and media filling (Varner and Nelson, 2014). However, these explanations cannot be applied in the study of sub-cellular structures such as dendritic spines.

In Turing theory, if the chemical substances involved in the interaction have diffusion, the original equilibrium state will be broken, which is called Turing instability (Turing, 1952). The reaction-diffusion model (Gierer and Meinhardt, 1972; Meinhardt, 1976), based on Turing's theory, illuminates the reactions between chemical substances in developing biological systems. It has been utilized to simulate Pomacanthus skin stripe patterns (Kondo and Asai, 1995), vascular mesenchymal cells patterns (Garfinkel et al., 2004), mouse limb development (Miura et al., 2006), lung branching patterns (Guo et al., 2014a; Hagiwara et al., 2015), and self-organizing morphogenesis (Okuda et al., 2018; Landge et al., 2020). In our previous work, side branching and tip branching of the lung were investigated using the reaction-diffusion model, which was verified by spatiotemporal parameters (Guo et al., 2014a). However, the patterns developed in previous work were not enough to describe the complex patterns in dendritic spines. Because different from the obtained side branches which were equally spaced, the dendritic spines studied in this paper are usually uneven. In spite of its potential use in simulate branching patterns, the strong non-linearity of the reaction-diffusion model makes it difficult to intuitively draw the relationship between parameter values and simulation results, which is inconvenient for the analysis of the inner mechanism of the model. Addressing this problem, dispersion relation was used to analyze Turing instability (Guo et al., 2014b; Saleem and Ali, 2018) to prove the mathematical mechanism of the simulation results. In previous research, we have investigated the mathematic mechanism through Turing instability analysis and found that different Turing wavelengths are underlying the different patterns in a lung (Xu et al., 2017). However, the relationship between Turing wavelength and branch density has not been investigated yet.

This paper reinterpreted the traditional reaction-diffusion model through the introduction of exogenous activator term and exogenous inhibitor term to simulate the formation process of dendritic spine, and further, study the factors affecting spine patterns. All four spine shapes, mushroom-type, stubby-type, thin-type, and branched-type, were reproduced using the model. Further, we found that the consumption rate of substrates by the cytoskeleton regulates the shape. Secondly, we found that the addition of an exogenous activator causes the spines to become denser, while the addition of an exogenous inhibitor causes the spines to become sparser, which provided a potential explanation for the anatomical results that spine decrease in hippocampal CA1 in SD rats with glioma. Finally, through Turing instability analysis, we found that Turing wavelength variation is the deep mathematical mechanism behind above parameters regulating spine density. Namely, the addition of an exogenous activator decreases the Turing wavelength, causing the density of the dendritic spines to increase, while the addition of an exogenous inhibitor increases the Turing wavelength, causing the density of the dendritic spines to decrease. Finally, the deep regulation mechanisms of several regulators of dendritic spine shape and density reported in other references were discussed based on our

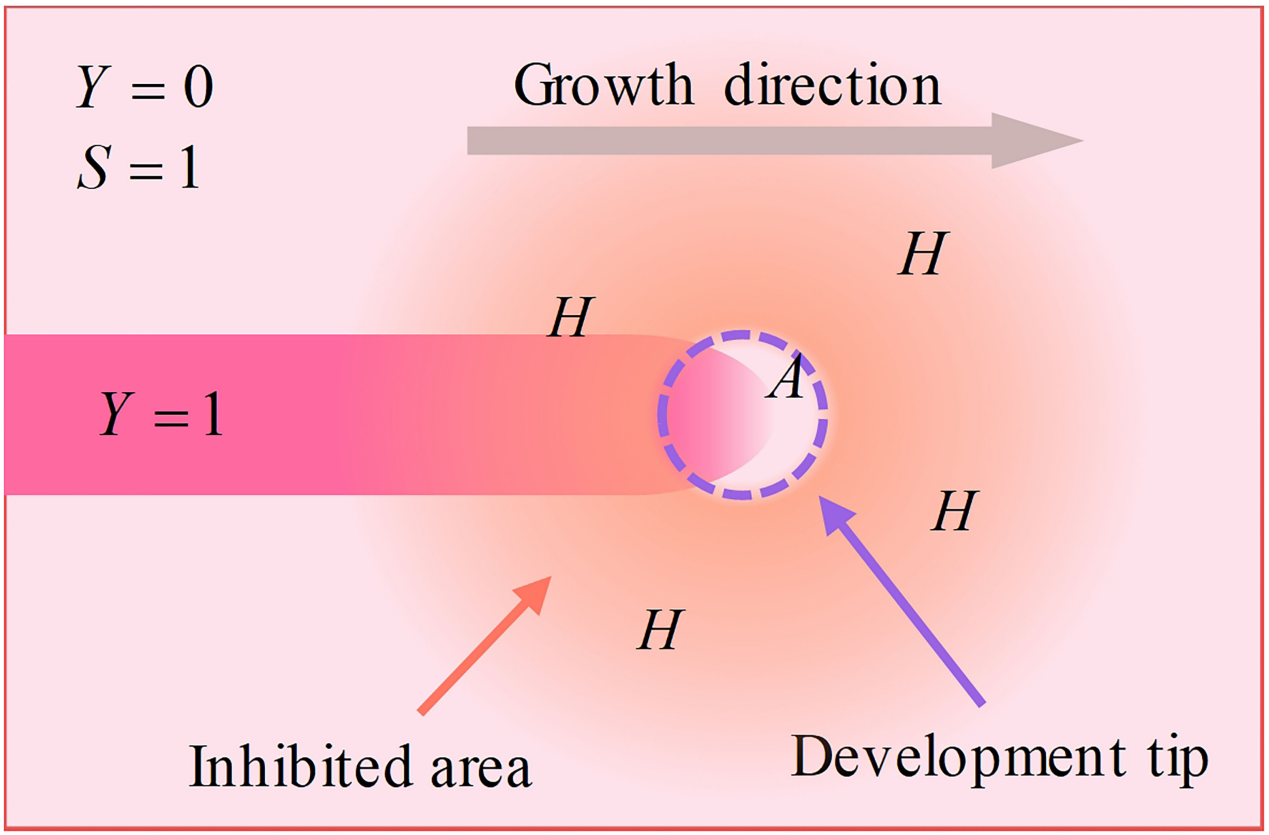


FIGURE 1 | Schematic of the development process. The neuron expresses activators and inhibitors. Activators gather at the tip, while inhibitors diffuse into the surrounding area due to a higher diffusion rate, making only the tip develops. This mechanism makes dendritic spines grow in a certain direction instead of exhibiting isotropous growth.

simulation results. We hope that our work could evoke attention to the mathematic model-based research for neuron diseases related to the dendritic spine pattern abnormalities and spark inspiration in the treatment research for these diseases.

MATERIALS AND METHODS

Reaction-Diffusion Model

The reaction-diffusion model is defined by Equation (1) (Meinhardt, 1976). It is a group of partial differential equations describing the reactions between activator A, inhibitor H, substrate S, and cytoskeleton Y.

$$\begin{cases} \frac{\partial A}{\partial t} = \frac{cA^2S}{H} - \mu A + \rho_A Y + D_A \nabla^2 A \\ \frac{\partial H}{\partial t} = cA^2S - \upsilon H + \rho_H Y + D_H \nabla^2 H \\ \frac{\partial S}{\partial t} = c_0 - \gamma S - \varepsilon YS + D_S \nabla^2 S \\ \frac{\partial Y}{\partial t} = dA - eY + \frac{Y^2}{1+fY^2} \end{cases} . \quad (1)$$

The reaction-diffusion model illuminates the reactions between chemical substances in developing biological systems. According to this model, neurons express activators (at a rate ρ_A) and inhibitors (at a rate ρ_H). Activators behave with self-catalysis (at a rate c) and catalyze inhibitors (at a rate c), while inhibitors inhibit activators. Simultaneously, activators and inhibitors behave with degradation and diffusion (activators degrade at a rate μ and diffuse at a rate D_A, whereas inhibitors degrade at a rate υ and diffuse at a rate D_H). High concentrations of activator accelerate the polymerization of cytoskeletons, inducing the development of dendritic spines. Because the diffusion rate of inhibitors is higher than that of activator, the polymerization of the cytoskeleton in the growth center is accelerated, and the polymerization of the cytoskeleton outside the growing center is inhibited. Thus, the dendritic spine grows in a certain direction, instead of displaying isotropous growth. The neuron creates substrate (at a rate c_0), while the cytoskeleton consumes substrate (at a rate ε). Substrate accelerates the catalysis of the activator. Similarly, the substrate behaves via degradation and diffusion (degrades at a rate γ and diffuses at a rate D_S), as well. Because the synthesis of the cytoskeleton consumes substrate, the peak concentration areas of activators and inhibitors, as well as the cytoskeleton, move in the direction of high substrate concentrations (**Figure 1**).

The development patterns of dendritic spines are determined by the neuron activity (Bloodgood and Sabatini, 2005) and the exogenous substances. The neuron activity is described by

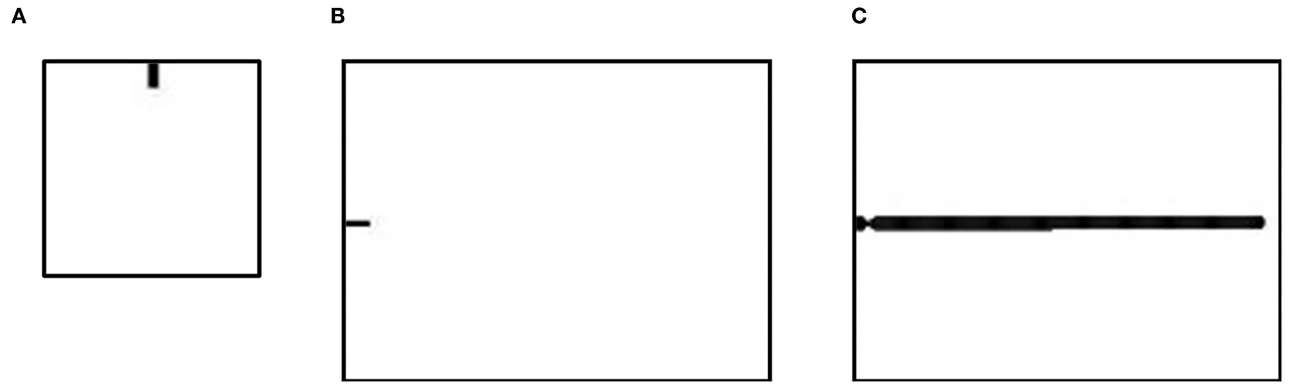


FIGURE 2 | The original state of the spine simulation and the dendrite simulation. **(A)** The original state of the spine simulation is used to simulate a single spine in different conditions. Simulations were performed on a 100×100 grid. The grid size of space is 0.3. Fixed parameters in Equation (2): $c = 0.002$, $\mu = 0.16$, $\upsilon = 0.04$, $\rho_A = 0.01$, $\rho_H = 0.00005$, $c_0 = 0.02$, $\gamma = 0.02$, $D_A = 0.02$, $D_H = 0.26$, $D_S = 0.06$, $d = 0.0035$, $e = 0.1$, and $f = 10$. **(B)** The first step in the dendrite simulation is used to simulate the dendrite trunk. Simulations were performed on a 150×200 grid. The grid size of the space is 0.3. Parameters in Equation (2): $c = 0.002$, $\mu = 0.16$, $\upsilon = 0.04$, $\rho_A = 0.03$, $\rho_H = 0.0001$, $\delta_A = 0$, $\delta_H = 0$, $c_0 = 0.02$, $\varepsilon = 0.017$, $\gamma = 0.02$, $D_A = 0.02$, $D_H = 0.26$, $D_S = 0.06$, $d = 0.0035$, $e = 0.1$, and $f = 10$. **(C)** The second step in the dendrite simulation grows from **(A)** and is used to simulate spines in different conditions. Fixed parameters in Equation (2): $c = 0.002$, $\mu = 0.16$, $\upsilon = 0.04$, $\rho_A = 0.02$, $\rho_H = 0.00005$, $c_0 = 0.05$, $\gamma = 0.02$, $D_A = 0.02$, $D_H = 0.26$, $D_S = 0.06$, $d = 0.0035$, $e = 0.1$, and $f = 10$. In **(A,B)**, black regions ($A = 2$, $H = 0.02$, $S = 1$, $Y = 1$) represent a part of a neuron, and white regions ($A = 0.001$, $H = 0.001$, $S = 1$, $Y = 0$) represent the environment surrounding the neuron.

the rate of substrate consumed by the cytoskeleton (ε) in our model. Exogenous substances include exogenous activators and exogenous inhibitors (Kommaddi et al., 2018). To describe the influence of exogenous substances, we added the exogenous activator term (δ_A) and the exogenous inhibitor term (δ_H) into the reaction-diffusion model:

$$\begin{cases} \frac{\partial A}{\partial t} = \frac{cA^2S}{H} - \mu A + (\rho_A + \delta_A)\, Y + D_A \nabla^2 A \\ \frac{\partial H}{\partial t} = cA^2S - \upsilon H + (\rho_H + \delta_H)\, Y + D_H \nabla^2 H \\ \frac{\partial S}{\partial t} = c_0 - \gamma S - \varepsilon YS + D_S \nabla^2 S \\ \frac{\partial Y}{\partial t} = dA - eY + \frac{Y^2}{1+fY^2} \end{cases} . \quad (2)$$

The new model includes 16 parameters, most of which are fixed parameters, such as reaction-term parameter c, degradation-term parameters μ, υ, and γ, diffusion-term parameters D_A, D_H, and D_S, and growth-term parameters d, e, and f. The values of fixed parameters are decided by the chemical characteristics of substances or cells, and the model has been proven to be robust to perturbations of fixed parameters (Murray, 1982). The other parameters are variable (ρ_A, δ_A, ρ_H, δ_H, c_0, and ε), whose values depend on the condition of the development system. In this work, we studied the effect of the neuron activity and the exogenous substances on dendritic spines. Thus, we set parameters δ_A, δ_H, and ε in Equation (2) as variable parameters.

The values of parameters were set according to previous research in lung branching patterns. In previous work, we set the values of parameters as: $c = 0.002$, $\mu = 0.18$, $\upsilon = 0.04$, $\rho_A = 0.063$, $\rho_H = 0.00005$, $c_0 = 0.02$, $\gamma = 0.02$, $\varepsilon = 0.045$, $D_A = 0.02$, $D_H = 0.32$, $D_S = 0.06$, $d = 0.0033$, $e = 0.1$, and $f = 10$. We verified the consistency of the mathematical model under certain parameters with the actual biological process by converting the time and space in the numerical simulation and comparing them with the spatiotemporal scale of real lung development (Guo et al., 2014a). The values of fixed parameters and the value ranges of variable parameters in the lung branching model provide references in our new model.

Numerical Simulation

In this work, we investigated the factors of shape and density of spines using a reaction-diffusion model on different spatial scales. First, we simulated a spine to explore the influence of model parameters on the shape of the spine (**Figure 2A**). This simulation was performed on a 100×100 grid, and the original state was a 10×5 pixels rectangular area. Second, we simulated a dendrite with spines to explore the influence of model parameters on the density of spines (**Figures 2B,C**). This simulation was performed on a 150×200 grid, and the original state was a 5×10 pixels rectangular area (**Figure 2B**). Then, a dendrite developed under certain conditions (**Figure 2C**).

Turing Instability Analysis Method

To verify the simulation results with mathematics, we explored Turing patterns underlying dendritic spine patterns with our previously developed decoupling method (Guo et al., 2014b). The substrate and cytoskeleton are considered dependent variables of time and space, written as S(x, y, t) and Y(x, y, t). Then, we put these variables into Equation (2) as parameters and obtained the model of an activator-inhibitor system as:

$$\begin{cases} \frac{\partial A}{\partial t} = \frac{cA^2S(x,y,t)}{H} - \mu A + (\rho_A + \delta_A)\, Y\left(x,y,t\right) + D_A \nabla^2 A \\ \frac{\partial H}{\partial t} = cA^2S\left(x,y,t\right) - \upsilon H + (\rho_H + \delta_H)\, Y\left(x,y,t\right) + D_H \nabla^2 H \end{cases} . \quad (3)$$

Branching is a system that can grow and form stable mode, which corresponds to the damped oscillation system of mathematical model. Some points in S-Y space correspond to damped oscillatory systems. The set of these points is called Turing instability space, and the wavelength of damped oscillation

system is called Turing wavelength (Turing, 1952). According to its definition, the mathematical expression of Turing space can be calculated. The detailed derivation process is in our previous work (Xu et al., 2017).

To explore dendritic spine development patterns according to Turing instability, a scheme was performed according to the following steps.

- Choose an interesting point (the branching point in the branched spine or a random point on the central axis in others) in a simulation result and plot the *S*-*Y* curve of this point.
- Calculate the Turing instability space using Equation (3).
- Find the intersection of the *S*-*Y* curve and Turing instability space.
- According to the form of the solution of Equation (2), we have

$$w(r, t) \sim \sum_k c_k e^{\lambda_k t} W_k(r), \tag{4}$$

and calculate the dispersion relation:

$$\lambda = \lambda(k). \tag{5}$$

- Record the maximum of the real part of the eigenvalue (λ_m) and corresponding wavenumbers (k_m).
- Calculate Turing wavelength (Λ) of the point in Step 1:

$$\Lambda = \frac{2\pi}{k_m}. \tag{6}$$

We used Turing instability analysis to explore the difference of mathematical mechanism behind different patterns of dendritic spines in section Turing Instability Underlying Dendritic Spines.

Anatomy of Hippocampal CA1 in SD Rat

In this study, images from Golgi-Cox-stained brain slices from SD rats were compared with simulation results. Golgi-Cox staining was carried out with a commercial Golgi staining kit (Keyijiaxin, Tianjin, China). As soon as they were taken from the skulls, the brains were stored in Golgi-Cox staining solution in a dark place for 2 weeks, and the solution was replaced at intervals of 48 h. Then, brain slices were produced using a vibratome (VT 1000S, Leica, Germany) with a thickness of 150 μm. The slices were placed on slides covered with 2% gelatine. Next, the slices were dyed with ammonia for 60 min; washed with water three times; fixed with Kodak film for 30 min; and then washed with water, dehydrated, cleared, and mounted. Later, dendritic spines in the CA1 region of the hippocampus were imaged under the 100× objective lens with a digital camera. Dendritic trees were detected along CA1 tertiary dendrites derived from secondary dendrites, which started at their point on the primary dendrite. All animal experiments were approved by the Animal Research Ethics Committee, School of Medicine, Nankai University and were performed in accordance with the Animal Management Rules of the Ministry of Health of the People's Republic of China.

RESULTS

Dendritic Spine Shape Factors Research Based on Reaction-Diffusion Model

There are four traditional types of dendritic spines: mushroom-type, stubby-type, thin-type, and branched-type (González-Tapia et al., 2016; Luczynski et al., 2016). In order to research the factors of dendritic spine shape, we firstly proposed a classification method of spine shape based on real spine microimages. Then, we classified a spine simulated by our reaction-diffusion model and found the change rule of dendritic spine shape in different conditions.

Classification Method of Dendritic Spine Shape

At present, the classification methods of dendritic spines shape are qualitative, expert experience-required. To study the shape of dendritic spines quantitatively, metrics to classify dendritic spines need to be determined. Given a branched-type dendritic spine is easy to identify, here we only propose a classification method for the three types of non-branched spines. First, we measured three geometric qualities of a dendritic spine, namely, the height (h), the extreme width of the head (w_{head}), and the extreme width of the neck (w_{neck}), as shown in **Figures 3A–D**. Then, based on these three values, we constructed two following dimensionless metrics:

- Relative average width (RAW) measures the overall thickness of spines, defined as

$$\text{RAW} = \frac{(w_{head} + w_{neck})}{2h}. \tag{7}$$

- Relative constriction width (RCW) measures the difference between the head width and the neck width, defined as

$$\text{RCW} = \frac{(w_{head} - w_{neck})}{h}. \tag{8}$$

We calculated the RAWs and RCWs of eight dendritic spines (including three mushroom-type spines, three stubby-type spines, and two thin-type spines, shown in **Figure 3E**). Thin-type spines have a thin head and neck, so the value of RAW is small. Both the head and neck of stubby-type spines are thick, and the head is thinner or slightly thicker than the neck, so for them, the value of RAW is usually large, and the value of RCW is small or even negative. For the mushroom-type spines, usually have a large head and a thin neck, their values of RAW and RCW are both large. Based on the above analysis, we set the metrics for three types of dendritic spines. As shown in **Figure 3F**, the shape differences among the three types of spines are obvious. We chose RAW = 0.4 and RCW = 0.25 as two criteria to classify the three types.

Finally, we presented a flow chart to distinguish the shapes of dendritic spines (**Figure 3G**). First, if the dendritic spine has a branching structure, it is recognized as a branched-type spine. Second, if the RAW value is lower than 0.4, it is regarded as a thin-type spine. Finally, if the RCW value is lower than 0.25, it is

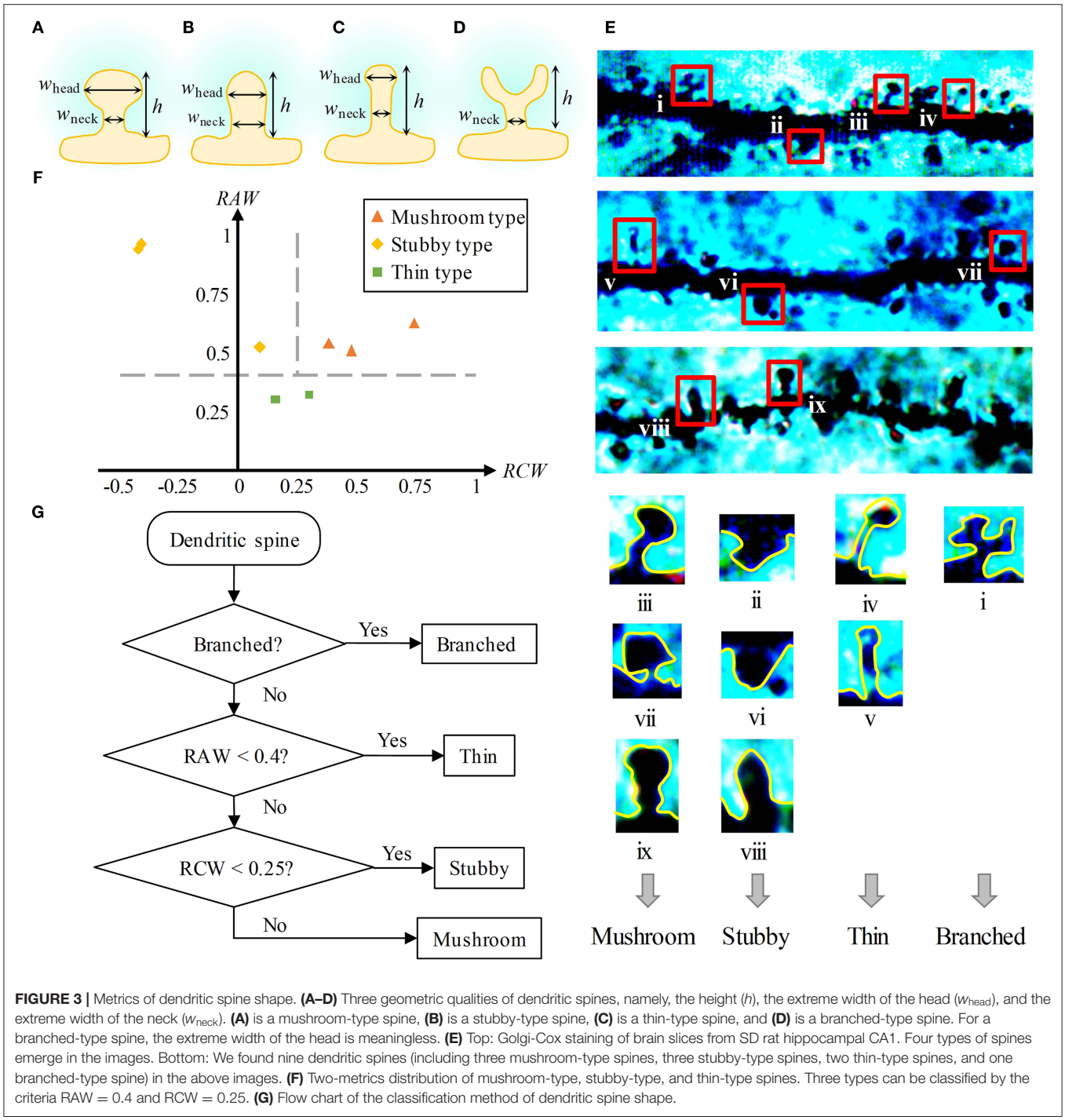


FIGURE 3 | Metrics of dendritic spine shape. **(A–D)** Three geometric qualities of dendritic spines, namely, the height (h), the extreme width of the head (w_{head}), and the extreme width of the neck (w_{neck}). **(A)** is a mushroom-type spine, **(B)** is a stubby-type spine, **(C)** is a thin-type spine, and **(D)** is a branched-type spine. For a branched-type spine, the extreme width of the head is meaningless. **(E)** Top: Golgi-Cox staining of brain slices from SD rat hippocampal CA1. Four types of spines emerge in the images. Bottom: We found nine dendritic spines (including three mushroom-type spines, three stubby-type spines, two thin-type spines, and one branched-type spine) in the above images. **(F)** Two-metrics distribution of mushroom-type, stubby-type, and thin-type spines. Three types can be classified by the criteria RAW = 0.4 and RCW = 0.25. **(G)** Flow chart of the classification method of dendritic spine shape.

identified as a stubby-type spine. Otherwise, it is recognized as a mushroom-type spine.

Consumption Rate of Substrate Dominates the Spine Shape Based on the Reaction-Diffusion Model

In our previous simulation, the rate that substrates are consumed by cells has been shown to play an important role in the branching pattern (Xu et al., 2017). Thus, we assumed that the consumption rate of substrates, namely, the neuron activity, has an effect on the spine shape. To verify this assumption, we performed the following single-spine simulations.

First, to investigate the influence of the consumption rate of substrates (ε) on the shape of dendritic spine, we adjusted the value of parameter ε in Equation (2). We varied the value of ε from 0.01 to 0.9, and part of the obtained results are shown in **Figure 4A** (with different amplification factors) (also

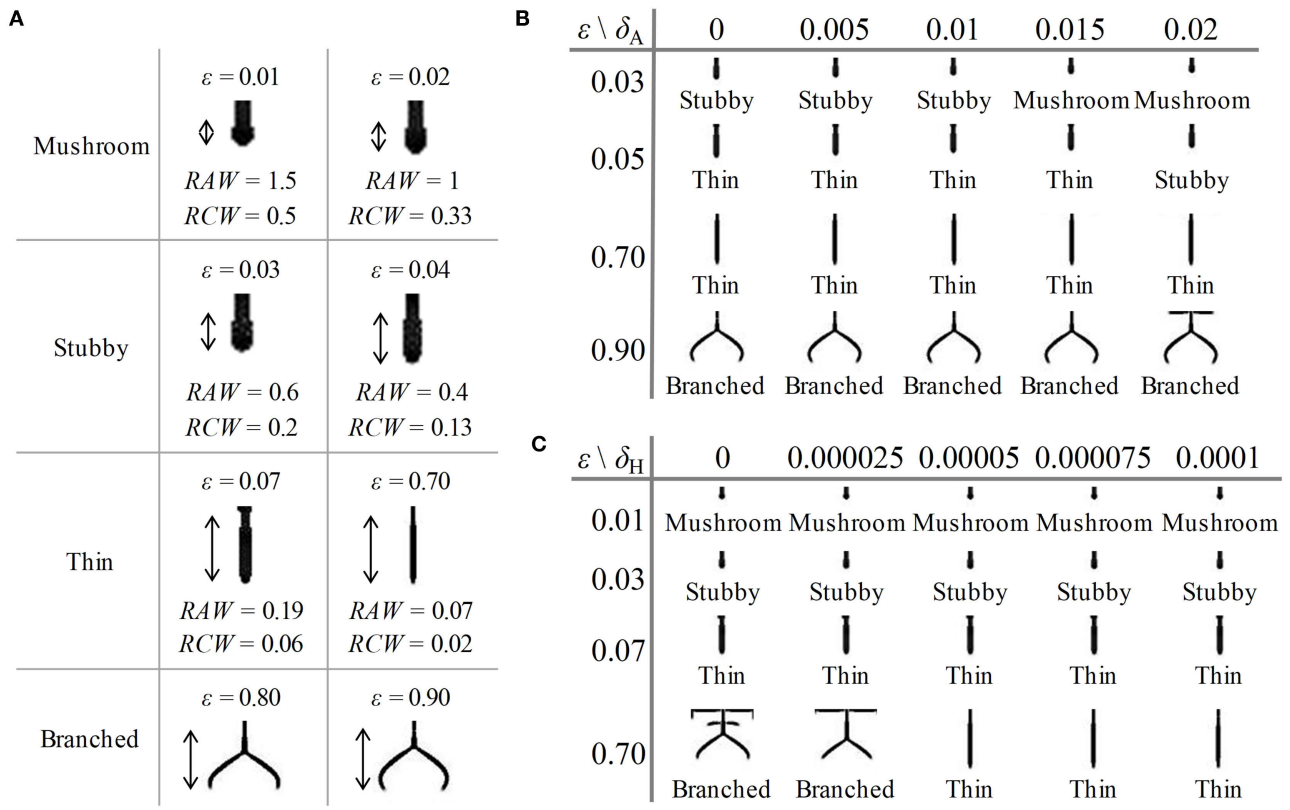


FIGURE 4 | Influence of parameters on the shape of dendritic spines. **(A)** Some of the results of spine simulations for the condition of $\delta_A = 0.01$, $\delta_H = 0.00005$, $0.01 \leq \varepsilon \leq 0.9$ (with different amplification factors). Arrows mark the newborn parts of spines growing from the original state in **Figure 2C**. With the enhancement of neuron activity, the values of RAW and RCW decrease, and all four shapes are obtained. The shape can be mushroom ($0 < \varepsilon \leq 0.02$), stubby ($0.02 < \varepsilon \leq 0.04$), thin ($0.04 < \varepsilon \leq 0.7$) or branched ($0.7 < \varepsilon \leq 0.9$) types. **(B)** The results of spine simulation in the condition of $\delta_A = 0$, 0.005, 0.01, 0.015, and 0.02, $\delta_H = 0.00005$, $\varepsilon =$ 0.03, 0.05, 0.7, and 0.9. With the addition of an exogenous activator, the stubby-type spine becomes mushroom-type, and the thin-type spine becomes stubby-type. **(C)** The results of spine simulation in the condition of $\delta_A = 0.01$, $\delta_H = 0$, 0.000025, 0.00005, 0.000075, and 0.0001, $\varepsilon = 0.01$, 0.03, 0.07, and 0.7. With the addition of an exogenous inhibitors, the branched-type spine becomes thin-type. In **(A–C)**, fixed parameters in Equation (2): $c = 0.002$, $\mu = 0.16$, $\upsilon = 0.04$, $\rho_A = 0.01$, $\rho_H =$ 0.00005, $c_0 = 0.02$, $\gamma = 0.02$, $D_A = 0.02$, $D_H = 0.26$, $D_S = 0.06$, $d = 0.0035$, $e = 0.1$, and $f = 10$.

see **Supplementary Videos 1–4**, respectively). As the value of ε increases, both RAW and RCW values decrease, and the dendritic spine shapes sequentially undergoes mushroom ($0 < \varepsilon \leq 0.02$), stubby ($0.02 < \varepsilon \leq 0.04$), thin ($0.04 < \varepsilon \leq 0.7$), and branched ($0.7 < \varepsilon \leq 0.9$) forms. All four dendritic spine shapes can be obtained with an increase in the consumption rate of substrates. This result indicated that neuron activity regulates the shape of dendritic spine.

In addition, to investigate the influence of exogenous activator (δ_A) and exogenous inhibitor (δ_H) on the shape of dendritic spine, we adjusted the value of parameter δ_A and δ_H in Equation (2), respectively. We varied the values of δ_A under the conditions of $\varepsilon = 0.03$, 0.05, 0.7, and 0.9 and the values of δ_H under the conditions of $\varepsilon = 0.01$, 0.03, 0.07, and 0.7, and the results are shown in **Figures 4B,C** (also see **Supplementary Figures 3, 4**, respectively). According to the results, we found that a stubby-type spine transforms to mushroom-type and a thin-type spine transforms to stubby-type with an increase in δ_A; additionally, a branched-type spine becomes thin-type with an increase in δ_H. However, there is no effect of δ_A on branched-type spines and no effect of δ_H on mushroom-type and stubby-type spines. These results indicated that both δ_A and δ_H also regulate the spine shape but they are not dominating factors compared to the consumption rate of substrates.

Therefore, dendritic spines sequentially undergo in-turn transformation of mushroom-type, stubby-type, thin-type, and branched-type, with an increase in the consumption rate of substrates. In contrast, exogenous activators affect non-branched dendritic spines, and exogenous inhibitors affect branched dendritic spines. Thus, the consumption rate of substrates (neuron activity) determines the shape of dendritic spines.

Dendritic Spine Densities Factors Research Based on Reaction-Diffusion Model

Dendritic spines participate in the formation of most excitatory axodendritic synapses, so the density of spines directly influences the density of synapses. In order to research the factors of dendritic spine density, we simulated a dendrite with spines using

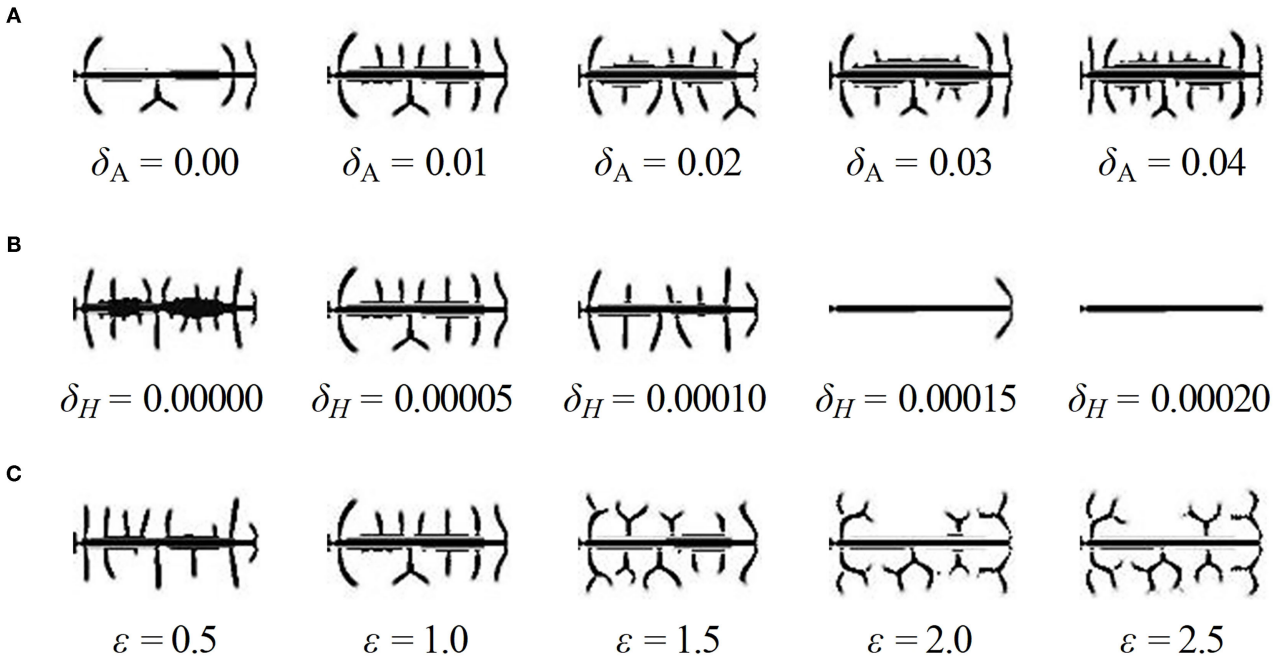


FIGURE 5 | Influence of parameters on the density of dendritic spines. **(A)** The results of dendrite simulation under the conditions of $\delta_A = 0$, 0.01, 0.02, 0.03, and 0.04, $\delta_H = 0.00005$, $\varepsilon = 1$. With the addition of an exogenous activator, the number of dendritic spines increases dramatically. **(B)** The results of dendrite simulation under the conditions of $\delta_A = 0.01$, $\delta_H = 0$, 0.00005, 0.0001, 0.00015, and 0.0002, $\varepsilon = 1$. With the addition of an exogenous inhibitor, the number of dendritic spines decreases dramatically. **(C)** The results of dendrite simulation under the conditions of $\delta_A = 0.01$, $\delta_H = 0.00005$, $\varepsilon = 0.5$, 1, 1.5, 2, and 2.5. With increased neuron activity, the shape varies from non-branched-dominant to branched-dominant. Branched-type spines take up more space; thus, the growth of surrounding spines is inhibited. In addition, the parameter ε has no effect on the density; for example, the densities are the same under the conditions of $\varepsilon = 2$ and $\varepsilon = 2.5$. In **(A–C)**, fixed parameters in Equation (2): $c = 0.002$, $\mu = 0.16$, $\upsilon = 0.04$, $\rho_A = 0.02$, $\rho_H = 0.00005$, $c_0 = 0.05$, $\gamma = 0.02$, $D_A = 0.02$, $D_H = 0.26$, $D_S = 0.06$, $d = 0.0035$, $e = 0.1$, and $f = 10$.

the reaction-diffusion model and found the relationship between dendritic spine density and key factors. Moreover, we observed the decrease of spine density in the hippocampal CA1 in rats with glioma and proposed a potential reason for this phenomenon by comparing the simulation results and observation results. Further, we used Turing instability to explain the mathematical mechanism behind the above parameters regulating spine density and found that an exogenous inhibitor and activator changes Turing wavelength through which to regulate spine densities.

Exogenous Substances Regulate Spine Density

To investigate the factors of dendritic spine density, we next simulated different spine densities which seen across multiple spines through dendrite simulations. In our previous research, we found that the rates of activator and inhibitor secretion from cells have been shown to play an important role in the density of side branching (Guo et al., 2014a). Similarly, it is reasonable for us to assume that exogenous activator and inhibitor are two key factors influencing the density of dendritic spines.

Firstly, in order to find out the effect of exogenous activator and inhibitor on the spine density, we adjusted the values of the two parameters δ_A and δ_H based on standard values of $\delta_A = 0.01$, $\delta_H = 0.00005$, and $\varepsilon = 1$, and we obtained two groups of results (**Figures 5A,B**). The results showed the density of dendritic spines is positively correlated with δ_A and negatively correlated with δ_H.

Next, we adjusted the values of the parameter ε to find whether the consumption rate of substrates is another factor of density, and the results are shown in **Figure 5C** (also see **Supplementary Figures 1, 2**, respectively). We noticed that the spine shape varied from non-branched-dominant to branched-dominant when ε varies from 0.5 to 2.0. Meanwhile, the spine densities have not significant changes when ε varies; for example, the densities are the same under the conditions of $\varepsilon = 2$ and $\varepsilon = 2.5$.

Through dendrite simulations, we found that exogenous activators increase the density of spines, while exogenous inhibitors have the opposite effect. In comparison to exogenous substances, neuron activity has no significant effect on the density.

Application in the Hippocampal CA1 of Rats

The hippocampus plays an important role in memory function and cognitive abilities (Muller et al., 1996). Certain diseases, such as glioma, affect the developmental patterns of dendritic spines on hippocampal neurons. It has also been reported that the impairment of neurocognitive function is a common consequence of glioma, in both glioma patients (Wefel et al.,

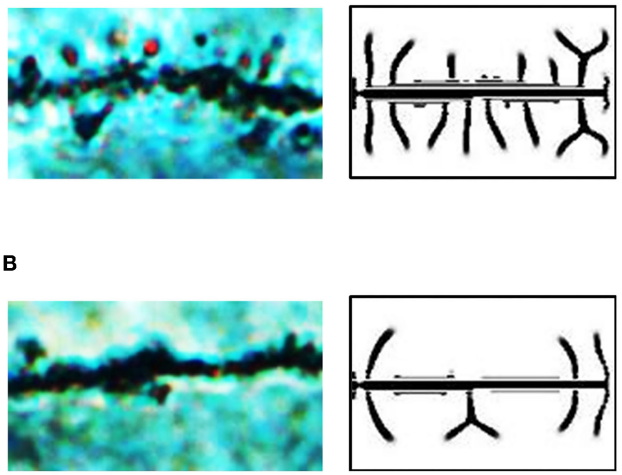


FIGURE 6 | Exogenous inhibitor decreases the spine density in hippocampal CA1 in SD rats with glioma. **(A)** A microscopic image of dendritic spines in the brain of a rat in the sham group. There are dense spines in this image, similar to those in our dendrite simulation results under the condition of $\delta_A = 0.01$, $\delta_H = 0$, $\varepsilon = 1$. **(B)** A microscopic image of dendritic spines in the brain of a rat in the glioma group. There are sparse spines in this image, similar to those in our dendrite simulation results under the condition of $\delta_A = 0.01$, $\delta_H = 0.0001$, $\varepsilon = 1$. In **(A,B)**, fixed parameters in Equation (2): $c = 0.002$, $\mu = 0.16$, $\upsilon = 0.04$, $\rho_A = 0.02$, $\rho_H = 0.00005$, $c_0 = 0.05$, $\gamma = 0.02$, $D_A = 0.02$, $D_H = 0.26$, $D_S = 0.06$, $d = 0.0035$, $e = 0.1$, and $f = 10$.

2016; Van Kessel et al., 2017) and glioma animal models (Wang et al., 2010; Hao et al., 2018). Through anatomy and neuron microimaging (see section Anatomy of Hippocampal CA1 in SD Rat for detail), we found that dendritic spines in rats with glioma were less dense (**Figures 6A,B**, also see **Supplementary Videos 5, 6**, respectively).

To study the reasons for various densities in the rat hippocampal CA1, we compared the microscopic images of neurons with our simulation results. It seems that the spine patterns in the brains of the rat sham group were similar to those in the simulation results under the condition of $\delta_A = 0.01$, $\delta_H = 0$, $\varepsilon = 1$, while the spine patterns in the brains of the rat glioma group were similar to those in the simulation results under the condition of $\delta_A = 0.01$, $\delta_H = 0.0001$, $\varepsilon = 1$ (**Figure 5B**). Thus, we considered that the addition of exogenous inhibitors is a potential reason for the decrease of dendritic spine density caused by glioma.

Turing Instability Underlying Dendritic Spines

Turing pointed out that the diffusion of chemical substances will break the original equilibrium state of substance concentration, which is called Turing instability (Turing, 1952). Branching patterns can only be generated from models with Turing instability. In order to qualitatively analyze the Turing instability, equilibrium position, and periodicity of the model solution, we have proposed a Turing instability analysis method using dispersion relation in previous research (Guo et al., 2014b) and found that the Turing wavelength is the internal factor causing the change of branching pattern of a lung (Xu et al., 2017) (see section Turing Instability Analysis Method for more details).

Exogenous substances have an effect on the Turing instability space in which a stable pattern can appear, and have no effect on the *S*-*Y* curve that shows the concentration relationship between substrate and cytoskeleton during development, which can be derived from Equation (3). We adjusted the values of the parameters δ_A from 0 to 0.2 and then drew an *S*-*Y* curve and Turing instability space in the *S*-*Y* space (**Figure 7A**). Three intersection points of Turing instability spaces and the corresponding *S*-*Y* curves were marked with black points. These points were substituted into the Equations (4–6) in order to calculate the Turing wavelengths (**Figure 7B**). An increase in parameter δ_A decreases the Turing wavelength. Similarly, we found the intersection points and calculated Turing wavelength under the conditions of $\delta_H = 0$, 0.00005, and 0.0001 (**Figures 7C,D**). An increase in parameter δ_H increases the Turing wavelength.

As the Turing wavelength implies the spatial periodicity of spines, it is negatively correlated to the density of dendritic spines. In conclusion, exogenous activators make the Turing wavelength smaller and cause an increase in density of dendritic spines, while exogenous inhibitors increase the Turing wavelength and cause a decrease in the density of dendritic spines.

DISCUSSION

In recent years, various chemicals have been reported to be capable of regulating the process of dendritic spine development. Our research may explore their regulation mechanism in a mathematical view. For example, the actin filaments (F-actin) were considered to be key in regulating the shape of dendritic spines (Miermans et al., 2017). We found the cytoskeleton was one key factor to regulate cell morphology. Hence, F-actin might be considered as the cytoskeleton (*Y*) in our model. It has been found that drebrin is an actin-binding protein in the dendritic spine, and its overexpression causes spine elongation (Hayashi and Shirao, 1999; Koganezawa et al., 2017; Hanamura et al., 2018). Bernstein reported that cofilin severs F-actin, contributing to actin dynamics (Bernstein and Bamburg, 2010). In addition, Calabrese suggested that dendritic spine growth correlates with decreased cofilin activity (Calabrese et al., 2014). According to our simulation results, drebrin and cofilin are similar to the functions of the activator (*A*) and the inhibitor (*H*) in our model respectively. Adenosine-triphosphate (ATP) is closely related to F-actin polymerization and depolymerization (Katkar et al., 2018; Merino et al., 2018), which implies that ATP may correspond to the substrate (*S*) in our model. Based on these hypotheses, we described our inferences as follows: (1) the overexpression of drebrin promotes the binding of F-actin and increases the density of dendritic spines, (2) the overexpression of cofilin hinders the binding of F-actin and decreases the density of dendritic spines, (3) the increase in ATP consumption during the process of creating F-actin results in a different F-actin pattern and causes spines to

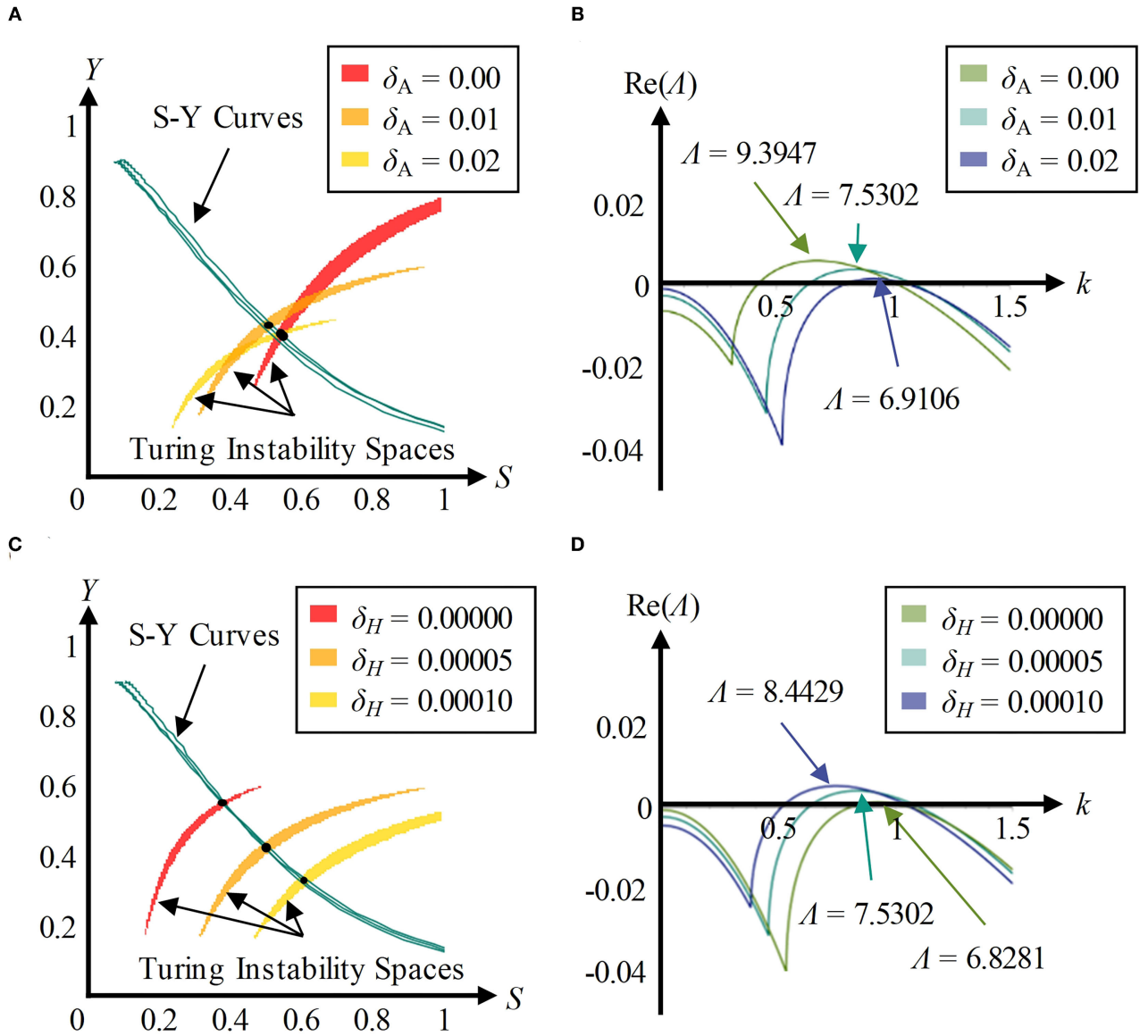


FIGURE 7 | Exogenous substances regulate dendritic spine density through Turing wavelength. **(A)** The S-Y curve and Turing instability space in three conditions of parameter δ_A ($\delta_H = 0.00005$, $\varepsilon = 1$). The yellow, orange, and red regions represent Turing instability spaces in the conditions of $\delta_A = 0$, 0.01, and 0.02, respectively. The cyan lines represent S-Y curves, and the black points represent the intersections. An exogenous activator affects the size and location of the Turing instability space, rather than the S-Y curves (artificial errors in choosing intersections result in non-coincidence). **(B)** The dispersion relationship curves in the conditions of $\delta_A = 0$, 0.01, and 0.02 ($\delta_H = 0.00005$, $\varepsilon = 1$). Adding an exogenous activator decreases the Turing wavelength. **(C)** The S-Y curve and Turing instability space in three conditions of parameter δ_H ($\delta_A = 0.01$, $\varepsilon = 1$). The yellow, orange, and red regions represent Turing instability spaces in the conditions of $\delta_H = 0$, 0.00005, and 0.0001, respectively. The cyan lines represent S-Y curves, and the black points represent the intersections. An exogenous inhibitor affects the size and location of the Turing instability space, rather than the S-Y curves (artificial errors in choosing intersections result in non-coincidence). **(D)** The dispersion relationship curves in the conditions of $\delta_H = 0$, 0.00005, and 0.0001 ($\delta_A = 0.01$, $\varepsilon = 1$). Adding an exogenous inhibitor increases the Turing wavelength.

become mushroom-type, stubby-type, thin-type and branched-type, in turn.

The verification experiments of morphogens is helpful to the correction of model parameters and the support of the conclusion in this work. Here, we proposed two ideas to verify the morphogens mentioned above: (1) research on the quantitative relationship between spine density and the addition of a substance that influences the expression of drebrin or cofilin, and (2) research on the quantitative relationship between spine shape distribution and ATP consumption during the process of creating F-actin. Moreover, in order to compare the spatiotemporal parameters between simulations and verification experiments quantitatively, 3D simulation is necessary.

With our method, certain diseases could be systematically investigated at the level of chemical reactions. For example, the anomalous rise of rho GTPase Rac1 activity inhibited cofilin in mice with Fragile X syndrome because of a trinucleotide expansion in the FMR1 gene on the X chromosome (Pyronneau et al., 2017). In our model, the decrease in δ_H decreases the concentration of the inhibitor (H), which results in dense dendritic spines. In another study, the intrathecal administration of latrunculin A, an actin-depolymerizing agent, in mice resulted in a decrease in F-actin levels and symptoms of Alzheimer's disease. Conversely, the intrathecal administration of jasplakinolide, a molecule that stabilizes F-actin, in mice restored F-actin levels and improved symptoms (Kommaddi et al., 2018). The effects of latrunculin A and jasplakinolide are similar to those of exogenous inhibitors and exogenous activators in this model, respectively. Exogenous activators promote the synthesis of the cytoskeleton, while exogenous inhibitors promote the decomposition of the cytoskeleton.

In conclusion, we were devoted to revealing the mechanism of the development patterns of dendritic spines. The results show that the consumption rate of substrate dominates the shape, while the addition of exogenous activators and exogenous inhibitors dominates the density. Our work provided a potential explanation for the phenomenon that sparser spines in the brains of SD rats with glioma and maybe also explain some diseases reported in the literature, such as Fragile X syndrome and Alzheimer's disease. Our research provides novel and fresh insight into the development patterns of dendritic spines, helping search treatment methods for related diseases.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The animal study was reviewed and approved by Animal Research Ethics Committee, School of Medicine, Nankai University.

AUTHOR CONTRIBUTIONS

YJ: conceptualization, methodology, software, formal analysis, investigation, writing-original draft, writing-review and editing, and visualization. QZ: conceptualization, methodology, writing-original draft, writing-review and editing, and funding acquisition. HY: methodology, validation, formal analysis, investigation, writing-original draft, writing-review and editing, and visualization. SG: methodology, software, formal analysis, data curation, and writing-review and editing. MS: validation, formal analysis, writing-review and editing, and funding acquisition. ZY: conceptualization, writing-review and editing, and supervision. XZ: conceptualization, writing-review and editing, supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by National Natural Science Foundation of China (62027812, U1813210, 61903201) and Opening Foundation of Beijing Advanced Innovation Center for Intelligent Robots and Systems (2019IRS05).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbot.2021.563682/full#supplementary-material

REFERENCES

Bernstein, B. W., and Bamburg, J. R. (2010). ADF/Cofilin: a functional node in cell biology. *Trends Cell Biol.* 20, 187–195. doi: 10.1016/j.tcb.2010.01.001

Bloodgood, B., and Sabatini, B. (2005). Neuronal activity regulates diffusion across the neck of dendritic spines. *Science* 310, 866–869. doi: 10.1126/science.1114816

Bourne, J., and Harris, K. (2007). Do thin spines learn to be mushroom spines? *Curr. Opin. Neurobiol.* 17, 381–386. doi: 10.1016/j.conb.2007.04.009

Calabrese, B., Saffin, J.-M., and Halpain, S. (2014). Activity-dependent dendritic spine shrinkage and growth involve downregulation of cofilin via distinct mechanisms. *PLoS ONE* 9:e94787. doi: 10.1371/journal.pone.0094787

Garfinkel, A., Tintut, Y., Petrasek, D., Boström, K., and Demer, L. L. (2004). Pattern formation by vascular mesenchymal cells. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9247–9250. doi: 10.1073/pnas.0308436101

Gierer, A., and Meinhardt, H. (1972). A theory of biological pattern formation. *Kybernetik* 12, 30–39. doi: 10.1007/BF00289234

Glantz, L. A., and Lewis, D. A. (2000). Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. *JAMA Psychiatry* 57, 65–73. doi: 10.1001/archpsyc.57.1.65

González-Tapia, D., Martínez-Torres, N. I., Hernández-González, M., Guevara, M. A., and González-Burgos, I. (2016). Plastic changes to dendritic spines on layer V pyramidal neurons are involved in the rectifying role of the prefrontal cortex during the fast period of motor learning. *Behav. Brain Res.* 298, 261–267. doi: 10.1016/j.bbr.2015.11.013

Guo, Y., Chen, T.-H., Zeng, X., Warburton, D., Boström, K. I., Ho, C.-M., et al. (2014a). Branching patterns emerge in a mathematical model of the dynamics of lung development. *J. Physiol.* 592, 313–324. doi: 10.1113/jphysiol.2013.261099

Guo, Y., Sun, M., Garfinkel, A., and Zhao, X. (2014b). Mechanisms of side branching and tip splitting in a model of branching morphogenesis. *PLoS ONE* 9:e102718. doi: 10.1371/journal.pone.0102718

Hagiwara, M., Peng, F., and Ho, C.-M. (2015). *In vitro* reconstruction of branched tubular structures from lung epithelial cells in high cell concentration gradient environment. *Sci. Rep.* 5:8054. doi: 10.1038/srep08054

Hanamura, K., Kamata, Y., Yamazaki, H., Kojima, N., and Shirao, T. (2018). Isoform-dependent regulation of drebrin dynamics in dendritic spines. *Neuroscience* 379, 67–76. doi: 10.1016/j.neuroscience.2018.02.038

Hao, S., Gao, J., Hui, W., Zhang, Y., Pavlov, A., Ge, H., et al. (2018). AG-1031 and AG-1503 improve cognitive deficits by promoting apoptosis and inhibiting autophagy in C6 glioma model rats. *Brain Res.* 1699, 1–8. doi: 10.1016/j.brainres.2018.06.026

Harris, K. M., and Kater, S. B. (1994). Dendritic spines: cellular specializations imparting both stability and flexibility to synaptic function. *Annu. Rev. Neurosci.* 17, 341–371. doi: 10.1146/annurev.ne.17.030194.002013

Hayashi, K., Ishikawa, R., Ye, L.-H., He, X.-L., Takata, K., Kohama, K., et al. (1996). Modulatory role of drebrin on the cytoskeleton within dendritic spines in the rat cerebral cortex. *J. Neurosci.* 16, 7161–7170. doi: 10.1523/JNEUROSCI.16-22-07161.1996

Hayashi, K., and Shirao, T. (1999). Change in the shape of dendritic spines caused by overexpression of drebrin in cultured cortical neurons. *J. Neurosci.* 19, 3918–3925. doi: 10.1523/JNEUROSCI.19-10-03918.1999

Kasai, H., Fukuda, M., Watanabe, S., Hayashi-Takagi, A., and Noguchi, J. (2010). Structural dynamics of dendritic spines in memory and cognition. *Trends Neurosci.* 33, 121–129. doi: 10.1016/j.tins.2010.01.001

Kasai, H., Matsuzaki, M., Noguchi, J., Yasumatsu, N., and Nakahara, H. (2003). Structure–stability–function relationships of dendritic spines. *Trends Neurosci.* 26, 360–368. doi: 10.1016/S0166-2236(03)00162-0

Katkar, H. H., Davtyan, A., Durumeric, A. E., Hocky, G. M., Schramm, A. C., Cruz, E. M. D. L., et al. (2018). Insights into the cooperative nature of ATP hydrolysis in actin filaments. *Biophys. J.* 115, 1589–1602. doi: 10.1016/j.bpj.2018.08.034

Koganezawa, N., Hanamura, K., Sekino, Y., and Shirao, T. (2017). The role of drebrin in dendritic spines. *Mol. Cell. Neurosci.* 84, 85–92. doi: 10.1016/j.mcn.2017.01.004

Kommaddi, R. P., Das, D., Karunakaran, S., Nanguneri, S., Bapat, D., Ray, A., et al. (2018). Aβ mediates F-actin disassembly in dendritic spines leading to cognitive deficits in Alzheimer's disease. *J. Neurosci.* 38, 1085–1099. doi: 10.1523/JNEUROSCI.2127-17.2017

Kondo, S., and Asai, R. (1995). A reaction–diffusion wave on the skin of the marine angelfish Pomacanthus. *Nature* 376, 765–768. doi: 10.1038/376765a0

Landge, A., Jordan, B., Diego, X., and Mueller, P. (2020). Pattern formation mechanisms of self-organizing reaction-diffusion systems. *Dev. Biol.* 460, 2–11. doi: 10.1016/j.ydbio.2019.10.031

Luczynski, P., Whelan, S. O., O'Sullivan, C., Clarke, G., Shanahan, F., Dinan, T. G., et al. (2016). Adult microbiota-deficient mice have distinct dendritic morphological changes: differential effects in the amygdala and hippocampus. *Eur. J. Neurosci.* 44, 2654–2666. doi: 10.1111/ejn.13291

McNeill, T. H., Brown, S. A., Rafols, J. A., and Shoulson, I. (1988). Atrophy of medium spiny I striatal dendrites in advanced Parkinson's disease. *Brain Res.* 455, 148–152. doi: 10.1016/0006-8993(88)90124-2

Meinhardt, H. (1976). Models and hypotheses. *Differentiation* 6, 117–123. doi: 10.1111/j.1432-0436.1976.tb01478.x

Merino, F., Pospich, S., Funk, J., Wagner, T., Küllmer, F., Arndt, H.-D., et al. (2018). Structural transitions of F-actin upon ATP hydrolysis at near-atomic resolution revealed by cryo-EM. *Nat. Struct. Mol. Biol.* 25, 528–537. doi: 10.1038/s41594-018-0074-0

Miermans, C. A., Kusters, R. P. T., Hoogenraad, C. C., and Storm, C. (2017). Biophysical model of therole of actin remodeling on dendritic spine morphology. *PLoS ONE* 12:e0170113. doi: 10.1371/journal.pone.0170113

Miura, T., Shiota, K., Morriss-Kay, G., and Maini, P. (2006). Mixed-mode pattern in doublefoot mutant mouse limb–turing reaction-diffusion model on a growing domain during limb development. *J. Theor. Biol.* 240, 562–573. doi: 10.1016/j.jtbi.2005.10.016

Muller, R., Stead, M., and Pach, J. (1996). The hippocampus as a cognitive graph. *J. Gen. Physiol.* 107, 663–694 doi: 10.1085/jgp.107.6.663

Murray, J. (1982). Parameter space for turing instability in reaction diffusion mechanisms: a comparison of models. *J. Theor. Biol.* 98, 143–163. doi: 10.1016/0022-5193(82)90063-7

Okuda, S., Miura, T., Inoue, Y., Adachi, T., and Eiraku, M. (2018). Combining turing and 3d vertex models reproduces autonomous multicellular morphogenesis with undulation, tubulation, and branching. *Sci. Rep.* 8:2386. doi: 10.1038/s41598-018-24858-2

O'Neal, M., Stallings, N., and Malter, J. (2018). Alzheimer's disease, dendritic spines, and calcineurin inhibitors: a new approach? *ACS Chem. Neurosci.* 9, 1233–1234. doi: 10.1021/acschemneuro.8b00213

Portera-Cailliau, C., Pan, D., and Yuste, R. (2003). Activity-regulated dynamic behavior of early dendritic protrusions: evidence for different types of dendritic filopodia. *J. Neurosci. Off. J. Soc. Neurosci.* 23, 7129–7142. doi: 10.1523/JNEUROSCI.23-18-07129.2003

Pyronneau, A., He, Q., Hwang, J.-Y., Porch, M., Contractor, A., and Zukin, R. S. (2017). Aberrant Rac1-cofilin signaling mediates defects in dendritic spines, synaptic function, and sensory perception in fragile X syndrome. *Science Signal.* 10:eaan0852. doi: 10.1126/scisignal.aan0852

Ratliff, W. A., Mervis, R. F., Citron, B. A., Schwartz, B., Rubovitch, V., Schreiber, S., et al. (2019). Mild blast-related TBI in a mouse model alters amygdalar neurostructure and circuitry. *Exp. Neurol.* 315, 9–14. doi: 10.1016/j.expneurol.2019.01.020

Richards, P., Didszun, C., Campesan, S., Simpson, A., Horley, B., Young, K., et al. (2011). Dendritic spine loss and neurodegeneration is rescued by rab11 in models of huntington's disease. *Cell Death Differ.* 18, 191–200. doi: 10.1038/cdd.2010.127

Saleem, M., and Ali, I. (2018). Numerical simulations of turing patterns in a reaction-diffusion model with the chebyshev spectral method. *Eur. Phys. J. Plus* 133:399. doi: 10.1140/epjp/i2018-12265-9

Sweet, R. A., Henteleff, R. A., Zhang, W., Sampson, A. R., and Lewis, D. A. (2008). Reduced dendritic spine density in auditory cortex of subjects with schizophrenia. *Neuropsychopharmacology* 34, 374–389. doi: 10.1038/npp.2008.67

Turing, A. M. (1952). The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond.* 237, 37–72. doi: 10.1098/rstb.1952.0012

Van Kessel, E. E., Baumfalk, A. J. E., van Zandvoort, M., Robe, P., and Snijders, T. J. (2017). Tumor-related neurocognitive dysfunction in patients with diffuse glioma: a systematic review of neurocognitive functioning prior to anti-tumor treatment. *J. Neurooncol.* 134, 9–18. doi: 10.1007/s11060-017-2503-z

Varner, V., and Nelson, C. (2014). Cellular and physical mechanisms of branching morphogenesis. *Development* 141, 2750–2759. doi: 10.1242/dev.104794

Wang, Y.-Y., Liu, S.-C., Yang, Z., and Zhang, T. (2010). Impaired hippocampal synaptic plasticity in C6 glioma-bearing rats. *J. Neurooncol.* 103, 469–477. doi: 10.1007/s11060-010-0447-7

Wefel, J., Noll, K., and Scheurer, M. (2016). Neurocognitive functioning and genetic variation in patients with primary brain tumors. *Lancet Oncol.* 17, e97–e108. doi: 10.1016/S1470-2045(15)00380-0

Xu, H., Sun, M., and Zhao, X. (2017). Turing mechanism underlying a branching model for lung morphogenesis. *PLoS ONE* 12:e0174946. doi: 10.1371/journal.pone.0174946

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 14 June 2021
doi: 10.3389/fonc.2021.657560

Breast Cancer Histopathological Images Recognition Based on Low Dimensional Three-Channel Features

Yan Hao[1], Shichang Qiao[2], Li Zhang[2], Ting Xu[2], Yanping Bai[2], Hongping Hu[2], Wendong Zhang[3] and Guojun Zhang[3]*

[1] School of Information and Communication Engineering, North University of China, Taiyuan, China, [2] Department of Mathematics, School of Science, North University of China, Taiyuan, China, [3] School of Instrument and Electronics, Key Laboratory of Dynamic Testing Technology, North University of China, Taiyuan, China

Breast cancer (BC) is the primary threat to women's health, and early diagnosis of breast cancer is imperative. Although there are many ways to diagnose breast cancer, the gold standard is still pathological examination. In this paper, a low dimensional three-channel features based breast cancer histopathological images recognition method is proposed to achieve fast and accurate breast cancer benign and malignant recognition. Three-channel features of 10 descriptors were extracted, which are gray level co-occurrence matrix on one direction (GLCM1), gray level co-occurrence matrix on four directions (GLCM4), average pixel value of each channel (APVEC), Hu invariant moment (HIM), wavelet features, Tamura, completed local binary pattern (CLBP), local binary pattern (LBP), Gabor, histogram of oriented gradient (Hog), respectively. Then support vector machine (SVM) was used to assess their performance. Experiments on BreaKHis dataset show that GLCM1, GLCM4 and APVEC achieved the recognition accuracy of 90.2%-94.97% at the image level and 89.18%-94.24% at the patient level, which is better than many state-of-the-art methods, including many deep learning frameworks. The experimental results show that the breast cancer recognition based on high dimensional features will increase the recognition time, but the recognition accuracy is not greatly improved. Three-channel features will enhance the recognizability of the image, so as to achieve higher recognition accuracy than gray-level features.

Keywords: breast cancer, histopathological images recognition, feature extraction, low dimensional features, three-channel features

OPEN ACCESS

Edited by:
Francesco Rundo, STMicroelectronics, Italy

Reviewed by:
Chen Liu, Army Medical University, China
Alceu Britto, Pontifical Catholic University of Parana, Brazil

****Correspondence:***
Yanping Bai
baiyp666@163.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *23 January 2021*
Accepted: *11 May 2021*
Published: *14 June 2021*

Citation:
Hao Y, Qiao S, Zhang L, Xu T, Bai Y, Hu H, Zhang W and Zhang G (2021) Breast Cancer Histopathological Images Recognition Based on Low Dimensional Three-Channel Features. Front. Oncol. 11:657560. doi: 10.3389/fonc.2021.657560

INTRODUCTION

Cancer has become one of the major public health problems that seriously threaten the health of people. The incidence and mortality of breast cancer have been rising continuously in recent years. Early accurate diagnosis is the key to improve the survival rate of patients. Mammogram is the first step of early diagnosis, but it is difficult to detect cancer in the dense breast of adolescent women, and the X-ray radiation poses a threat to the health of patients and radiologists. Computed tomography (CT) is a localized examination, which can not be used to judge that a patient is

suffering from breast cancer according to the observed abnormalities. The gold standard for breast cancer diagnosis is still pathological examination. Pathological examinations usually obtain tumor specimens through puncture, excision, etc. And then stain them with hematoxylin and eosin (H&E) stains. Hematoxylin binds deoxyribonucleic acid (DNA) to highlight the nucleus, while eosin binds proteins and highlights other structures. Accurate diagnosis of breast cancer requires experienced histopathologists, and it requires a lot of time and effort to complete this task. In addition, the diagnosis results of different histopathologists are not the same, which strongly depends on the prior knowledge of histopathologists. It resulting in lower diagnosis consistency, and the average diagnosis accuracy is only 75% (1).

Currently, breast cancer diagnosis based on histopathological images is facing three major challenges. Firstly, there is a shortage of experienced histopathologists around the world, especially in some underdeveloped areas and small hospitals. Secondly, the diagnosis of histopathologist is subjective and there is no objective evaluation basis. Whether the diagnosis is correct or not depends entirely on the histopathologists' prior knowledge. Thirdly, the diagnosis of breast cancer based on histopathological images is very complicated, time-consuming and labor-intensive, which is inefficient in the era of big data. In face of these problems, an efficient and objective breast cancer diagnosis method is urgently needed to alleviate the workload of histopathologists.

With the rapid development of computer-aided diagnosis (CAD), it has been gradually applied to the clinical field. The CAD system cannot completely replace the doctor, but it can be used as a "second reader" to assist doctors in diagnosing diseases. However, there are many false positive areas detected by the computer, which will take a lot of time of doctors to re-evaluate the results prompted by the computer, resulting in a decrease in the accuracy and efficiency. Therefore, how to improve the sensitivity of computer-aided tumor detection method, while greatly reducing the false positive detection rate, improve the overall performance of the detection method is a subject to be studied.

In recent years, machine learning has been successfully applied to image recognition, object recognition, and text classification. With the advancement of computer-aided diagnosis technology, machine learning has also been successfully applied to breast cancer diagnosis (2–8). There are two common methods, histopathological images classification based on artificial feature extraction and traditional machine learning methods, and histopathological images classification based on deep learning methods. Histopathological images classification based on artificial feature extraction and traditional machine learning methods needs manual design of features, but it does not require equipment with high performance and has advantages in computing time. However, histopathological images classification based on deep learning, especially convolutional neural network (CNN), often requires a large number of labeled training samples, while the labeled data is difficult to obtain. The labeling of lesions is a time-consuming and laborious work, which takes a lot of time even for very experienced histopathologists.

The key of traditional histopathological images classification is feature extraction. The common features include color features, morphological features, texture features, statistical features etc. Spanhol et al. (9) introduced a publicly available breast cancer histopathology dataset (BreaKHis), and they extracted LBP, CLBP, gray level co-occurrence matrix (GLCM), Local phase quantization (LPQ), parameter-free threshold adjacency statistics (PFTAS) and one keypoint descriptor named ORB features, and 1-nearest neighbor (1-NN), quadratic linear analysis (QDA), support vector machines (SVMs), and random forests (RF) were used to assess the aforementioned features, with an accuracy range from 80% to 85%. Pendar et al. (10) introduced a representation learning-based unsupervised domain adaptation on the basis of (9) and compared it with the results of CNN. Anuranjeeta et al. (11) proposed a breast cancer recognition method based on morphological features. 16 morphological features were extracted, and 8 classifiers were used for recognition, the accuracy is about 80%. The authors in (12–14) proposed breast cancer recognition methods based on texture features. Particularly, Carvalho et al. (14) used phylogenetic diversity indexes to characterize the types of breast cancer. Sudharshan et al. (15) compared 12 multi-instance learning methods based on PFTAS and verified that multi-instance learning is more effective than single-instance learning. But none of them considered the color channel of the image. Fang et al. (16) proposed a framework called Local Receptive Field based Extreme Learning Machine with Three Channels (3C-LRF-ELM), which can automatically extract histopathological features to diagnose whether there is inflammation. In addition, in order to reduce the recognition time and the complexity of the algorithms, this paper is committed to achieving high recognition accuracy with low dimensional features.

Deep learning methods, especially CNN, can achieve more accurate cancer recognition (17–25) for it's ability to extract powerful high-level features compared with traditional image recognition methods. For example, Spanhol et al. (17) used the existing AlexNet to test the BreaKHis dataset, and its recognition accuracy was significantly higher than their previous work (9). The authors in (18–21, 25) used different CNN frameworks and obtained the recognition accuracy of more than 90% on the two-class problem of the BreaKHis dataset. Benhammou et al. (22) comprehensively surveyed the researches based on BreaKHis datasets from the magnification-specific binary, magnification independent binary, magnification specific multi-category and magnification independent multi-category four aspects, and proposed a magnification independent multi-category method based on CNN, which is rarely considered in previous studies. The works (23–26) also achieved good performance on the Bioimaging 2015 dataset. Both the BreaKHis and Bioimaging 2015 are the challenging datases for breast cancer detection. Due to the drawbacks of model training, most researchers' research were based on models that have been well trained through other datasets and verified by histopathological images. Few people trained a complete model with histopathological images for the lack of labeled data.

In order to reduce the workload of histopathologists and allow them to spend more time on the diagnosis of more complex diseases, efficient and fast computer-aided diagnosis methods are of urgent need. This paper proposed a breast cancer histopathological images recognition method based on low dimensional three-channel features. The features of the three channels of the image were extracted respectively, then the three-channel features were fused to realize better breast cancer histopathological images recognition for the image level and the patient level. The framework is shown in **Figure 1**.

The contributions of this paper are as follows:

1) proposed a histopathological images recognition method based on three-channel features,
2) proposed a histopathological images recognition method based on low dimensional features,
3) it is a method with high accuracy and fast recognition speed,
4) it is a method easy to implement.

The rest of the paper is organized as follows: in Section 2 the feature extraction methods are introduced, the experiments and results analysis are given in Section 3, and Section 4 concludes the work.

FEATURE EXTRACTION

Gray Level Co-Occurrence Matrix

Gray level co-occurrence matrix is a common method to describe the texture of an image by studying its spatial correlation characteristics. In 1973, Haralick et al. first used GLCM to describe texture features (27). In our experiments, we calculated the GLCM of 256 gray levels in one direction 0° and four directions 0°, 45°, 90°, 135°, respectively. Then, according to the GLCM, 22 related features were calculated: autocorrelation, contrast, 2 correlation, cluster probability, cluster shade, dissimilarity, energy, entropy, 2 homogeneity, maximum probability, sum of squares, sum average, sum variance, sum entropy, difference variance, difference entropy, 2 information measures of correlation, inverse difference, inverse difference moment (27–29).

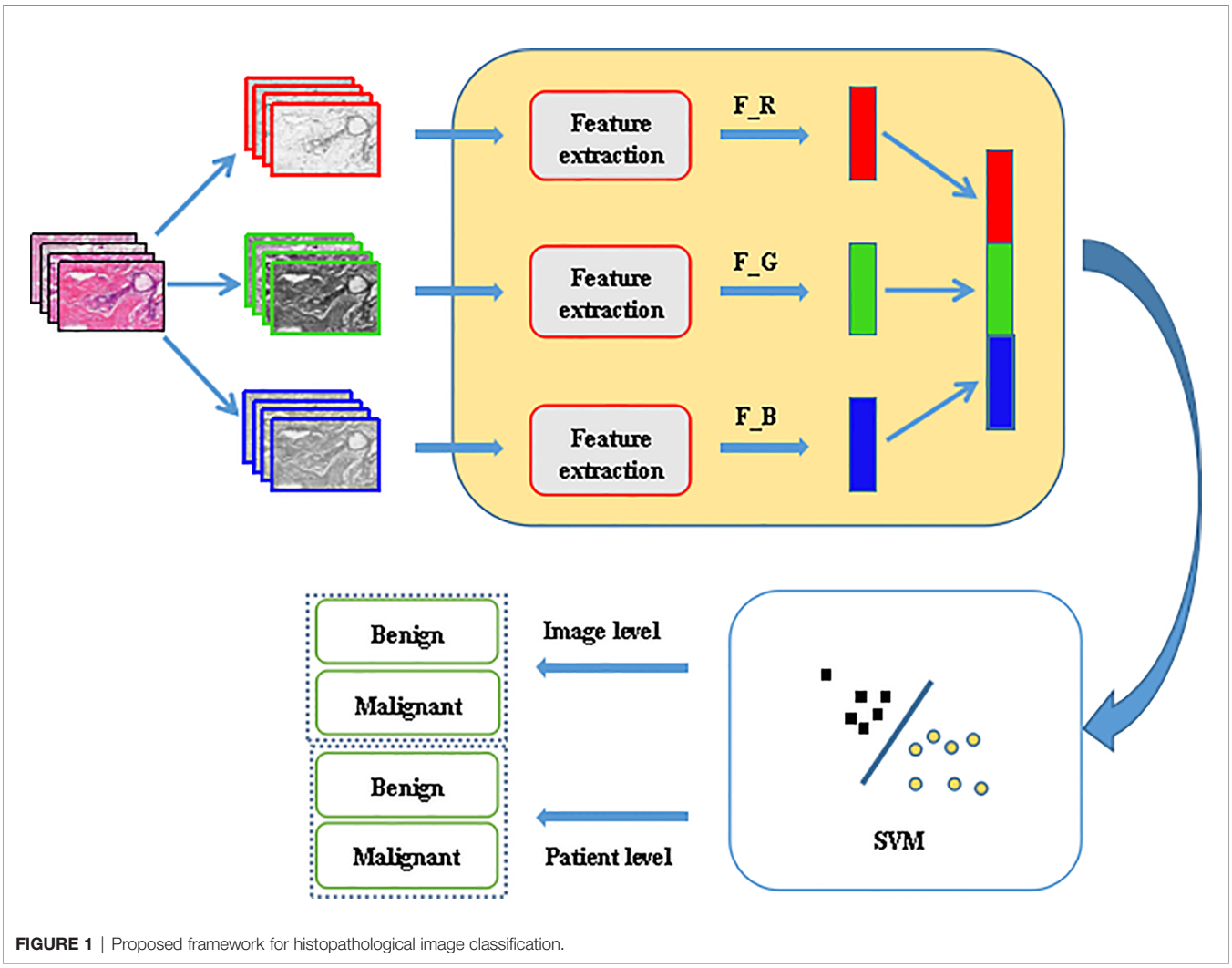


FIGURE 1 | Proposed framework for histopathological image classification.

Average Pixel Value of Each Channel

The average value reflects the centralized tendency of the data and is an important amplitude feature of images. For an image, the average pixel value of each color channel is expressed as

$$f_{mean} = \frac{1}{MN}\sum_{x_c=1}^{M}\sum_{y_c=1}^{N} f(x_c, y_c), \quad (1)$$

where $f(x_c, y_c)$ represents the pixel value of (x_c, y_c).

Hu Invariant Moment

Geometric moments were proposed by Hu.M.K (30) in 1962. They constructed seven invariant moments according to second-order and third-order normalized central moments, and proved that they are invariant to rotation, scaling and translation. Hu invariant moment is a region-based image shape descriptor. In the construction of Hu invariant moments, the central moment is used to eliminate the influence of image translation, the normalization eliminates the influence of image scaling, and the polynomial is constructed to realize the invariant characteristics of rotation. Different order moments reflect different characteristics, the low order reflects the basic shape of the target, and the high order reflects the details and complexity.

Wavelet Features

The result of two-dimensional wavelet decomposition reflects the frequency changes in different directions and the texture characteristics of the image. Since the detail subgraph is the high-frequency component of the original image and contains the main texture information, the energy of the individual detail subgraph is taken as the texture feature, which reflects the energy distribution along the frequency axis with respect to the scale and direction. In this paper, 5-layer wavelet decomposition was carried out, and the energy of high-frequency components in each layer was taken as the feature vector.

Tamura

Tamura et al. (31) proposed a texture feature description method based on the psychological research of texture visual perception, and defined six characteristics to describe texture. Namely, coarseness, contrast, directionality, line likeness, regu larity, and roughness. Coarseness reflects the change intensity of image gray level. The larger the texture granularity is, the coarser the texture image is. Contrast reflects the lightest and darkest gray levels in a gray image, and the range of differences determines the contrast. Directionality reflects the intensity of image texture concentration along a certain direction. Lineality reflects whether the image texture has a linear structure. Regulation reflects the consistency of texture features between a local region and the whole image. Roughness is the sum of roughness and contrast.

Local Binary Pattern

Local Binary Pattern (32) is an operator used to describe local texture features of an image. It has significant advantages such as rotation invariance and gray level invariance. The original LBP operator is defined as comparing the gray values of eight adjacent pixels with the threshold value namely the center pixel in a 3×3 window. If the value of the adjacent pixel is greater than or equal to the value of the center pixel, the position of the pixel is marked as 1, otherwise it is 0. That is, for a pixel (x_c, y_c) on the image

$$LBP_{P,R} \quad (x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \quad (2)$$

Where P is the number of sampling points in the neighborhood of the center pixel, R is the radius of the neighborhood, g_c is the gray value of the center pixel; g_p is the gray value of the pixel adjacent to the center pixel.

In this way, 8 points in the neighborhood can be compared to generate a total of 256 8-bit binary numbers, that is, the LBP value of the center pixel of the 3×3 window is obtained, and this value is used to reflect the texture information of the region.

Completed Local Binary Pattern

Completed local binary pattern (33) is a variant of LBP. The local area of the CLBP operator is represented by its center pixel and local differential sign magnitude transformation. After the center pixel is globally thresholded, it is coded with a binary string as CLBP_Center (CLBP_C). At the same time, the local difference sign magnitude transformation is decomposed into two complementary structural components: difference sign CLBP-Sign (CLBP_S) and difference magnitude CLBP-Magnitude (CLBP_M). For a certain pixel (x_c, y_c) on the image, the components are expressed as:

$$\begin{cases} CLBP\_C_{P,R}(x_c, y_c) = s(g_c - g_N) \\ CLBP\_S_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \qquad s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} . \\ CLBP\_M_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(D_p - D_c)2^p \end{cases} \quad (3)$$

Where, N is the number of windows, $g_N = \frac{1}{N}\sum_{n=0}^{N-1} g_n$ represents the mean gray value about g_c when the center point is constantly moving, and $D_p = |g_p - g_c|, D_c = \frac{1}{P}\sum_{p=0}^{P-1}|g_p - g_c|$ represents the mean magnitude. $CLBP\_S_{P,R}(x_c, y_c)$ is equivalent to the traditional LBP operator, which describes the difference sign characteristics of the local window. $CLBP\_M_{P,R}(x_c, y_c)$ describes the difference magnitude characteristics of the local window. $CLBP\_C_{P,R}(x_c, y_c)$ is the gray level information reflected by the pixel at the center. In our experiments, we worked with rotation-invariant uniform patterns, with a standard value of $P = 8$, $R = 1$, yielding a 20-D feature vector for each channel.

Gabor

Gabor feature is a kind of feature that can be used to describe the texture information of image. The frequency and direction of Gabor filter are similar to human visual system, and it is particularly suitable for texture representation and

discrimination. Gabor features mainly rely on Gabor kernel to window the signal in frequency domain, so as to describe the local frequency information of the signal. Different textures generally have different center frequencies and bandwidths. According to these frequencies and bandwidths, a set of Gabor filters can be designed to filter texture images. Each Gabor filter only allows the texture corresponding to its frequency to pass smoothly, while the energy of other textures is suppressed. Texture features are analyzed and extracted from the output results of each filter for subsequent classification tasks. we used the Gabor filters with five scales and eight orientations, the size of the filter bank is 39×39, the block size is 46×70, yielding a 4000-D feature vector for each channel.

Histogram of Oriented Gradient

Histogram of Oriented Gradient (34) is a feature descriptor used for object detection in computer vision and image processing. It constructs features by calculating and counting the histogram of the gradient direction in the local area of the image. The use of gradient information can well reflect the edge information of the target, the local appearance and shape of the image can be characterized by the size of the local gradient. It is generally used in pedestrian detection, face recognition and other fields, but it does not perform well on images with complex texture information. It is introduced as a comparison in this paper.

EXPERIMENTS AND RESULTS

Dataset

The BreaKHis dataset (9) contains biopsy images of benign and malignant breast tumors, which were collected through clinical studies from January 2014 to December 2014. During the period, all patients with clinical symptoms of BC were invited to the Brazilian P&D laboratory to participate in the study. Samples were collected by surgical open biopsy (SOB) and stained with hematoxylin and eosin. Hematoxylin is alkaline, mainly making the chromatin in the nucleus and nucleic acid in the cytoplasm stained blue-purple. eosin is acidic, mainly making the components in the cytoplasm and extracellular matrix stained pink. These images can be used for histological studies and marked by pathologists in the P&D laboratory. The BreaKHis dataset consists of 7909 breast tumor tissue microscopic images of 82 patients, divided into benign and malignant tumors, including 2480 benign (24 patients) and 5429 malignant (58 patients). The image is obtained in a three-channel RGB (red-green-blue) true color space with magnification factors of 40X, 100X, 200X, 400X, and the size of each image is 700×460. **Tables 1** and **2** summarize the image distribution. And **Figure 2** shows the representative examples of BreaKHis dataset.

Protocol

All of the experiments were conducted on a platform with an Intel Core i7-5820K CPU and 16G memory. The BreaKHis dataset has been randomly divided into a training set (70%, 56 patients) and a testing set (30%, 26 patients). We guarantee that patients use to build the training set are not used for the testing set. The results presented in this work are the average of five trials.

All the images we used were without any preprocessing before feature extraction. For the SVM, we chose the RBF kernel. The best penalty factor c=2 and kernel function parameter g=1 were obtained by cross validation. For wavelet function, we selected coif5 wavelet function, which has better symmetry than dbN, has the same support length as db3N and sym3N, and has the same number of vanishing moments as db2N and sym2N.

Here, we report the recognition accuracy at both the image level and the patient level. For the image level, let N_{rec_I} be the

TABLE 1 | Image distribution by magnification factor and class.

Magnification	Benign	Malignant	Total
40X	625	1370	1995
100X	644	1437	2081
200X	623	1390	2013
400X	588	1232	1820
Total	2480	5429	7909
Patients	24	58	82

TABLE 2 | Image distribution by magnification factor and subclass.

Class	Sub-class	Magnification			
		40X	100X	200X	400X
Benign	Adenosis (A)	114	113	111	106
	Fibroadenoma(F)	253	260	264	237
	Phyllodes_tumor(PT)	109	121	108	115
	Tubular_adenoma(TA)	149	150	140	130
Malignant	Ductal_carcinoma(DC)	864	903	896	788
	Lobular_carcinoma(LC)	156	170	163	137
	Mucinous_carcinoma(MC)	205	222	196	169
	Papillary_carcinoma(PC)	145	142	135	138

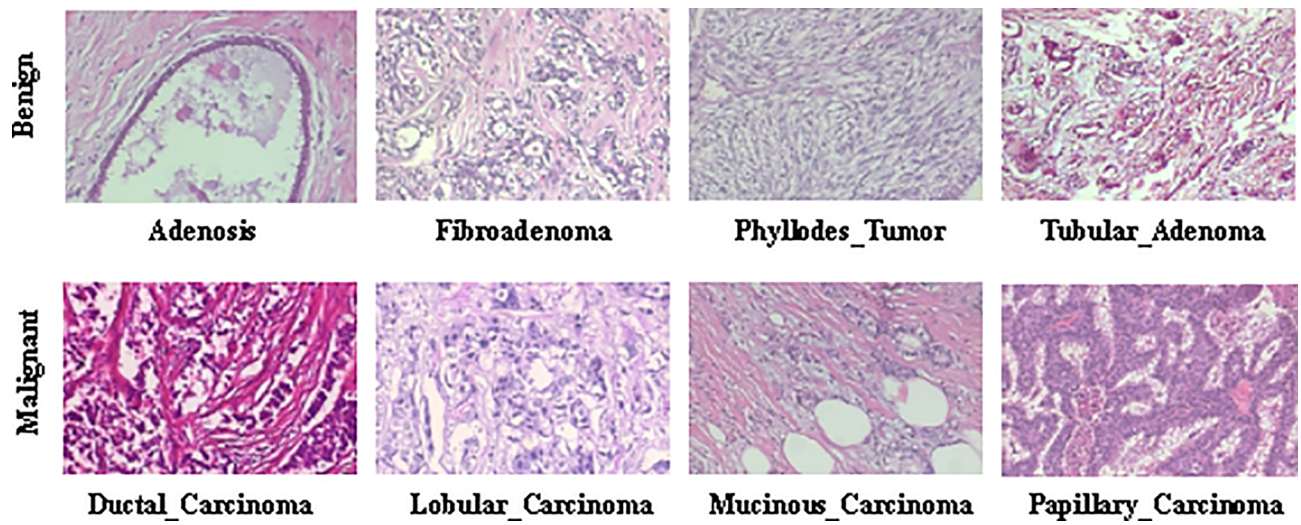


FIGURE 2 | Representative examples of BreaKHis dataset.

number of images correctly classified, N represents all the test samples, then the recognition accuracy of the image level can be defined as

$$Image\_accuracy = \frac{N_{rec\_I}}{N}. \tag{4}$$

For the patient level, we followed the definition of (9). Let N_P be the image of patient P, S is the total number of patients, and N_{rec_P} images of patient P were correctly classified, then the patient score can be defined as

$$Patient \quad score = \frac{N_{rec\_P}}{N_P}, \tag{5}$$

and define the recognition accuracy of the patient level as

$$Patient\_accuracy = \frac{\sum Patient \quad score}{S}. \tag{6}$$

To further assess the performance of the proposed framework, sensitivity (Se), precision (Pr) and F1-score metrics were used and the formulations of the metrics are described as

$$Se = \frac{TP}{TP + FN}, \tag{7}$$

$$Pr = \frac{TP}{TP + FP}, \tag{8}$$

$$F1 - score = \frac{2 \times TP}{2 \times TP + FP + FN}, \tag{9}$$

where true positive (TP) represents the number of malignant samples classified as malignant, whereas true negative (TN) represents the number of benign samples classified as benign. Also, false positive (FP) represents the number of benign samples incorrectly classified as malignant while false negative (FN) represents the number of malignant samples misclassified as benign.

Experiment Results

Table 3 reports the performance of all descriptors we have assessed. The image level recognition accuracy, the patient level recognition accuracy, sensitivity, precision and F1-score of 10 different three-channel descriptors under 4 magnifications were compared. The descriptors are GLCM1, GLCM4, APVEC, HIM, wavelet feature, Tamura, CLBP. In order to show the effectiveness of low dimensional features, LBP, Gabor, and Hog were introduced for comparison.

For images at 40X magnification, GLCM1 achieved the highest recognition accuracy of 94.12 ± 2.19% at the image level and 93.48 ± 2.7% at the patient level, as well as the highest precision and F1_score. The second was GLCM4 with which the image_accuracy and the patient_accuracy were 93.4 ± 3.54% and 92.95 ± 4.02, respectively. Followed by APVEC achieving the image_accuracy of 92.12 ± 1.09%, and the patient_accuracy of 90.55 ± 0.84%. The same conclusion was drawn for 100X. The image level recognition accuracy and the patient level recognition accuracy of GLCM1, GLCM4, and APVEC were 92.65 ± 3.08%, 91.74 ± 3.89%, 91.98 ± 3.79%, 91.16 ± 3.88%, 90.2 ± 2.33%, 89.18 ± 3.45%, respectively. However, for 200X, APVEC achieved the highest image level recognition accuracy of 94.97 ± 1.35%, followed by GLCM1 and GLCM4. GLCM1 performed best at the patient level with an accuracy of 94.24 ± 2.86%, which is 0.3% higher than APVEC. As for 400X, APVEC performed best at both the image level (92.78 ± 3.14%) and the patient level (93.3 ± 3.25%) followed by GLCM1 and GLCM4. On the whole, GLCM1, GLCM4 and APVEC performed well at both the image level and the patient level, followed by HIM. The four descriptors all get the highest recognition accuracy at 200X, and all descriptors except Gabor and

TABLE 3 | Classification performance of different descriptors based on three-channel features.

Features	Magnification	Image_accuracy (%)	Patient_accuracy (%)	Sensitivity (%)	Precision (%)	F1-score (%)
GLCM1	40X	94.12 ± 2.19	93.48 ± 2.70	94.99 ± 1.53	**97.01 ± 2.31**	95.98 ± 1.73
	100X	92.65 ± 3.08	91.74 ± 3.89	95.61 ± 4.81	94.53 ± 1.56	95.01 ± 2.47
	200X	94.67 ± 2.02	**94.24 ± 2.86**	97.17 ± 3.49	95.81 ± 2.01	96.44 ± 1.54
	400X	90.98 ± 2.17	91.43 ± 2.19	91.80 ± 4.09	95.76 ± 3.01	93.66 ± 1.81
GLCM4	40X	93.42 ± 3.54	92.95 ± 4.02	94.72 ± 3.94	96.26 ± 2.22	95.46 ± 2.83
	100X	91.98 ± 3.79	91.16 ± 3.88	96.57 ± 3.62	92.85 ± 2.76	94.65 ± 2.86
	200X	93.53 ± 3.50	92.81 ± 4.72	96.14 ± 3.98	95.36 ± 2.75	95.70 ± 2.37
	400X	90.65 ± 2.93	90.76 ± 3.28	93.54 ± 3.49	93.72 ± 3.76	93.56 ± 2.20
APVEC	40X	92.12 ± 1.09	90.55 ± 0.84	94.67 ± 2.04	94.83 ± 1.31	94.72 ± 0.66
	100X	90.20 ± 2.33	89.18 ± 3.45	94.05 ± 2.36	93.16 ± 4.15	93.52 ± 1.40
	200X	**94.97 ± 1.35**	94.21 ± 2.37	**98.30 ± 0.73**	95.21 ± 1.79	**96.72 ± 0.84**
	400X	92.78 ± 3.14	93.30 ± 3.25	94.07 ± 4.91	96.04 ± 2.54	94.96 ± 2.35
HIM	40X	89.21 ± 1.59	86.44 ± 3.46	93.96 ± 2.36	91.81 ± 2.46	92.85 ± 2.31
	100X	88.99 ± 2.45	87.67 ± 3.14	94.17 ± 3.38	91.62 ± 3.84	92.78 ± 1.48
	200X	92.93 ± 1.89	92.19 ± 2.87	95.06 ± 2.28	95.59 ± 1.60	95.30 ± 1.20
	400X	88.64 ± 3.97	88.61 ± 4.94	91.77 ± 6.21	92.73 ± 2.11	92.14 ± 3.03
Wavelet	40X	80.98 ± 4.23	80.03 ± 7.16	97.66 ± 4.23	80.78 ± 8.00	88.38 ± 4.81
	100X	80.36 ± 3.66	80.24 ± 1.03	97.02 ± 4.62	80.82 ± 4.39	88.04 ± 2.53
	200X	78.99 ± 4.47	76.50 ± 3.33	97.37 ± 3.63	79.46 ± 4.22	87.43 ± 2.86
	400X	76.08 ± 2.22	76.79 ± 1.45	89.65 ± 4.31	80.27 ± 3.99	84.56 ± 1.74
Tamura	40X	78.91 ± 3.30	78.62 ± 1.27	97.23 ± 3.93	79.30 ± 6.16	87.31 ± 2.17
	100X	78.68 ± 4.03	78.09 ± 1.17	99.27 ± 0.43	78.18 ± 4.08	87.43 ± 2.66
	200X	77.37 ± 1.89	76.00 ± 2.09	94.43 ± 3.02	79.41 ± 2.38	86.23 ± 1.59
	400X	75.88 ± 2.86	75.66 ± 1.72	94.04 ± 2.18	77.68 ± 2.64	85.06 ± 2.18
LBP	40X	84.38 ± 2.32	86.51 ± 2.43	93.23 ± 2.92	87.07 ± 3.48	89.87 ± 2.37
	100X	83.91 ± 4.84	85.20 ± 3.78	95.95 ± 2.63	84.66 ± 5.15	89.89 ± 3.33
	200X	83.26 ± 4.04	84.05 ± 3.27	92.24 ± 4.23	86.39 ± 4.03	89.15 ± 3.13
	400X	82.35 ± 5.56	82.76 ± 4.84	91.64 ± 4.12	85.57 ± 6.45	88.35 ± 3.81
CLBP	40X	82.63 ± 3.54	83.18 ± 3.68	93.29 ± 3.77	85.03 ± 6.84	88.89 ± 2.95
	100X	82.64 ± 4.69	84.31 ± 3.66	95.46 ± 3.93	83.83 ± 5.31	89.14 ± 3.10
	200X	78.72 ± 2.61	78.20 ± 2.36	95.65 ± 4.81	80.13 ± 3.58	87.08 ± 1.91
	400X	75.26 ± 4.01	75.97 ± 2.28	94.39 ± 5.78	77.40 ± 5.36	84.81 ± 2.42
Gabor	40X	86.11 ± 4.46	84.87 ± 4.85	97.00 ± 1.34	86.45 ± 6.92	91.29 ± 3.20
	100X	89.98 ± 2.15	89.79 ± 2.74	93.45 ± 2.87	93.37 ± 3.78	93.33 ± 1.39
	200X	91.04 ± 2.66	89.65 ± 3.97	97.14 ± 1.86	91.58 ± 3.78	94.23 ± 1.72
	400X	88.94 ± 2.87	87.84 ± 2.69	96.96 ± 1.74	88.94 ± 3.07	92.75 ± 1.99
Hog	40X	76.59 ± 4.42	76.82 ± 7.21	95.42 ± 2.97	78.68 ± 10.21	85.94 ± 4.74
	100X	76.06 ± 4.00	75.54 ± 2.64	95.13 ± 5.36	78.21 ± 5.80	85.59 ± 2.50
	200X	76.83 ± 3.35	76.28 ± 2.18	93.32 ± 3.38	79.60 ± 4.38	85.81 ± 2.30
	400X	77.88 ± 3.23	78.20 ± 1.72	89.69 ± 4.25	82.28 ± 5.96	85.59 ± 1.94

The bold values indicate the best value of each metric.

Hog obtain the worst performance at 400X, which is same as the conclusion of (18, 35). Although the recognition accuracy of LBP and Gabor is above 82%, which is also acceptable, it also needs more recognition time due to the high feature dimension, as shown in **Table 4**. Tamura and Hog performed slightly worse compared to other descriptors.

The reason for the above results is that the distributions of features extracted by different feature descriptors are different. The high dispersion of feature distribution will increase the difficulty of image recognition, and the feature with more concentrated distribution will achieve better recognition performance. **Figure 3** is the best illustration of the results.

TABLE 4 | Running time for feature extraction of each image and classification of different descriptors.

Methods	Feature dimensions	Running time for feature extraction of each image (s)	Running time for classification (s)
GLCM1	22×3-D	0.26	0.30
GLCM4	88×3-D	0.05	1.34
APVEC	1×3-D	10.29	0.05
HIM	7×3-D	0.19	0.12
Wavelet	5×3-D	0.41	0.10
Tamura	6×3-D	59.17	0.55
CLBP	20×3-D	0.61	0.19
LBP	256×3-D	0.47	3.09
Gabor	4000×3-D	12.48	66.80
Hog	288×3-D	0.21	4.10

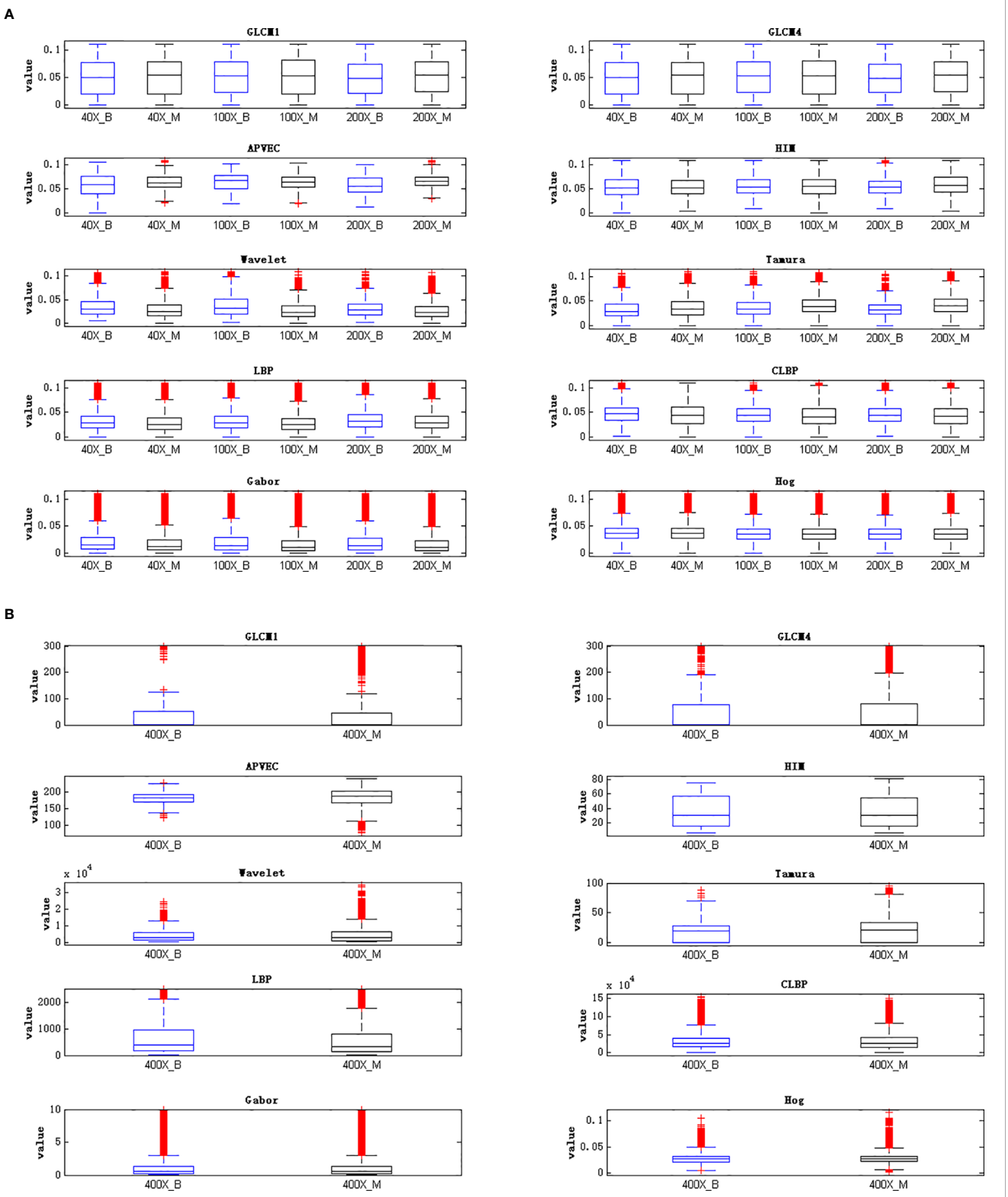


FIGURE 3 | Visualization of feature distribution. **(A)** Feature distribution of 40X, 100X, 200X, **(B)** feature distribution of 400X.

Figure 3 is the visualization of feature distribution. The ordinate represents the feature values. Since the feature values of 40X, 100X, and 200X are relatively small, while the feature values of 400X are relatively large, the feature distribution cannot be displayed in the same figure at the same time. Here are two figures showing the data distribution, **Figure 3**(A) shows the feature distribution of 40X, 100X, 200X, and **Figure 3**(B) shows the feature distribution of 400X. It can be seen from **Figure 3** that for 40X, 100X, 200X, the outliers of GLCM1, GLCM4, APVEC, and HIM are much less than other feature descriptors, indicating that the distributions of these four features are relatively concentrated, which is beneficial for breast cancer identification. In addition, comparing the feature distributions of benign and malignant samples under different magnifications, it can be found that the data distribution of benign and malignant samples of Hog are very similar, indicating the weak ability to discriminate between benign and malignant, which is also the reason for its poor performance. The outliers of GLCM1 and GLCM4 under 400X are obviously more compared to other magnifications, and the similarity of the benign and malignant feature distributions of all descriptors is relatively high, resulting in the poor performance of 400X.

Compared with RGB images, grayscale images only retain the brightness information of the images, but lose the chroma and saturation information of the images. Three-channel features can make up for the lost information of single-channel features, increasing the recognition capability of features, so as to achieve better recognition performance. To further illustrate the advantages of three-channel features, **Table 5** shows the performance of different descriptors of gray-level features.

Comparing **Table 3** and **Table 5**, it can be seen that the performance of the three-channel features is much better than that of gray-level features, especially GLCM1, GLCM4, APVEC, HIM and Gabor. The accuracy for most of them has increased by more than 10% for both the image level and the patient level. **Figure 4** shows the average recognition accuracy of three-

TABLE 5 | Classification performance of different gray-level features.

Features	Magnification	Image_accuracy (%)	Patient_accuracy (%)	Sensitivity (%)	Precision (%)	F1-score (%)
GLCM1	40X	82.88 ± 5.27	83.48 ± 3.65	97.39 ± 4.75	83.08 ± 6.41	89.47 ± 2.98
	100X	83.71 ± 4.66	85.07 ± 2.69	99.45 ± 0.69	82.53 ± 4.93	90.12 ± 2.87
	200X	77.56 ± 4.64	76.44 ± 4.02	99.31 ± 0.83	77.44 ± 4.35	86.95 ± 2.54
	400X	78.51 ± 6.32	80.02 ± 4.46	96.45 ± 2.45	79.42 ± 5.33	86.86 ± 3.22
GLCM4	40X	82.06 ± 4.14	82.71 ± 4.75	98.61 ± 0.94	81.48 ± 3.75	89.14 ± 2.20
	100X	82.78 ± 4.69	83.58 ± 3.50	99.66 ± 1.17	81.56 ± 4.02	89.64 ± 2.42
	200X	77.38 ± 5.02	76.35 ± 3.46	99.18 ± 1.12	77.38 ± 3.95	86.85 ± 2.30
	400X	80.93 ± 6.43	82.37 ± 6.44	93.09 ± 3.70	83.28 ± 4.88	87.75 ± 3.46
APVEC	40X	74.83 ± 3.31	73.06 ± 4.12	99.86 ± 0.30	74.88 ± 5.58	85.56 ± 3.51
	100X	75.41 ± 3.78	73.73 ± 3.22	99.5 ± 0.66	75.51 ± 5.37	85.81 ± 3.33
	200X	75.39 ± 3.11	73.27 ± 0.49	99.86 ± 0.98	75.42 ± 3.51	85.91 ± 2.04
	400X	75.17 ± 3.62	75.58 ± 2.52	99.16 ± 1.43	75.06 ± 4.69	85.40 ± 2.60
HIM	40X	74.89 ± 3.27	73.17 ± 2.05	100.00 ± 1.83	74.88 ± 5.01	85.60 ± 2.54
	100X	76.06 ± 4.60	74.70 ± 2.91	99.06 ± 1.19	76.31 ± 5.51	86.11 ± 3.10
	200X	75.36 ± 3.08	73.27 ± 0.33	99.54 ± 0.94	75.51 ± 2.95	85.85 ± 2.03
	400X	73.97 ± 3.31	74.03 ± 1.41	99.48 ± 1.63	74.00 ± 4.42	84.83 ± 2.36
Wavelet	40X	81.47 ± 4.95	81.38 ± 1.56	99.41 ± 0.74	80.50 ± 3.77	88.90 ± 2.58
	100X	80.54 ± 3.78	80.51 ± 1.16	99.80 ± 0.16	79.50 ± 3.66	88.45 ± 2.27
	200X	77.50 ± 4.04	76.55 ± 0.73	99.09 ± 2.14	77.41 ± 4.10	86.87 ± 1.90
	400X	74.82 ± 3.87	74.77 ± 2.90	98.15 ± 1.54	75.15 ± 5.11	85.08 ± 2.90
Tamura	40X	78.55 ± 4.12	77.83 ± 1.05	99.08 ± 1.75	78.16 ± 3.94	87.33 ± 2.17
	100X	79.45 ± 4.07	79.16 ± 1.61	99.27 ± 1.28	78.85 ± 4.40	87.83 ± 2.41
	200X	76.54 ± 2.14	75.33 ± 2.56	98.60 ± 1.59	76.81 ± 2.50	86.33 ± 1.53
	400X	73.30 ± 2.98	73.08 ± 2.16	100.00 ± 0.64	73.30 ± 4.24	84.57 ± 2.58
LBP	40X	84.97 ± 3.54	87.04 ± 3.96	94.61 ± 4.40	86.51 ± 1.54	90.31 ± 2.79
	100X	84.31 ± 4.98	86.05 ± 3.24	97.31 ± 1.54	84.31 ± 4.31	90.25 ± 3.03
	200X	83.12 ± 4.99	84.17 ± 5.72	92.48 ± 4.47	86.04 ± 2.53	89.09 ± 3.39
	400X	80.05 ± 5.54	80.72 ± 4.94	91.65 ± 4.58	82.93 ± 4.95	86.99 ± 3.97
CLBP	40X	83.19 ± 5.15	84.63 ± 3.17	96.40 ± 1.73	83.89 ± 6.26	89.57 ± 3.07
	100X	84.06 ± 4.80	86.12 ± 2.23	97.77 ± 2.25	83.89 ± 6.08	90.17 ± 3.01
	200X	78.56 ± 3.55	78.58 ± 3.30	96.38 ± 3.92	79.75 ± 8.05	87.12 ± 3.13
	400X	74.79 ± 3.80	74.94 ± 4.33	97.74 ± 4.66	75.19 ± 8.13	84.98 ± 3.59
Gabor	40X	80.06 ± 4.64	79.84 ± 2.61	98.03 ± 2.59	79.96 ± 5.82	88.01 ± 2.55
	100X	81.25 ± 4.89	81.57 ± 2.57	98.91 ± 1.94	80.6 ± 5.84	88.75 ± 3.01
	200X	76.94 ± 2.88	75.43 ± 1.90	98.17 ± 1.71	77.36 ± 4.68	86.49 ± 2.23
	400X	77.40 ± 4.68	78.70 ± 3.54	92.37 ± 2.37	80.19 ± 7.64	85.71 ± 3.28
Hog	40X	76.46 ± 4.03	76.30 ± 2.29	97.50 ± 1.98	77.18 ± 4.00	86.09 ± 2.31
	100X	75.80 ± 3.51	74.31 ± 1.73	97.63 ± 2.91	76.56 ± 3.87	85.76 ± 2.27
	200X	76.22 ± 2.59	74.74 ± 1.60	97.78 ± 1.86	76.85 ± 2.43	86.05 ± 1.81
	400X	76.02 ± 3.23	75.84 ± 1.84	91.15 ± 4.68	79.30 ± 4.34	84.70 ± 2.76

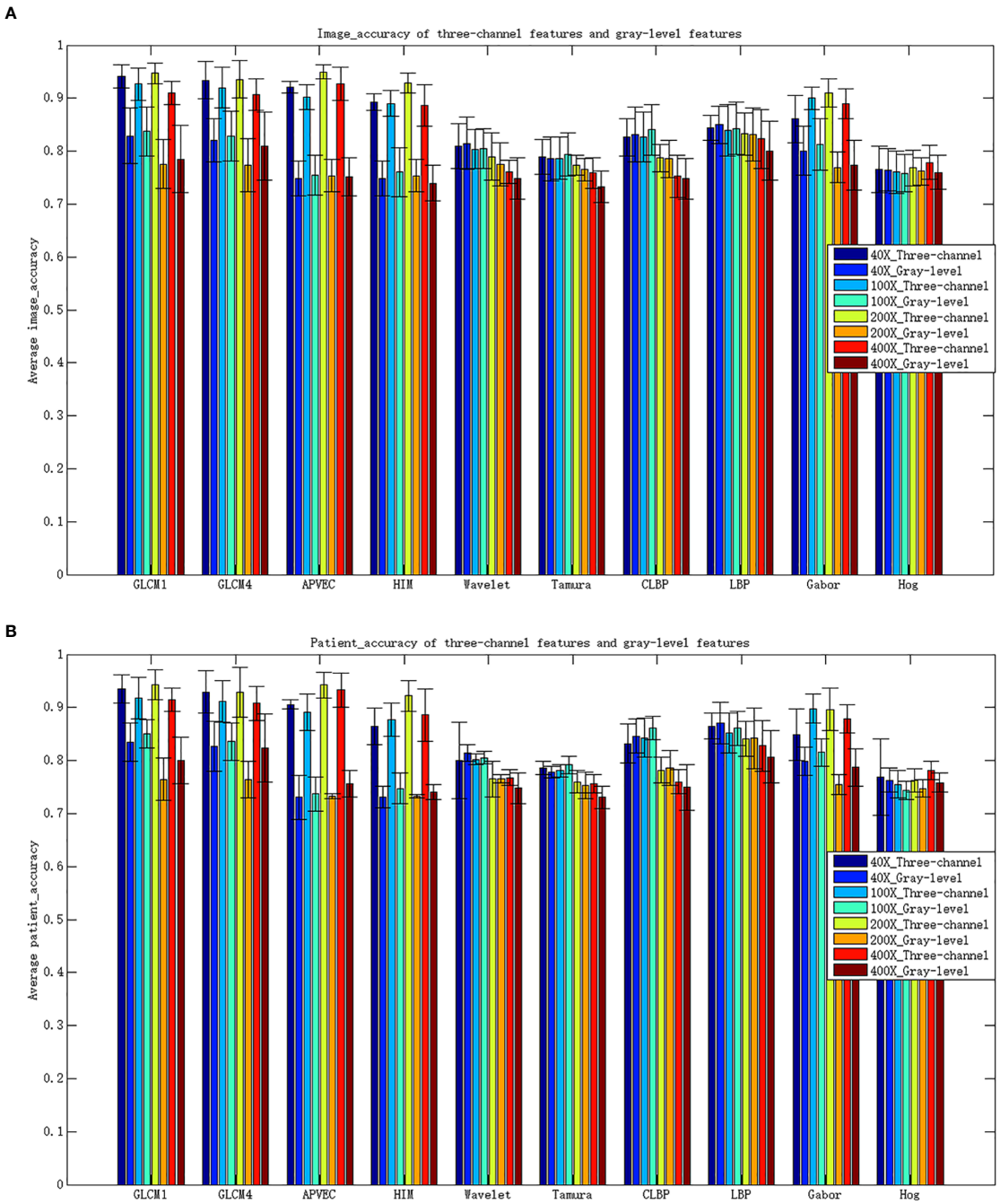


FIGURE 4 | Classification accuracy for different features. **(A)** Image_accuracy for three-channel features and gray-level features, **(B)** patient_accuracy for three-channel features and gray-level feature\s.

channel features and gray-level features for the image level and the patient level. The advantages of the three-channel features can be seen more clearly from **Figure 4**.

Although the advantages of the three-channel features are obvious, we still have no idea about which channel plays a more important role in the classification results. **Table 6** shows the classification performance of single-channel features under different magnifications. Observing the experimental results, we can find that R channel have a greater impact on the classification results under 40X, 100X, 200X magnifications, while B channel performs better under 400X. This is consistent with the actual situation of H&E histopathological images under different magnifications. The images of 40X, 100X, and 200X have more cytoplasm and appear pink. The image of 400X

TABLE 6 | Classification performance of single-channel features under different magnifications.

Features	40X					
	R		G		B	
	Image_accuracy (%)	Patient_accuracy (%)	Image_accuracy (%)	Patient_accuracy (%)	Image_accuracy (%)	Patient_accuracy (%)
GLCM1	89.03 ± 2.28	87.46 ± 2.23	82.25 ± 4.51	82.03 ± 2.30	83.62 ± 3.07	82.84 ± 2.12
GLCM4	88.47 ± 1.99	86.06 ± 1.58	81.53 ± 4.12	82.05 ± 2.68	83.58 ± 3.62	82.00 ± 3.23
APVEC	83.43 ± 1.51	79.73 ± 1.30	77.00 ± 3.95	77.39 ± 2.60	74.97 ± 3.32	73.18 ± 0.23
HIM	81.08 ± 1.03	77.26 ± 0.86	77.30 ± 4.28	77.81 ± 2.54	74.83 ± 3.31	73.08 ± 0.00
Wavelet	80.74 ± 4.69	80.24 ± 2.28	81.35 ± 5.17	81.21 ± 3.16	81.28 ± 5.12	80.97 ± 2.89
Tamura	76.90 ± 3.50	75.57 ± 0.64	79.32 ± 4.22	78.84 ± 1.85	77.17 ± 3.06	75.97 ± 0.77
LBP	84.70 ± 3.75	86.65 ± 3.63	84.11 ± 3.52	85.53 ± 2.93	84.64 ± 3.01	86.35 ± 2.02
CLBP	82.64 ± 4.17	83.29 ± 2.78	83.79 ± 3.75	84.90 ± 2.81	83.36 ± 4.40	84.18 ± 3.11
Gabor	83.46 ± 4.47	82.09 ± 3.56	78.63 ± 3.25	78.31 ± 3.32	80.60 ± 4.86	80.12 ± 2.49
Hog	76.59 ± 4.22	76.09 ± 3.13	76.38 ± 3.95	75.77 ± 1.91	76.07 ± 3.79	75.33 ± 2.18
Features	**100X**					
	R		G		B	
	Image_accuracy(%)	Patient_accuracy(%)	Image_accuracy(%)	Patient_accuracy(%)	Image_accuracy(%)	Patient_accuracy(%)
GLCM1	90.32 ± 1.59	88.52 ± 1.87	84.00 ± 4.99	84.95 ± 4.6	87.43 ± 1.10	85.00 ± 1.05
GLCM4	88.94 ± 1.46	86.83 ± 2.17	83.17 ± 4.99	84.62 ± 3.67	87.14 ± 1.54	85.58 ± 1.03
APVEC	77.70 ± 4.37	74.96 ± 2.26	78.29 ± 5.93	78.91 ± 4.77	75.00 ± 3.74	73.08 ± 0.00
HIM	78.83 ± 2.58	75.92 ± 1.32	78.32 ± 4.74	77.11 ± 3.99	76.17 ± 2.90	73.70 ± 0.44
Wavelet	79.94 ± 3.77	79.73 ± 1.38	81.00 ± 3.84	81.08 ± 1.01	79.46 ± 4.64	79.13 ± 2.12
Tamura	78.01 ± 3.83	77.31 ± 1.67	79.91 ± 4.28	79.69 ± 1.72	78.83 ± 4.26	78.41 ± 1.99
LBP	84.48 ± 4.98	85.68 ± 3.83	84.61 ± 5.02	86.51 ± 3.21	83.25 ± 4.49	84.54 ± 2.32
CLBP	81.75 ± 4.38	82.78 ± 3.12	81.41 ± 4.13	83.49 ± 1.73	81.83 ± 3.58	82.66 ± 1.55
Gabor	88.42 ± 2.99	87.27 ± 2.69	82.25 ± 5.07	83.63 ± 3.32	83.45 ± 3.64	83.31 ± 1.63
Hog	76.04 ± 3.86	74.73 ± 1.97	75.76 ± 4.05	74.40 ± 0.65	75.67 ± 3.93	74.34 ± 1.15
Features	**200X**					
	R		G		B	
	Image_accuracy(%)	Patient_accuracy(%)	Image_accuracy(%)	Patient_accuracy(%)	Image_accuracy(%)	Patient_accuracy(%)
GLCM1	91.31 ± 3.14	90.11 ± 3.82	79.05 ± 3.82	77.86 ± 2.81	88.62 ± 3.76	86.74 ± 5.37
GLCM4	90.45 ± 2.90	89.35 ± 3.65	78.18 ± 3.86	78.42 ± 3.27	88.77 ± 3.30	86.91 ± 4.72
APVEC	83.95 ± 0.97	81.71 ± 2.93	79.40 ± 5.22	79.00 ± 4.83	76.46 ± 1.57	73.89 ± 1.12
HIM	83.89 ± 1.56	81.72 ± 2.61	78.46 ± 5.16	77.63 ± 4.28	79.09 ± 2.46	77.02 ± 1.37
Wavelet	77.67 ± 4.12	76.68 ± 2.27	77.69 ± 3.76	76.58 ± 2.48	76.59 ± 3.99	75.16 ± 2.18
Tamura	75.98 ± 2.10	74.17 ± 2.44	76.48 ± 1.77	74.84 ± 2.52	76.01 ± 2.08	74.05 ± 2.18
LBP	83.21 ± 4.75	84.35 ± 3.95	82.66 ± 4.97	83.36 ± 4.19	82.48 ± 4.07	82.87 ± 3.39
CLBP	79.21 ± 3.43	79.24 ± 2.83	77.08 ± 3.15	75.97 ± 2.15	77.59 ± 3.37	77.19 ± 1.29
Gabor	88.38 ± 1.91	86.82 ± 2.72	78.46 ± 4.83	77.66 ± 3.66	84.03 ± 1.9	80.46 ± 2.54
Hog	75.75 ± 2.97	74.3 ± 1.25	76.39 ± 2.55	74.99 ± 1.12	75.99 ± 2.91	74.16 ± 1.06
Features	**400X**					
	R		G		B	
	Image_accuracy(%)	Patient_accuracy(%)	Image_accuracy(%)	Patient_accuracy(%)	Image_accuracy(%)	Patient_accuracy(%)
GLCM1	83.48 ± 1.11	81.67 ± 1.35	80.22 ± 5.84	81.96 ± 5.05	83.09 ± 2.20	82.28 ± 3.23
GLCM4	83.14 ± 1.11	81.24 ± 1.55	82.25 ± 5.36	83.43 ± 5.00	84.51 ± 2.73	83.75 ± 3.71
APVEC	73.34 ± 2.99	73.10 ± 0.05	79.31 ± 6.57	80.75 ± 5.89	73.37 ± 2.85	73.03 ± 0.11
HIM	74.27 ± 3.01	74.07 ± 1.10	78.08 ± 5.87	79.33 ± 5.53	76.01 ± 1.36	74.51 ± 2.39
Wavelet	75.47 ± 3.87	75.80 ± 2.41	75.18 ± 3.90	75.20 ± 2.50	75.54 ± 4.67	75.83 ± 3.40
Tamura	73.34 ± 2.99	73.10 ± 0.05	73.47 ± 2.76	73.25 ± 0.29	75.37 ± 4.01	75.92 ± 2.36
LBP	80.81 ± 5.01	81.87 ± 3.61	80.78 ± 5.77	82.31 ± 3.97	82.3 ± 6.32	83.35 ± 5.19
CLBP	74.88 ± 3.31	75.00 ± 1.70	74.07 ± 3.55	74.40 ± 1.21	74.63 ± 3.50	75.11 ± 1.60
Gabor	78.71 ± 3.73	76.82 ± 4.35	80.12 ± 5.34	81.46 ± 4.02	79.27 ± 2.88	77.86 ± 2.79
Hog	75.41 ± 3.02	74.66 ± 0.76	76.75 ± 3.68	76.91 ± 1.54	77.03 ± 3.31	76.62 ± 1.83

contains more information about the precise lesion locations, which is usually presented through the nucleus, and appear blue-purple.

Different descriptors extract different features. It often cannot obtain all the effective information of the image only by one method. There may be a complementary relationship between different methods, and sometimes more redundant information may be added. In this paper, GLCM1 with the best recognition performance is combined with 8 other methods except GLCM4. Different features are fused in a cascade way. The results are shown in **Table 7**.

Table 7 shows that after the combination of GLCM1 and APVEC, the recognition accuracy of 40X and 100X is better than a single method whether it is for the image level or the patient level, and the accuracy of 200X and 400X is slightly lower than that of APVEC. The combination of GLCM1 and HIM improves

TABLE 7 | Classification performance of GLCM1 combined with other descriptors.

Features	Magnification	Image_accuracy (%)	Patient_accuracy (%)	Sensitivity (%)	Precision (%)	F1-score (%)
GLCM1+APVEC	40X	**94.22 ± 2.18**	**93.70 ± 2.68**	95.77 ± 2.19	96.39 ± 1.57	96.07 ± 1.77
	100X	**92.78 ± 2.88**	**91.80 ± 3.67**	95.47 ± 4.45	94.83 ± 1.84	95.09 ± 2.29
	200X	94.77 ± 2.18	94.33 ± 2.99	97.23 ± 3.64	95.89 ± 2.13	96.50 ± 1.65
	400X	91.20 ± 2.17	91.59 ± 1.94	91.82 ± 3.90	96.03 ± 2.9	93.80 ± 1.80
GLCM1+ HIM	40X	**94.13 ± 1.85**	93.15 ± 2.52	95.26 ± 2.24	96.75 ± 1.33	95.99 ± 1.56
	100X	**92.69 ± 2.71**	91.17 ± 3.58	95.46 ± 4.74	94.75 ± 1.71	95.03 ± 2.22
	200X	**94.88 ± 2.39**	**94.30 ± 3.17**	96.98 ± 4.00	96.29 ± 2.47	96.56 ± 1.76
	400X	**91.19 ± 2.73**	**91.54 ± 2.26**	91.55 ± 4.75	96.26 ± 2.44	93.76 ± 2.27
GLCM1+Wavelet	40X	93.24 ± 3.45	92.85 ± 3.75	94.67 ± 3.7	96.06 ± 2.14	95.35 ± 2.61
	100X	92.37 ± 3.47	91.38 ± 4.29	95.03 ± 4.67	94.66 ± 1.98	94.80 ± 2.73
	200X	94.25 ± 1.64	93.47 ± 2.39	96.35 ± 3.45	96.06 ± 2.33	96.15 ± 1.27
	400X	90.62 ± 1.45	90.35 ± 1.52	91.49 ± 3.33	95.59 ± 3.08	93.42 ± 1.27
GLCM1+Tamura	40X	93.76 ± 2.89	93.35 ± 3.24	94.81 ± 3.14	96.64 ± 1.80	95.71 ± 2.35
	100X	92.28 ± 3.73	91.07 ± 4.44	95.63 ± 5.63	94.04 ± 1.97	94.75 ± 2.95
	200X	94.88 ± 1.89	94.45 ± 2.90	97.36 ± 2.98	95.98 ± 2.71	96.61 ± 1.30
	400X	90.97 ± 1.76	91.26 ± 1.84	92.23 ± 3.63	95.34 ± 2.91	93.69 ± 1.52
GLCM1+LBP	40X	88.02 ± 3.43	88.28 ± 3.79	90.60 ± 3.33	93.33 ± 3.70	91.83 ± 2.82
	100X	89.03 ± 4.85	89.12 ± 4.08	93.05 ± 2.69	92.30 ± 5.04	92.64 ± 3.45
	200X	89.71 ± 3.10	89.26 ± 3.22	92.49 ± 3.25	93.68 ± 2.35	93.06 ± 2.33
	400X	87.18 ± 4.14	86.42 ± 3.79	89.18 ± 3.76	93.08 ± 5.15	90.99 ± 3.15
GLCM1+CLBP	40X	93.86 ± 2.62	93.07 ± 3.55	95.69 ± 4.12	96.03 ± 2.36	95.84 ± 2.14
	100X	91.72 ± 3.07	90.46 ± 3.82	94.06 ± 3.75	94.70 ± 1.62	94.36 ± 2.39
	200X	94.34 ± 1.99	93.51 ± 2.61	97.48 ± 1.60	95.15 ± 2.21	96.29 ± 1.29
	400X	89.66 ± 3.23	88.94 ± 3.64	90.87 ± 5.22	94.83 ± 3.75	92.69 ± 2.66
GLCM1+Gabor	40X	87.63 ± 4.43	87.22 ± 4.8	93.86 ± 2.38	90.47 ± 6.57	91.92 ± 3.15
	100X	90.28 ± 2.04	89.95 ± 2.72	93.44 ± 2.97	93.71 ± 3.22	93.51 ± 1.34
	200X	91.28 ± 2.64	89.81 ± 4.09	97.06 ± 1.88	91.92 ± 3.66	94.37 ± 1.70
	400X	89.02 ± 2.83	87.80 ± 2.54	96.72 ± 1.77	89.22 ± 3.37	92.79 ± 1.96
GLCM1+Hog	40X	85.55 ± 5.78	86.56 ± 5.32	93.02 ± 2.53	88.39 ± 6.82	90.56 ± 4.20
	100X	86.62 ± 5.08	88.34 ± 3.18	94.06 ± 3.22	89.23 ± 7.32	91.35 ± 3.14
	200X	90.47 ± 3.78	90.04 ± 3.61	93.68 ± 4.40	93.55 ± 2.85	93.57 ± 2.92
	400X	88.15 ± 4.05	88.13 ± 3.72	91.45 ± 2.54	92.28 ± 4.56	91.82 ± 3.01

The bold values indicate that the recognition accuracy of combining the two features is better than that of a single feature.

the image level accuracy, while for the patient level, the accuracy of 40X and 100X is slightly lower than GLCM1. This shows the complementary relationship between GLCM1 and APVEC, HIM. The performance of the combination of GLCM1 and other methods is lower than that of single GLCM1, which shows that the fusion of different texture features increases the redundancy of features and reduces the recognizability.

The recognition accuracy of GLCM1, GLCM4, APVEC, and HIM based on the three-channel features is better than many existing studies, particularly, better than the performance of some deep learning models. **Table 8** shows that the method proposed in this paper is superior to many state-of-the-art methods in benign and malignant tumor recognition, both for the image level and the patient level. It is worth mentioning that works (35–43) did not split training and test set according to the protocol of (9), works (44, 45) adopted the existed protocol, and works (46, 47) randomly divided training set (70%) and test set (30%), but they did not mention whether it was the same as the protocol. Although the recognition accuracy of the works (37, 39, 41–43, 46, 47) is significantly higher than that of our method, they all use deep learning model, which requires a large number of labeled training samples and consumes longer training time. In addition, in these works, except (42), they only calculated the image level recognition accuracy. George et al. even only tested their method based on the data of 200X.

CONCLUSION

In this paper, a breast cancer histopathological images recognition method based on low dimensional three-channel features is proposed. There have been many related studies, but in traditional methods, most scholars did not consider the color channel of the image, so that the extracted features lost part of the effective information. This paper compares the performance of 10 different feature descriptors in the recognition of breast cancer histopathological images. We extracted the three-channel features of different descriptors and fused the features of each channel. Then SVM was used to assess their performance. The experimental results show that the recognition accuracy of GLCM1, GLCM4, APVEC can reach more than 90% regardless of the image level or the patient level. And the performance based on three-channel features is much better than that of gray-level features, especially for GLCM1, GLCM4. We also proved that the R channel has a greater impact on the classification results of 40X, 100X, and 200X, while for 400X, it is more dependent on the B channel. In addition, high dimensional features consume more recognition time, this paper dedicates to achieving accurate recognition based on low dimensional features. Experiment results verify that the high dimensional features extracted by LBP, Hog, and Gabor require more recognition time, but the accuracy has not been greatly improved. Our method is based on the existing traditional methods and is easy to implement without

TABLE 8 | Comparison of the proposed methods with other state-of-the-art methods.

Methods	Training/test	Magnification	Image_accuracy (%)	Patient_accuracy (%)
Gupta V et al. (35)	70%/30%	40X	\	87.2 ± 3.74
		100X	\	88.22 ± 3.28
		200X	\	88.89 ± 2.51
		400X	\	85.82 ± 3.81
Das K et al. (36)	80%/20%	40X	\	89.52
		100X	\	89.06
		200X	\	88.84
		400X	\	87.67
Das K et al. (37)	80%/20%	40X	94.82	\
		100X	94.38	\
		200X	94.67	\
		400X	93.49	\
Cascianelli S et al. (38)	25%/75%	40X	\	87.0
		100X	\	85.2
		200X	\	85.0
		400X	\	81.3
Wei B et al. (39)	75%/25%	40X	97.89	97.02
		100X	97.64	97.23
		200X	97.56	97.89
		400X	97.97	97.50
Zhi W et al. (40)	80%/20%	40X	91.28	\
		100X	91.45	\
		200X	88.57	\
		400X	84.58	\
Nahid AA et al. (41)	85%/15%	40X	95.0	\
		100X	96.6	\
		200X	93.5	\
		400X	94.2	\
Han Z et al. (42)	50%/50%	40X	95.8 ± 3.1	97.1 ± 1.5
		100X	96.9 ± 1.9	95.7 ± 2.8
		200X	96.7 ± 2.0	96.5 ± 2.1
		400X	94.9 ± 2.8	95.7 ± 2.2
Boumaraf S et al. (43)	80%/20%	40X	99.25	\
		100X	99.04	\
		200X	99.00	\
		400X	98.08	\
Song Y et al. (44)	70%/30%	40X	87.0 ± 2.6	90.0 ± 3.2
		100X	86.2 ± 3.7	88.9 ± 5.0
		200X	85.2 ± 2.1	86.9 ± 5.2
		400X	82.9 ± 3.7	86.3 ± 7.0
Saxena S et al. (45)	70%/30%	40X	86.41	89.46
		100X	88.92	92.61
		200X	90.05	93.92
		400X	83.16	89.78
Wang P et al. (46)	70%/30%	40X	92.71 ± 0.16	\
		100X	94.52 ± 0.11	\
		200X	94.03 ± 0.25	\
		400X	93.54 ± 0.24	\
George K et al. (47)	70%/30%	40X	\	\
		100X	\	\
		200X	96.66 ± 0.77	\
		400X	\	\

complex image preprocessing. Experimental results and comparison with other methods confirm that our method requires less training time than deep learning methods, which cannot be ignored in practical applications.

In the future work, we will continue to propose more efficient and rapid methods for breast cancer recognition. The target is to realize multi-class recognition of breast cancer based on the research of benign and malignant tumor recognition. In addition to improving the recognition accuracy, we also hope to extract more effective information about cancer, which can help doctors find the lesion faster and reduce the workload on doctors.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/.

AUTHOR CONTRIBUTIONS

Data processing: TX and HH. Methodology: YH, YB, and SQ. Software: YH, SQ, and LZ. Supervision: YB, HH, TX, WZ and GZ. Original draft: YH. Review & editing: YH, SQ, YB, WZ and GZ. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China as National Major Scientific Instruments Development Project (Grant No. 61927807), the National Natural Science Foundation of China (Grant No. 51875535, 61774137), the Key Research and Development Projects of Shanxi Province (Grant No. 201903D121156) and the National Key Research and Development Project (Grant No. 2019YFC0119800).

ACKNOWLEDGMENTS

We especially thank for the support of Shanxi Provincial Key Laboratory for Biomedical Imaging and Big Data and the fund for Shanxi '1331 Project' Key Subject Construction and Innovation Special Zone Project.

REFERENCES

1. Elmore JG, Longton GM, Carney PA, Geller BM, Onega T, Tosteson ANA, et al. Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens. *JAMA* (2015) 313(11):1122–32. doi: 10.1001/jama.2015.1405
2. Xu J, Xiang L, Liu Q, Gilmore H, Wu J, Tang J, et al. Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images. *IEEE Trans Med Imaging* (2016) 35(1):119–30. doi: 10.1109/TMI.2015.2458702
3. Reis S, Gazinska P, Hipwell J, Mertzanidou T, Naidoo K, Williams N, et al. Automated Classification of Breast Cancer Stroma Maturity From Histological Images. *IEEE Trans Biomed Eng* (2019) 64(10):2344–52. doi: 10.1109/TBME.2017.2665602
4. Niazi MKK, Lin Y, Liu F, Ashok A, Bilgin A. Pathological Image Compression for Big Data Image Analysis: Application to Hotspot Detection in Breast Cancer. *Artif Intell Med* (2018) 95:82–7. doi: 10.1016/j.artmed.2018.09.002
5. Anji RV, Soni B, Sudheer RK. Breast Cancer Detection by Leveraging Machine Learning. *ICT Express* (2020) 6(4):320–4. doi: 10.1016/j.icte.2020.04.009
6. Rahman MM, Ghasemi Y, Suley E, Zhou Y, Wang S, Rogers J. Machine Learning Based Computer Aided Diagnosis of Breast Cancer Utilizing Anthropometric and Clinical Features. *IRBM* (2020). doi: 10.1016/j.irbm.2020.05.005
7. Das A, Nair MS, Peter SD. Sparse Representation Over Learned Dictionaries on the Riemannian Manifold for Automated Grading of Nuclear Pleomorphism in Breast Cancer. *IEEE Trans Image Process* (2019) 28:1248–60. doi: 10.1109/TIP.2018.2877337
8. Wang Z, Li M, Wang H, Jiang H, Yao Y, Zhang H. Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion With CNN Deep Features. *IEEE Access* (2019) 7:105146–58. doi: 10.1109/ACCESS.2019.2892795
9. Spanhol FA, Oliveira LS, Petitjean C, Heutte L. A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Trans Biomed Eng* (2016) 63 (7):1455–62. doi: 10.1109/TBME.2015.2496264
10. Pendar A, Behzad H, Alireza ME, Abdolhossein F. Representation Learning-Based Unsupervised Domain Adaptation for Classification of Breast Cancer Histopathology Images. *Biocybern Biomed Eng* (2018) 38:S0208521617304448. doi: 10.1016/j.bbe.2018.04.008
11. Anuranjeeta A, Shukla KK, Tiwari A, Sharma S. Classification of Histopathological Images of Breast Cancerous and Non Cancerous Cells Based on Morphological Features. *Biomed Pharmacol J* (2017) 10(1):353–66. doi: 10.13005/bpj/1116
12. Belsare AD, Mushrif MM, Pangarkar MA, Meshram N. Classification of Breast Cancer Histopathology Images Using Texture Feature Analysis. In: *Tencon 2015-2015 IEEE Region 10 Conference*. Macao, China: IEEE (2016) 1–5. doi: 10.1109/TENCON.2015.7372809
13. Sharma M, Singh R, Bhattacharya M. Classification of Breast Tumors as Benign and Malignant Using Textural Feature Descriptor. In: *2017 IEEE International Conference on Bioinformatics & Biomedicine*. Kansas City, MO, USA: IEEE (2017) 1110–3. doi: 10.1109/BIBM.2017.8217811
14. Carvalho ED, Filho AOC, Silva RRV, Araújo FHD, Diniz JOB, Silva AC. Breast Cancer Diagnosis from Histopathological Images Using Textural Features and CBIR. *Artif Intell Med* (2020) 105:101845. doi: 10.1016/j.artmed.2020.101845
15. Sudharshan PJ, Petitjean C, Spanhol F, Oliveira LE, Heutte L, Honeine P. Multiple Instance Learning for Histopathological Breast Cancer Image Classification. *Expert Syst Appl* (2019) 117:103–11. doi: 10.1016/j.eswa.2018.09.049
16. Fang J, Xu X, Liu H, Sun F. Local Receptive Field Based Extreme Learning Machine with Three Channels for Histopathological Image Classification. *Int J Mach Learn Cybern* (2019) 10(6):1437–47. doi: 10.1007/s13042-018-0825-6
17. Spanhol FA, Oliveira LS, Petitjean C, Heutte L. Breast Cancer Histopathological Image Classification Using Convolutional Neural Networks. In: *International Joint Conference on Neural Networks (Ijcnn 2016)*. IEEE (2016) 2560–67. doi: 10.1109/IJCNN.2016.7727519
18. Kumar A, Singh KS, Saxena S, Lakshmanan K, Sangaiah AK, Chauhan H. Deep Feature Learning for Histopathological Image Classification of Canine Mammary Tumors and Human Breast Cancer. *Inf Sci* (2020) 508:405–21. doi: 10.1016/j.ins.2019.08.072
19. Bardou D, Zhang K, Ahmad SM. Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks. *IEEE Access* (2018) 6:24680–93. doi: 10.1109/ACCESS.2018.2831280
20. Alom MZ, Yakopcic C, Nasrin MS, Taha TM, Asari VK. Breast Cancer Classification From Histopathological Images With Inception Recurrent Residual Convolutional Neural Network. *J Digital Imaging* (2019) 32 (5):605–17. doi: 10.1007/s10278-019-00182-7
21. Toğaçar M, ÖZkurt KB, Ergen B, Cömert Z. Breastnet: A Novel Convolutional Neural Network Model Through Histopathological Images for the Diagnosis of Breast Cancer. *Phys A: Statal Mech its Appl* (2019) 545:123592. doi: 10.1016/j.physa.2019.123592
22. Benhammou Y, Achchab B, Herrera F, Tabik S. BreakHis Based Breast Cancer Automatic Diagnosis Using Deep Learning: Taxonomy, Survey and Insights. *Neurocomputing* (2020) 375:9–24. doi: 10.1016/j.neucom.2019.09.044
23. Arau´jo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, et al. Classification of Breast Cancer Histology Images Using Convolutional Neural Networks. *PloS One* (2017) 12(6):e0177544. doi: 10.1371/journal.pone.0177544
24. Li Y, Wu J, Wu Q. Classification of Breast Cancer Histology Images Using Multi-Size and Discriminative Patches Based on Deep Learning. *IEEE Access* (2019) 7:21400–8. doi: 10.1109/ACCESS.2019.2898044
25. Vo DM, Nguyen NQ, Lee SW. Classification of Breast Cancer Histology Images Using Incremental Boosting Convolution Networks. *Inf Sci* (2019) 482:123–38. doi: 10.1016/j.ins.2018.12.089
26. Yan R, Ren F, Wang Z, Wang L, Zhang T, Liu Y. Breast Cancer Histopathological Image Classification Using a Hybrid Deep Neural Network. *Methods* (2020) 173:52–60. doi: 10.1016/j.ymeth.2019.06.014
27. Haralick RM, Shanmugam K, Dinstein I. Textural Features of Image Classification. *IEEE Trans Sys Man Cybern* (1973) 6:610–21. doi: 10.1109/TSMC.1973.4309314
28. Soh LK, Tsatsoulis C. Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices. *IEEE Trans Geosci Remote Sens* (1999) 37 (2):780–95. doi: 10.1109/36.752194

29. Clausi DA. An Analysis of Co-Occurrence Texture Statistics as a Function of Grey Level Quantization. *Can J Remote Sens* (2002) 28(1):45–62. doi: 10.5589/m02-004
30. Hu MK. Visual Pattern Recognition by Moment Invariants. *IRE Trans Inf Theory* (1962) 8(2):179–87. doi: 10.1109/TIT.1962.1057692
31. Tamura H, Mori S, Yamawaki T. Textural Features Corresponding to Visual Perception. *IEEE Trans Sys Man Cybern* (1978) 8(6):460–73. doi: 10.1109/TSMC.1978.4309999
32. Ojala T, Pietikainen M, Maenpaa T. Multiresolution Gray-Scale and Rotation Invariant Texture and Classification with Local Binary Patterns. *IEEE Trans Pattern Anal Mach Intell* (2002) 24(7):971. doi: 10.1007/3-540-44732-6_41
33. Guo Z, Zhang L, Zhang D. A Completed Modeling of Local Binary Pattern Operator for Texture Classification. *IEEE Trans Image Process* (2010) 19 (6):1657–63. doi: 10.1109/TIP.2010.2044957
34. Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. In: *2005 IEEE Computer Society Conference on Computer Vision & Pattern Recognition (CVPR'05)*. San Diego, CA, USA: IEEE (2005) 1:886–93. doi: 10.1109/CVPR.2005.177
35. Gupta V, Bhavsar A. Breast Cancer Histopathological Image Classification: Is Magnification Important? In: *2017 IEEE Conference on Computer Vision & Pattern Recognition Workshops (CVPRW)*. Honolulu, HI, USA: IEEE (2017):769–76. doi: 10.1109/CVPRW.2017.107
36. Das K, Conjeti S, Roy AG, Chatterjee J, Sheet D. Multiple Instance Learning of Deep Convolutional Neural Networks for Breast Histopathology Whole Slide Classification. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. Washington, DC, USA: IEEE (2018) 578–81. doi: 10.1109/ISBI.2018.8363642
37. Das K, Karri S, Roy AG, Chatterjee J, Sheet D. Classifying Histopathology Whole-Slides Using Fusion of Decisions From Deep Convolutional Network on a Collection of Random Multi-Views At Multi-Magnification. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. Melbourne, VIC, Australia: IEEE (2017):1024–7. doi: 10.1109/ISBI.2017.7950690
38. Cascianelli S, Bello-Cerezo R, Bianconi F, Fravolini ML, Kather JN. Dimensionality Reduction Strategies for CNN-Based Classification of Histopathological Images. *Int Conf Intelligent Interact Multimed Syst Serv* (2018) 76:21–30. doi: 10.1007/978-3-319-59480-4_3
39. Wei B, Han Z, He X, Yin Y. Deep Learning Model Based Breast Cancer Histopathological Image Classification. In: *2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (Icccbda)*. Chengdu, China: IEEE (2017) 348–53. doi: 10.1109/ICCCBDA.2017.7951937
40. Zhi W, Yueng HWF, Chen Z, Zandavi SD, Lu Z, Chung YY. Using Transfer Learning With Convolutional Neural Networks to Diagnose Breast Cancer From Histopathological Images. In: *The 24th International Conference On Neural Information Processing*. Guangzhou, China: Springer, Cham (2017) 10637:669–76. doi: 10.1007/978-3-319-70093-9_71
41. Nahid AA, Kong Y. Histopathological Breast-Image Classification Using Concatenated R–G–B Histogram Information. *Ann Data Sci* (2018 6:513–29. doi: 10.1007/s40745-018-0162-3
42. Han Z, Wei B, Zheng Y, Yin Y, Li K, Li S. Breast Cancer Multi-Classification From Histopathological Images with Structured Deep Learning Model. *Sci Rep* (2017) 7(1):4172. doi: 10.1038/s41598-017-04075-z
43. Boumaraf S, Liu X, Zheng Z, Ma X, Ferkous C. A New Transfer Learning Based Approach to Magnification Dependent and Independent Classification of Breast Cancer in Histopathological Images. *Biomed Signal Process Control* (2021) 63:102192. doi: 10.1016/j.bspc.2020.102192
44. Song Y, Zou J, Chang H, Cai W. Adapting Fisher Vectors for Histopathology Image Classification. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. Melbourne, VIC, Australia: IEEE (2017) 600–3. doi: 10.1109/ISBI.2017.7950592
45. Saxena S, Shukla S, Gyanchandani M. Pre-Trained Convolutional Neural Networks as Feature Extractors for Diagnosis of Breast Cancer Using Histopathology. *Int J Imaging Syst Technol* (2020) 30:577–91. doi: 10.1002/ima.22399
46. Wang P, Wang J, Li Y, Li P, Li L. Automatic Classification of Breast Cancer Histopathological Images Based on Deep Feature Fusion and Enhanced Routing. *Biomed Signal Process Control* (2021) 65(6):102341. doi: 10.1016/j.bspc.2020.102341
47. George K, Sankaran P, Joseph KP. Computer Assisted Recognition of Breast Cancer in Biopsy Images Via Fusion of Nucleus-Guided Deep Convolutional Features. *Comput Methods Prog Biomed* (2020) 194:105531. doi: 10.1016/j.cmpb.2020.105531

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 22 July 2021
doi: 10.3389/fonc.2021.606677

Application of CT-Based Radiomics in Discriminating Pancreatic Cystadenomas From Pancreatic Neuroendocrine Tumors Using Machine Learning Methods

Xuejiao Han[1†], Jing Yang[2,3†], Jingwen Luo[1†], Pengan Chen[4], Zilong Zhang[4], Aqu Alu[1], Yinan Xiao[4] and Xuelei Ma[1]*

[1] Department of Biotherapy, Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [2] State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China, [3] Melanoma and Sarcoma Medical Oncology Unit, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China, [4] West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China

Objectives: The purpose of this study aimed at investigating the reliability of radiomics features extracted from contrast-enhanced CT in differentiating pancreatic cystadenomas from pancreatic neuroendocrine tumors (PNETs) using machine-learning methods.

Methods: In this study, a total number of 120 patients, including 66 pancreatic cystadenomas patients and 54 PNETs patients were enrolled. Forty-eight radiomic features were extracted from contrast-enhanced CT images using LIFEx software. Five feature selection methods were adopted to determine the appropriate features for classifiers. Then, nine machine learning classifiers were employed to build predictive models. The performance of the forty-five models was evaluated with area under the curve (AUC), accuracy, sensitivity, specificity, and F1 score in the testing group.

Results: The predictive models exhibited reliable ability of differentiating pancreatic cystadenomas from PNETs when combined with suitable selection methods. A combination of DC as the selection method and RF as the classifier, as well as Xgboost+RF, demonstrated the best discriminative ability, with the highest AUC of 0.997 in the testing group.

Conclusions: Radiomics-based machine learning methods might be a noninvasive tool to assist in differentiating pancreatic cystadenomas and PNETs.

Keywords: pancreatic cystadenomas, pancreatic neuroendocrine tumors, radiomics, machine learning, differentiation, pNETs, CT

OPEN ACCESS

Edited by:
Giuseppe Luigi Banna, United Lincolnshire Hospitals NHS Trust, United Kingdom

Reviewed by:
Isaac Shiri, Geneva University Hospitals (HUG), Switzerland
Jiawen Zhang, Fudan University, China

****Correspondence:***
Xuelei Ma
drmaxuelei@gmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *15 September 2020*
Accepted: *05 July 2021*
Published: *22 July 2021*

Citation:
Han X, Yang J, Luo J, Chen P, Zhang Z, Alu A, Xiao Y and Ma X (2021) Application of CT-Based Radiomics in Discriminating Pancreatic Cystadenomas From Pancreatic Neuroendocrine Tumors Using Machine Learning Methods. Front. Oncol. 11:606677. doi: 10.3389/fonc.2021.606677

INTRODUCTION

Pancreatic neuroendocrine tumors (PNETs), a rare group of heterogeneous tumors originated from ductal pluripotent stem cells, account for less than 5% of pancreatic neoplasms and 7% of all NETs (1–3). Its incidence has increased in recent years, reaching 0.48 per 100 000 persons per year in the United States. This increase is probably due to the improvement in modern imaging and endoscopic technologies (4). Based on the clinical manifestations, PNETs have two subtypes, functional and non-functional PNETs. The most common functional PNETs includes insulinoma, gastrinoma, and glucagonoma. However, about 2/3 of all PNETs are non-functional, making the diagnosis difficult in clinical practice (5). For localized and advanced, but resectable PNETs, surgery is the first-line therapy capable of improving the clinical outcome (6–8). Pancreatic cystadenomas including serous pancreatic cystadenomas and mucinous pancreatic cystadenomas, account for approximately 46.3% of all surgically removed pancreatic cystic tumors (9). Serous cystadenomas are rare glycogen-rich lesions, which arise from pancreatic ductal epithelium (10). But mucinous cystadenomas are cystic epithelial tumors, consisting of ovarian stromata and mucus-producing columnar epithelium (11). In clinical practice, most pancreatic cystadenoma patients are asymptomatic or manifest non-specific symptoms (12). The management of patients is different due to the biological differences of PNETs and pancreatic cystadenomas. The endoscopic ultrasound fine-needle aspiration (EUS-FNA) is considered the best approach to diagnosis pancreatic tumors, but it is invasive and not completely accurate due to the small size of samples (13, 14). Therefore, a preoperative differential diagnosis is vital to identify the most appropriate therapies and improve clinical management.

Generally, computed tomography (CT) is the most effective imaging technique for initial detection and tumor staging among pancreatic patients (15–18). Previous studies showed that CT could clearly show the tumor site and boundary, maximum diameter, cyst wall characteristics, enhancement degree and other imaging signs, which may contribute to differentiating pancreatic cystadenoma from PNETs (18–21). Radiomics is a high-throughput method to extract quantitative imaging features, which can conduct mining and analysis of image feature data in depth. The strength of radiomics is to provide an objective, repeatable, non-invasive and low-risk diagnostic tool, which helps to derive a comprehensive characterization of tumors heterogeneity. It has plenty of applications including biological feature prediction and tumor classification (14, 22, 23). According to recent studies, the radiological features extracted from CT images are helpful for differentiating pancreatic neoplasms (19, 24, 25), as well as the prediction of PNETs grading (21). For example, He et al. developed three models to differentiate non-functional PNET and pancreatic ductal adenocarcinoma (PDAC), which all showed good performance (26). The AUC of the validation cohort was 0.884 in the LASSO based model that integrated clinicoradiological features and radiomic. Another study built a radiomics model based on non-contrast MRI to predict grades of non-functional PNETs and the AUC was 0.769 in the training cohort and 0.729 in the validation cohort (27). However, there are no studies to differentiate pancreatic cystadenomas from PNETs. Therefore, we conducted this study to identify pancreatic cystadenomas and PNETs using machine learning methods based on enhanced CT image features.

MATERIALS AND METHODS

Patient Selection

We retrospectively included all patient with pancreatic cystadenomas including pancreatic mucinous and serous cystadenoma or PNETs in West China hospital from January 2013 to May 2018 in this study. We initially identified 356 eligible patients according to criteria as followed: (1) pathological diagnosis of pancreatic cystadenomas or PNETs; (2) enhanced-contrast CT examination before biopsy or surgery. Then 92 of 356 patients without exact pathological evidence supporting pancreatic cystadenomas or PNETs were excluded. In addition, 144 patients lacking abdominal enhanced-contrast CT images before surgery were also excluded. Finally, we enrolled 120 qualified patients, consisting of 66 pancreatic cystadenomas patients and 54 PNETs patients. The patient selection process was illustrated in **Figure 1**. All procedures conformed to the Declaration of Helsinki, as well as its later amendments.

CT Image Acquisition

All the contrast-enhanced CT images were collected from the patients before they received any treatment. A single 64-MDCT scanner (Brilliance64, Philips Medical Systems, Eindhoven, The Netherlands) were used for scanning. The tube voltage was 120 kVp, and tube current was 200 mAs. The gantry rotation time was 0.42 s, and the detector collimation was 0.75 mm. The matrix was 512×512 and beam pitch was 0.891. The slice thickness was 1.0 mm and reconstruction increment was 5.0 mm. Before the examination, patients were administrated with 1.5-2.0 mL/kg of nonionic contrast material (Omnipaque 350, GE Healthcare) intravenously at 3 mL/s. Images were obtained at 60-65 s for portal venous phase.

ROI Segmentation and Texture Extraction

All the contrast-enhanced CT images (400-bit gray scale) were obtained through searching the picture archiving and communication system (PACS). The lesion segmentation and texture analysis were acquired with a local image features extraction software (LIFEx v3.74, CEA-SHFJ, Orsay, France) (9). Briefly, the two-dimensional region of interest (ROI) was firstly obtained by drawing the outline of each tumor in the portal vein phase CT images (**Figure 2**) and then three-dimensional ROIs were automatically generated with default setting (28, 29). In order to reduce bias in the evaluation of derived radiomic features, the whole process was performed

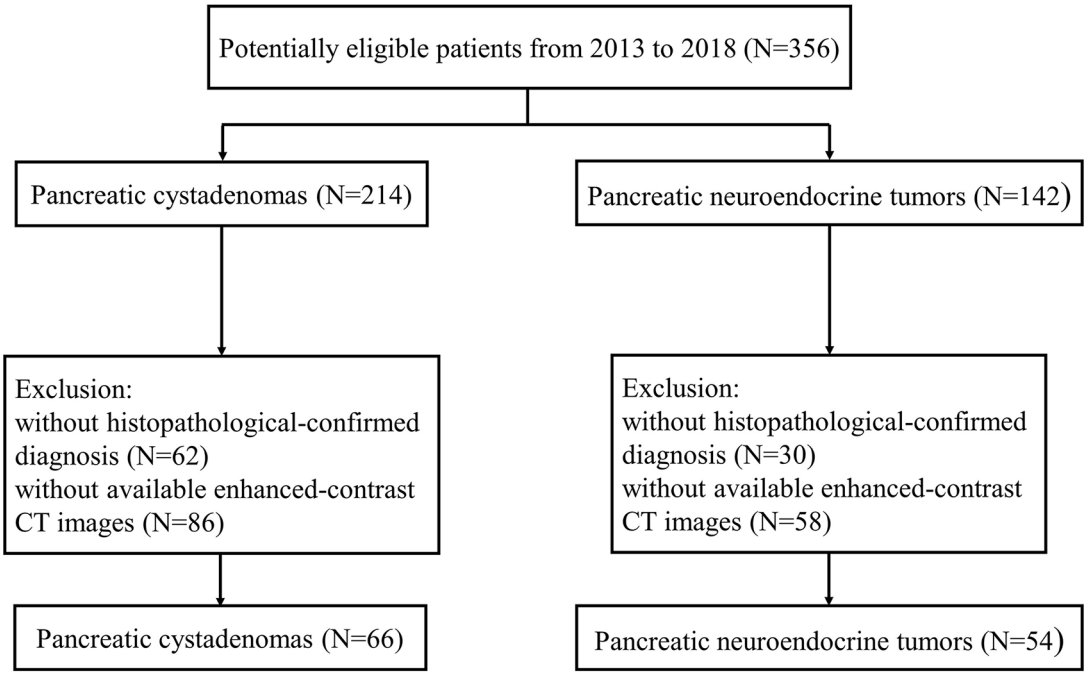


FIGURE 1 | Flowchart shows selection of study population and exclusion criteria.

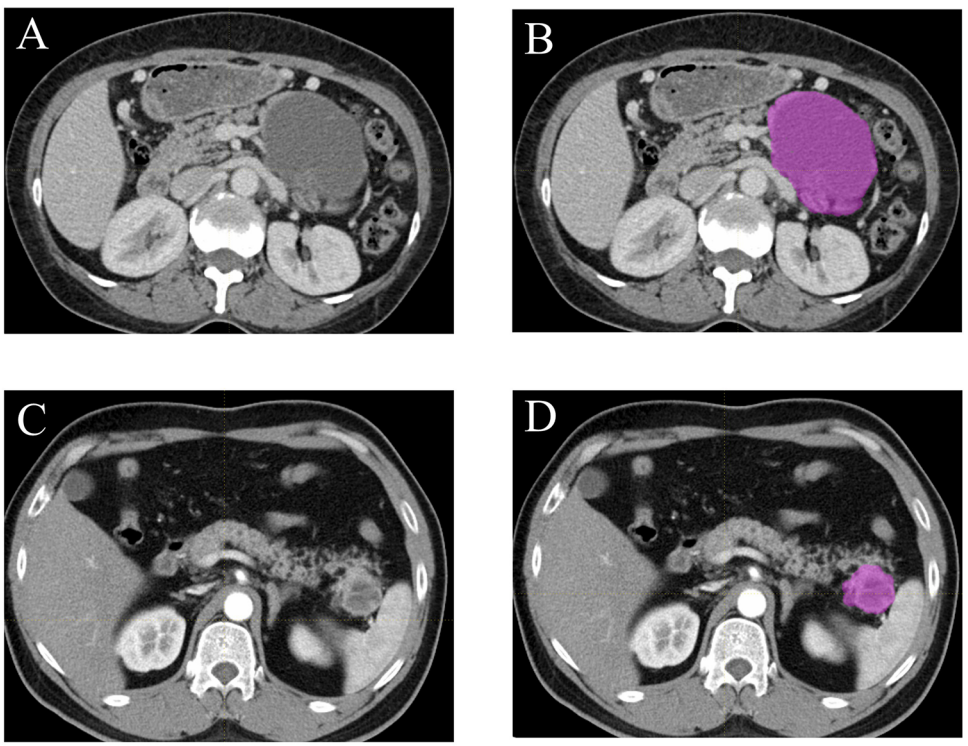


FIGURE 2 | Examples of lesion segmentation and contouring on contrast-enhanced CT images. Portal vein phase CT images of a patient with histopathologically proved pancreatic cystadenomas **(A, B)** and pancreatic neuroendocrine tumors **(C, D)**.

independently by two experienced radiologists without relevant knowledge of pancreatic tumor diagnosis and a third radiologist evaluated and selected the more accurate ROIs. The ROIs included the whole tumor while avoiding vascular shadows and surrounding adipose tissue. Then the textural parameters were calculated automatically based on the ROIs by the computer software LIFEx. Forty-eight texture features were extracted, including histogram-based matrix, shape-based matrix, gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), neighborhood gray-level dependence matrix (NGLDM) and gray-level zone length matrix (GLZLM). The association among texture parameters was analyzed by Pearson correlation coefficient test.

Machine Learning Model

In this study, five feature selection methods, namely, random forest (RF), distance correlation (DC), least absolute shrinkage and selection operator (LASSO), gradient boosting decision tree (GBDT) and eXtreme gradient boosting (Xgboost) were used to analyze the texture parameters and several clinicoradiological features (gender, age, size, location of lesions and calcification of lesions). Then, discriminative models were built by nine machine learning classifiers, which covers linear discriminant analysis (LDA), RF, Adaboost, support vector machine (SVM), GaussianNB, k-nearest neighbor (KNN), GBDT, logistic regression (LR) and decision tree (DT). Patients meeting the criteria were randomly assigned to either training group or validation group. The ration of patient number in training group to validation group was 4:1. The classification models were generated from the training group and the discriminative capability of models were tested in the validation group. We performed 10-fold cross validation with 1000 repetition to guarantee the real distribution of classification. The sensitivity, specificity, accuracy and F1 score were calculated accordingly after establishing a confusion matrix. Besides, the area of receiver operating characteristic curve (AUC) was recorded to assess the discriminative ability of the classification models. The machine learning algorithm was completed by the Python programming language and sklear-Package. The whole study process was shown in **Figure 3**.

RESULTS

Patient Characteristics

In this study, a total number of 120 patients (including 66 individuals with pancreatic cystadenomas and 54 patients with PNETs) were enrolled. The median age of all patients was 50.26 (ranging from 24 to 77) in the group of pancreatic cystadenomas, and 54 (ranging from 29 to 73) in the PNETs group. Of the 66 patients with pancreatic cystadenomas, 15 (22.7%) were males and 51 (77.3%) were females. In PNETs group, there were 32 (59.3%) male and 22 (40.7%) female patients. In the patients with pancreatic cystadenomas, 40.9% of lesions were located in the pancreatic head, whereas 59.1% were in the pancreatic body-tail. Of the 54 patients with PNETs, the tumor was located in the head of pancreas in 23 patients (42.6%) and in the pancreatic body-tail in 31 patients. The average size of pancreatic cystadenomas and PNETs was 4.1 cm (range 0.8-12.0 cm) and 4.19 cm (range 1.0-12.0 cm) respectively. The summary of patient and lesion characteristics were recorded in **Table 1**.

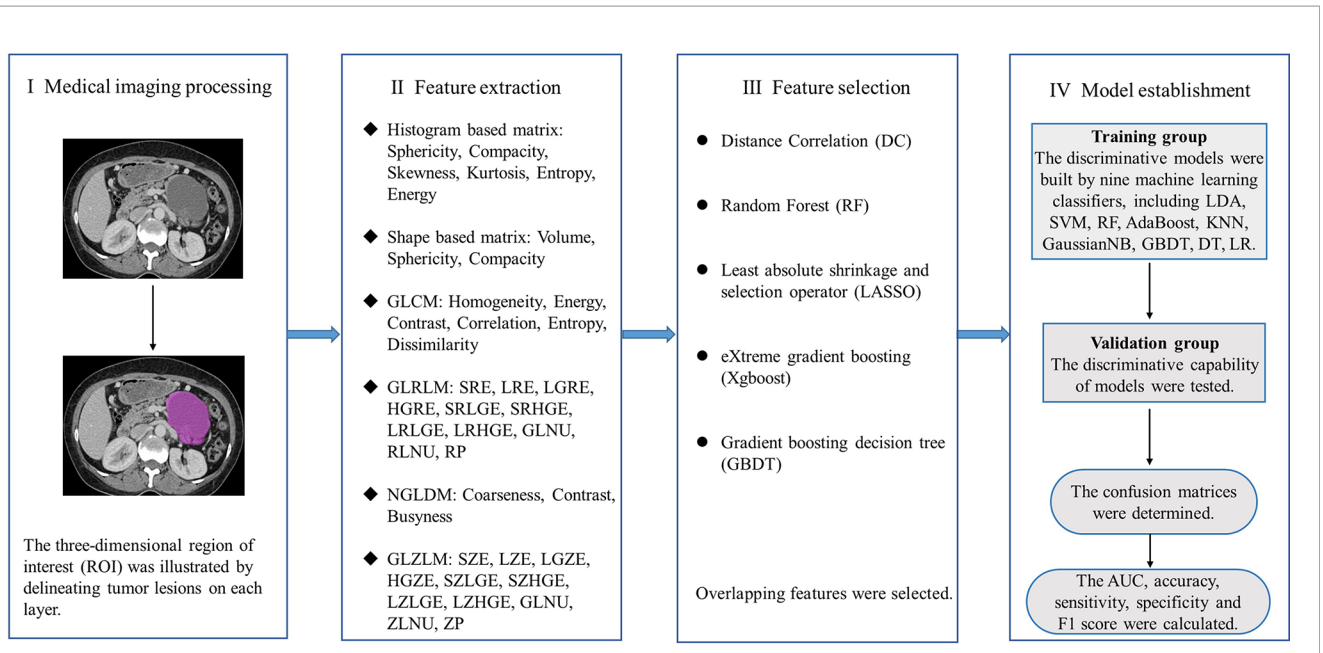


FIGURE 3 | Flowchart depicts the workflow of the whole study.

TABLE 1 | Characteristics of patients and lesions.

Characteristics	Pancreatic cystadenomas	Pancreatic neuroendocrine tumors
Number	66	54
Mean age (range) (year)	50.26 (24-77)	50.48 (29-73)
Gender		
Male	15 (22.7%)	32 (59.3%)
Female	51 (77.3%)	22 (40.7%)
Location		
Head	27 (40.9%)	23 (42.6%)
Body or tail	39 (59.1%)	31 (57.4%)
Maximum diameter (range) (cm)	4.1 (0.8-12)	4.19 (1-12)

Texture Parameters Selection

Most texture parameters were uncorrelated or weakly correlated according to the result of Pearson correlation coefficient test, only a few parameters showing strong positive or negative correlation (**Figure 4**). Different texture parameters were selected in five feature selection methods, however, no clinicoradiological feature was selected using the five feature selection methods (**Table 2**). LASSO selected the most number of parameters, including maxValue, HISTO_Skewness, SHAPE_Sphericity, GLRLM_SRLGE, GLRLM_GLNU, GLRLM_RLNU, GLRLM_RP, GLZLM_SZE, GLZLM_LGZE, GLZLM_SZLGE, GLZLM_LZLGE, GLZLM_GLNU, GLZLM_ZLNU, GLZLM_ZP and Xgboost selected only four parameters.

Machine-Learning Model Evaluation

Through a random combination of five feature selection methods and nine machine-learning classifiers, we acquired a total of forty-five predictive models for distinguishing pancreatic cystadenomas from PNETs. The AUC, accuracy, sensitivity, specificity and F1 score of all models in the testing group were shown in **Table 3**. The result revealed that radiomics-based machine learning models were able to differentiate pancreatic cystadenoma from PNETs, with AUC more than 0.743 in the validation cohort (**Figure 5**). It is noteworthy that the combination of DC as the selection method and RF as the classifier, as well as Xgboost+RF, demonstrated the best discriminative ability, with the highest AUC of 0.997 in the testing group. The receiver operating characteristic (ROC) curves of 10 fold for DC+RF and Xgboost+RF in the testing group were shown in **Figure 6**. For the model (DC+RF) in the testing group, the accuracy, sensitivity, specificity and F1 score were 0.983, 0.980, 0.986 and 0.980, respectively. The mean AUC for DC+RF was 0.9977 (Std= 0.0024; 95% CI, 0.9976 to 0.9979) after 1000 repetition. As for the model of Xgboost+RF, the accuracy, sensitivity, specificity and F1 score were 0.992, 0.980, 1.000 and

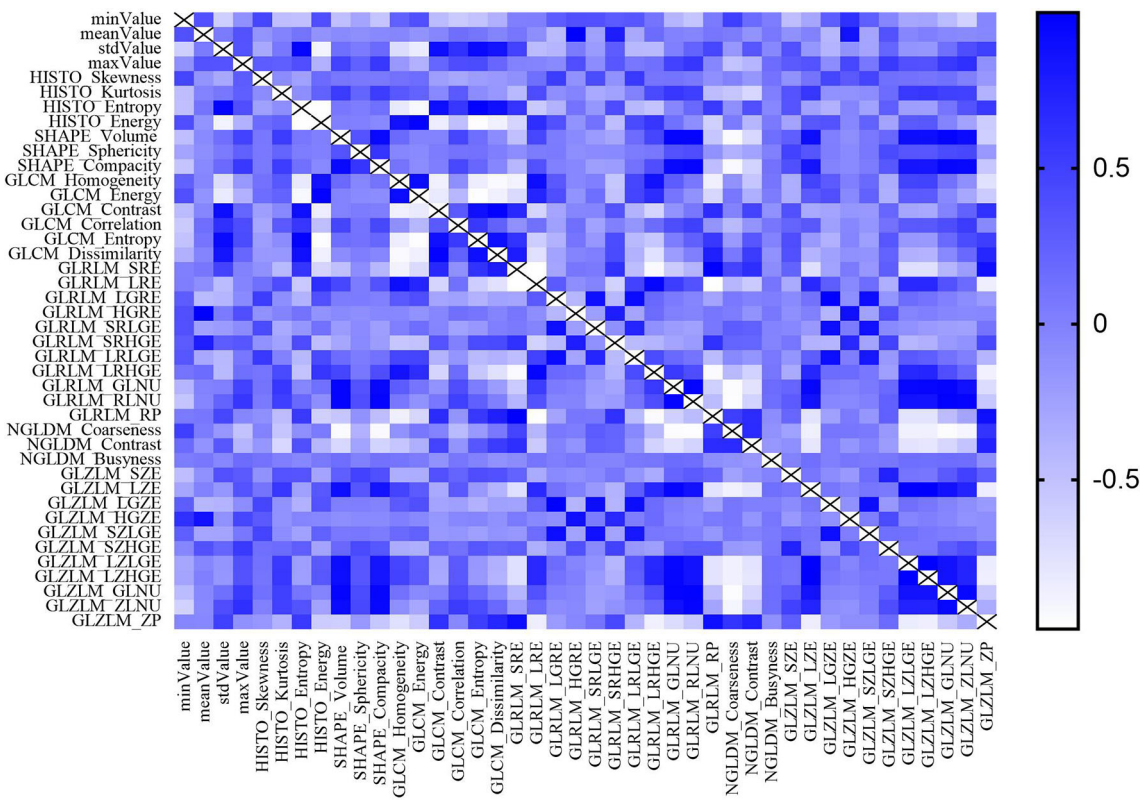


FIGURE 4 | The result of Pearson correlation coefficients test between radiomics features.

TABLE 2 | Selected features of five selection methods.

DC	RF	LASSO	Xgboost	GBDT
meanValue	meanValue	maxValue	meanValue	HISTO_Kurtosis
HISTO_Kurtosis	HISTO_Kurtosis	HISTO_Skewness	HISTO_Kurtosis	GLRLM_SRHGE
GLZLM_HGZE	PARAMS_YSpatialResampling	SHAPE_Sphericity	GLRLM_HGRE	GLRLM_LRHGE
GLZLM_SZHGE	GLRLM_HGRE	GLRLM_SRLGE	GLRLM_LRHGE	GLZLM_HGZE
GLRLM_HGRE	GLRLM_SRHGE	GLRLM_GLNU		GLZLM_SZHGE
GLRLM_SRHGE	GLRLM_LRHGE	GLRLM_RLNU		
GLZLM_SZE	GLZLM_SZE	GLRLM_RP		
	GLZLM_HGZE	GLZLM_SZE		
	GLZLM_SZHGE	GLZLM_LGZE		
		GLZLM_SZLGE		
		GLZLM_LZLGE		
		GLZLM_GLNU		
		GLZLM_ZLNU		
		GLZLM_ZP		

DC, distance correlation; RF, random forest; LASSO, least absolute shrinkage and selection operator; Xgboost, eXtreme gradient boosting; GBDT, gradient boosting decision tree; GLZLM, gray-level zone length matrix; HGZE, High Grey Level Zone Emphasis; GLRLM, Gray Level Run Length Matrix; SZHGE, Short Zone High Grey Level Emphasis; HGRE, High Gray Level Run Emphasis; SRHGE, Short-Run High Grey Level Emphasis; SZE, Short Zone Emphasis; LRHGE, Long-Run High Grey Level Emphasis; SRLGE, Short-Run Low Grey Level Emphasis; GLNU, Grey Level Non-Uniformity; RLNU, Run Length Non-Uniformity; RP, Run Percentage; LGZE, Low Gray Level Zone Emphasis; SZLGE, Short Zone Low Grey Level Emphasis; LZLGE, Long Zone Low Grey Level Emphasis; ZP, Zone Percentage.

0.989 respectively in the validation group and the mean AUC was 0.9962 (Std=0.0033; 95% CI, 0.9960 to 0.9964) after 1000 repetition.

DISCUSSION

Contrast-enhanced CT and MRI are widely used for identification of pancreatic neoplasms, which is not sufficiently specific due to the overlapping of anatomical imaging features between different pancreatic tumors (30). Furthermore, PNETs identification is to exploit somatostatin analog radiotracers for CT and MRI imaging, which is easily affected by intra-/inter-observer variability (31). The accuracy for manual identification of tumor lesions is only 60%-70%, even by highly-trained radiologists (32). To date, EUS-FNA is considered the best approach to diagnosis pancreatic tumors based on cytopathologic features. However, it is invasive with the possibility for puncture failure, as well as unable to reflect tumor heterogeneity accurately due to the small size and limited number of samples (14, 33).

Radiomics combine quantitative image analysis with machine learning or artificial intelligence methods to select out and classify target features in the sample images. It has been applied in modern medical care including diagnosis, risk stratification, virtual biopsy and radiogenomics (34). Several studies have investigated the utility of machine learning based-radiomics on the differentiation of pancreatic mucinous cystadenomas from pancreatic serous cystadenomas (35, 36) and the prediction of PNETs grading (13, 21, 37–39). However, no studies reported how to differentiate pancreatic cystadenomas from PNETs. Given the similar incidence, nonspecific symptoms and various biological behaviors of pancreatic cystadenomas and PNETs, we conducted this study by combining radiomics and machine learning method to distinguish these two types of pancreatic lesions. This is the first study that utilized enhanced CT images features and machine learning methods to differentiate pancreatic cystadenomas from PNETs so far.

Different radiomic features were obtained from preoperative contrast-enhanced CT images. Then the establishment of predictive models were based on five feature-selection methods and nine machine-learning classifiers. Our results proved that the combination of predictive models and appropriate selection methods could differentiate pancreatic cystadenomas from PNETs, with AUC more than 0.743 in the validation cohort. Notably, the combination of DC as the selection method and RF as the classifier, as well as Xgboost+RF, seemed to be optimal in differentiating the two types of pancreatic lesions, with the highest AUC of 0.997 in the testing group. For the model (DC+RF) in the testing group, the accuracy, sensitivity, specificity and F1 score were 0.983, 0.980, 0.986 and 0.980, respectively. In addition, the accuracy, sensitivity, specificity and F1 score were 0.992, 0.980, 1.000 and 0.989 respectively in the validation group for the model of Xgboost+RF. RF classifier is an excellent classification algorithm that has been widely used in many studies (40–42). Previous studies have investigated the diagnostic performance of PNETs from PDACs based on CT features and texture analysis. Yu et al. used LASSO and univariate logistic regression (ULR) analyses to select CT radiomic features and generated four multivariate logistic regression (MLR) models. The highest AUC was 0.926 in the model built with CT radiomic features extracted from portal venous phase (43). Another study developed a MLR model to discriminate PNETs from solid pseudopapillary tumors (SPTs) (44). The model incorporating MRI radiomics features and sex and age of patients exhibited the best discriminative ability with AUC of 0.97 and 0.86 in the training and validation cohort separately. Compared with previous studies, we developed more predictive models by employing more selection methods and classifiers to identify the best model.

However, our study existed several limitations. First of all, this study was conducted in a retrospective fashion, which had unavoidable selection bias. Second, it was a single-center study with a small group of patients. Multicenter studies with more patients would be helpful to confirm these findings. Third, there

TABLE 3 | Results of discriminative models in distinguishing pancreatic cystadenomas and pancreatic neuroendocrine tumors in the testing group.

	DC					RF					LASSO					Xgboost					GBDT				
	AUC	Accuracy	Sensitivity	Specificity	F1-score	AUC	Accuracy	Sensitivity	Specificity	F1-score	AUC	Accuracy	Sensitivity	Specificity	F1-score	AUC	Accuracy	Sensitivity	Specificity	F1-score	AUC	Accuracy	Sensitivity	Specificity	F1-score
LDA	0.907	0.858	0.797	0.912	0.832	0.915	0.917	0.847	0.971	0.890	0.901	0.867	0.763	0.955	0.832	0.947	0.917	0.817	1.000	0.894	0.918	0.875	0.810	0.929	0.850
SVM	0.833	0.700	0.633	0.757	0.653	0.972	0.725	0.430	0.971	0.542	0.777	0.642	0.267	0.957	0.365	0.946	0.833	0.650	0.986	0.764	0.966	0.908	0.817	0.986	0.886
RF	0.997	0.983	0.980	0.986	0.980	0.994	0.975	0.960	0.986	0.969	0.975	0.960	0.947	0.952	0.948	0.997	0.992	0.980	1.000	0.989	0.989	0.992	0.980	1.000	0.989
Adaboost	0.990	0.967	0.960	0.967	0.961	0.990	0.967	0.960	0.967	0.961	0.976	0.975	1.000	0.952	0.977	0.990	0.975	0.940	1.000	0.964	0.990	0.975	0.943	1.000	0.966
KNN	0.959	0.925	0.870	0.971	0.908	0.983	0.967	0.967	0.971	0.964	0.760	0.675	0.647	0.700	0.634	0.932	0.925	0.890	0.957	0.912	0.969	0.975	0.940	1.000	0.967
GaussianNB	0.973	0.942	0.893	0.986	0.928	0.973	0.775	0.523	0.986	0.654	0.986	0.975	0.963	0.986	0.971	0.926	0.633	0.223	0.971	0.292	0.926	0.608	0.167	0.971	0.209
LR	0.882	0.692	0.613	0.757	0.633	0.946	0.708	0.393	0.971	0.492	0.743	0.675	0.360	0.940	0.458	0.935	0.708	0.377	0.986	0.502	0.909	0.625	0.170	1.000	0.276
GBDT	0.989	0.975	0.940	0.969	0.972	0.979	0.983	0.980	0.986	0.980	0.976	0.975	1.000	0.952	0.977	0.993	0.983	0.980	0.986	0.980	0.983	0.963	0.980	0.983	0.981
DT	0.975	0.975	0.940	0.969	0.972	0.983	0.983	0.980	0.986	0.980	0.976	0.975	1.000	0.952	0.977	0.990	0.992	0.980	1.000	0.989	0.982	0.963	0.980	0.983	0.981

DC, distance correlation; RF, random forest; LASSO, least absolute shrinkage and selection operator; Xgboost, eXtreme gradient boosting; GBDT, gradient boosting decision tree; LDA, linear discriminant analysis; SVM, support vector machine; KNN, k-nearest neighbor; LR, logistic regression; DT, decision tree; AUC, area under curve.

FIGURE 5 | The results of AUC **(A)**, sensitivity **(B)** and specificity **(C)** in the testing group.

was subjectivity when manually defining the tumor boundary on CT images.

CONCLUSION

This study concluded that radiomics-based machine learning method was able to preoperatively differentiate pancreatic cystadenomas from PNETs, which may guide clinical decision-making for better treatment. The feature selection method DC or Xgboost and classifier RF showed a good prospect in discriminating pancreatic cystadenomas from PNETs. However,

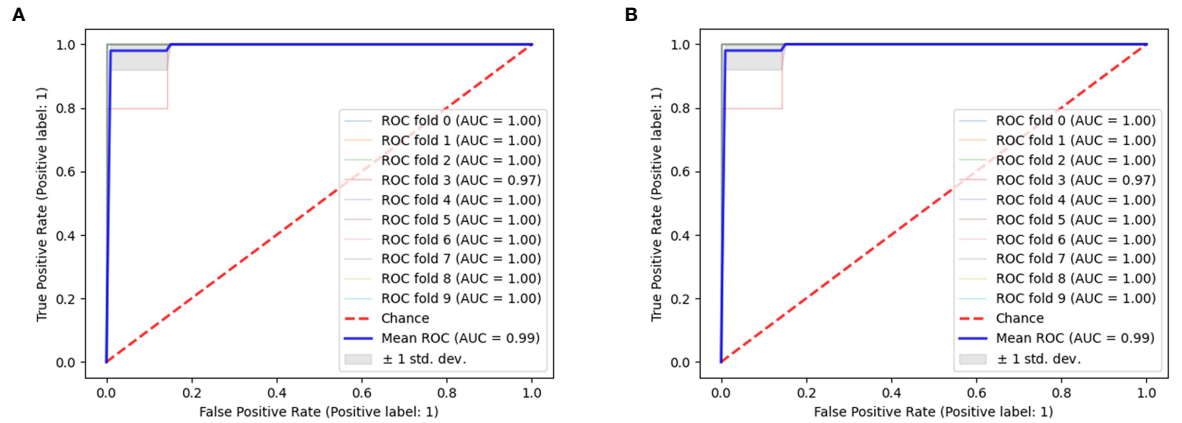


FIGURE 6 | The ROC curves of 10 fold for DC+RF **(A)** and Xgboost+RF **(B)** in the testing group.

more large-scale multicenter studies are required to supplement more evidence concerning the feasibility of this method in clinical practice.

DATA AVAILABILITY STATEMENT

The dataset generated for this study can be obtained from the correspondence author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Administration Office of the West China Hospital, Sichuan University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

XH and JY collected images, extracted features, performed statistical analysis and drafted the manuscript. JL collected image data and extracted texture feature. PC and ZZ collected image data and drafted the manuscript. AA and YX revised the manuscript. XM participated in study conception and manuscript revision. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Lewis RB, Lattin GE Jr, Paal E. Pancreatic Endocrine Tumors: Radiologic-Clinicopathologic Correlation. *Radiographics* (2010) 30(6):1445–64. doi: 10.1148/rg.306105523
2. Öberg K, Knigge U, Kwekkeboom D, Perren A. Neuroendocrine Gastro-Entero-Pancreatic Tumors: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann Oncol* (2012) 23 Suppl 7:vii124–30. doi: 10.1093/annonc/mds295
3. Lawrence B, Gustafsson BI, Chan A, Svejda B, Kidd M, Modlin IM. The Epidemiology of Gastroenteropancreatic Neuroendocrine Tumors. *Endocrinol Metab Clin North Am* (2011) 40(1):1–18, vii. doi: 10.1016/j.ecl.2010.12.005
4. Dasari A, Shen C, Halperin D, Zhao B, Zhou S, Xu Y, et al. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States. *JAMA Oncol* (2017) 3 (10):1335–42. doi: 10.1001/jamaoncol.2017.0589
5. Dumlu EG, Karakoç D, Özdemir A. Nonfunctional Pancreatic Neuroendocrine Tumors: Advances in Diagnosis, Management, and Controversies. *Int Surg* (2015) 100(6):1089–97. doi: 10.9738/INTSURG-D-14-00204.1
6. Falconi M, Bartsch DK, Eriksson B, Klöppel G, Lopes JM, O'Connor JM, et al. Enets Consensus Guidelines for the Management of Patients With Digestive Neuroendocrine Neoplasms of the Digestive System: Well-Differentiated Pancreatic non-Functioning Tumors. *Neuroendocrinology* (2012) 95(2):120–34. doi: 10.1159/000335587
7. Hill JS, McPhee JT, McDade TP, Zhou Z, Sullivan ME, Whalen GF, et al. Pancreatic Neuroendocrine Tumors: The Impact of Surgical Resection on Survival. *Cancer* (2009) 115(4):741–51. doi: 10.1002/cncr.24065
8. Falconi M, Eriksson B, Kaltsas G, Bartsch DK, Capdevila J, Caplin M, et al. Enets Consensus Guidelines Update for the Management of Patients With Functional Pancreatic Neuroendocrine Tumors and Non-Functional Pancreatic Neuroendocrine Tumors. *Neuroendocrinology* (2016) 103 (2):153–71. doi: 10.1159/000443171
9. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. Lifex: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res* (2018) 78(16):4786–9. doi: 10.1158/0008-5472.CAN-18-0125
10. Valsangkar NP, Morales-Oyarvide V, Thayer SP, Ferrone CR, Wargo JA, Warshaw AL, et al. 851 Resected Cystic Tumors of the Pancreas: A 33-Year Experience at the Massachusetts General Hospital. *Surgery* (2012) 152(3 Suppl 1):S4–12. doi: 10.1016/j.surg.2012.05.033
11. Zhao ZM, Jiang N, Gao YX, Yin ZZ, Zhao GD, Tan XL, et al. Clinical Diagnosis and Management of Pancreatic Mucinous Cystadenoma and Cystadenocarcinoma: Single-center Experience With 82 Patients. *World J Gastrointest Oncol* (2020) 12(6):642–50. doi: 10.4251/wjgo.v12.i6.642
12. Bai XL, Zhang Q, Masood N, Masood W, Zhang Y, Liang TB. Pancreatic Cystic Neoplasms: A Review of Preoperative Diagnosis and Management. *J Zhejiang Univ Sci B* (2013) 14(3):185–94. doi: 10.1631/jzus.B1200283

13. Zhao Z, Bian Y, Jiang H, Fang X, Li J, Cao K, et al. Ct-Radiomic Approach to Predict G1/2 Nonfunctional Pancreatic Neuroendocrine Tumor. *Acad Radiol* (2020) 27(12):e272–81. doi: 10.1016/j.acra.2020.01.002
14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
15. Del Frate C, Zanardi R, Mortele K, Ros PR. Advances in Imaging for Pancreatic Disease. *Curr Gastroenterol Rep* (2002) 4(2):140–8. doi: 10.1007/s11894-002-0051-x
16. Hasan A, Visrodia K, Farrell JJ, Gonda TA. Overview and Comparison of Guidelines for Management of Pancreatic Cystic Neoplasms. *World J Gastroenterol* (2019) 25(31):4405–13. doi: 10.3748/wjg.v25.i31.4405
17. Yuan D, Yu W, Ren XB, Pan WD, Zhang LH. Characterization and Diagnostic Accuracy of Serous Cystadenomas and Mucinous Neoplasms of the Pancreas With Multi-Slice Helical Computed Tomography. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao* (2007) 29(2):232–7.
18. Al-Hawary M. Role of Imaging in Diagnosing and Staging Pancreatic Cancer. *J Natl Compr Canc Netw* (2016) 14(5 Suppl):678–80. doi: 10.6004/jnccn.2016.0191
19. Guo C, Zhuge X, Wang Q, Xiao W, Wang Z, Wang Z, et al. The Differentiation of Pancreatic Neuroendocrine Carcinoma From Pancreatic Ductal Adenocarcinoma: The Values of CT Imaging Features and Texture Analysis. *Cancer Imaging* (2018) 18(1):37. doi: 10.1186/s40644-018-0170-8
20. Canellas R, Burk KS, Parakh A, Sahani DV. Prediction of Pancreatic Neuroendocrine Tumor Grade Based on CT Features and Texture Analysis. *AJR Am J Roentgenol* (2018) 210(2):341–6. doi: 10.2214/AJR.17.18417
21. Gu D, Hu Y, Ding H, Wei J, Chen K, Liu H, et al. CT Radiomics may Predict the Grade of Pancreatic Neuroendocrine Tumors: A Multicenter Study. *Eur Radiol* (2019) 29(12):6880–90. doi: 10.1007/s00330-019-06176-x
22. Meng X, Xia W, Xie P, Zhang R, Li W, Wang M, et al. Preoperative Radiomic Signature Based on Multiparametric Magnetic Resonance Imaging for Noninvasive Evaluation of Biological Characteristics in Rectal Cancer. *Eur Radiol* (2019) 29(6):3200–9. doi: 10.1007/s00330-018-5763-x
23. Zhang Q, Xiao Y, Suo J, Shi J, Yu J, Guo Y, et al. Sonoelastomics for Breast Tumor Classification: A Radiomics Approach With Clustering-Based Feature Selection on Sonoelastography. *Ultrasound Med Biol* (2017) 43(5):1058–69. doi: 10.1016/j.ultrasmedbio.2016.12.016
24. Reinert CP, Baumgartner K, Hepp T, Bitzer M, Horger M. Complementary Role of Computed Tomography Texture Analysis for Differentiation of Pancreatic Ductal Adenocarcinoma From Pancreatic Neuroendocrine Tumors in the Portal-Venous Enhancement Phase. *Abdom Radiol (NY)* (2020) 45(3):750–8. doi: 10.1007/s00261-020-02406-9
25. Lin X, Xu L, Wu A. Differentiation of Intrapancreatic Accessory Spleen From Small Hypervascular Neuroendocrine Tumor of the Pancreas: Textural Analysis on Contrast-Enhanced Computed Tomography. *Acta Radiol* (2019) 60: (5):553–60. doi: 10.1177/0284185118788895
26. He M, Liu Z, Lin Y, Wan J, Li J, Xu K, et al. Differentiation of Atypical non-Functional Pancreatic Neuroendocrine Tumor and Pancreatic Ductal Adenocarcinoma Using CT Based Radiomics. *Eur J Radiol* (2019) 117:102–11. doi: 10.1016/j.ejrad.2019.05.024
27. Bian Y, Zhao Z, Jiang H, Fang X, Li J, Cao K, et al. Noncontrast Radiomics Approach for Predicting Grades of Nonfunctional Pancreatic Neuroendocrine Tumors. *J Magn Reson Imaging* (2020) 52(4):1124–36. doi: 10.1002/jmri.27176
28. Nardone V, Tini P, Pastina P, Botta C, Reginelli A, Carbone SF, et al. Radiomics Predicts Survival of Patients With Advanced non-Small Cell Lung Cancer Undergoing PD-1 Blockade Using Nivolumab. *Oncol Lett* (2020) 19(2):1559–66. doi: 10.3892/ol.2019.11220
29. Chen C, Guo X, Wang J, Guo W, Ma X, Xu J. The Diagnostic Value of Radiomics-Based Machine Learning in Predicting the Grade of Meningiomas Using Conventional Magnetic Resonance Imaging: A Preliminary Study. *Front Oncol* (2019) 9:1338. doi: 10.3389/fonc.2019.01338
30. Al-Hawary MM, Francis IR, Anderson MA. Pancreatic Solid and Cystic Neoplasms: Diagnostic Evaluation and Intervention. *Radiol Clin North Am* (2015) 53(5):1037–48. doi: 10.1016/j.rcl.2015.05.005
31. Li J, Lu J, Liang P, Li A, Hu Y, Shen Y, et al. Differentiation of Atypical Pancreatic Neuroendocrine Tumors From Pancreatic Ductal Adenocarcinomas: Using Whole-Tumor CT Texture Analysis as Quantitative Biomarkers. (2018) 7(10):4924–31. doi: 10.1002/cam4.1746
32. Sahani DV, Sainani NI, Blake MA, Crippa S, Mino-Kenudson M, del-Castillo CF. Prospective Evaluation of Reader Performance on MDCT in Characterization of Cystic Pancreatic Lesions and Prediction of Cyst Biologic Aggressiveness. *AJR Am J Roentgenol* (2011) 197(1):W53–61. doi: 10.2214/AJR.10.5866
33. Strobel O, Hinz U, Gluth A, Hank T, Hackert T, Bergmann F, et al. Pancreatic Adenocarcinoma: Number of Positive Nodes Allows to Distinguish Several N Categories. *Ann Surg* (2015) 261(5):961–9. doi: 10.1097/SLA.0000000000000814
34. Dalal V, Carmicheal J, Dhaliwal A, Jain M, Kaur S, Batra SK. Radiomics in Stratification of Pancreatic Cystic Lesions: Machine Learning in Action. *Cancer Lett* (2020) 469:228–37. doi: 10.1016/j.canlet.2019.10.023
35. Yang J, Guo X, Ou X, Zhang W, Ma X. Discrimination of Pancreatic Serous Cystadenomas From Mucinous Cystadenomas With Ct Textural Features: Based on Machine Learning. *Front Oncol* (2019) 9:494. doi: 10.3389/fonc.2019.00494
36. Xie H, Ma S, Guo X, Zhang X, Wang X. Preoperative Differentiation of Pancreatic Mucinous Cystic Neoplasm From Macrocystic Serous Cystic Adenoma Using Radiomics: Preliminary Findings and Comparison With Radiological Model. *Eur J Radiol* (2020) 122:108747. doi: 10.1016/j.ejrad.2019.108747
37. Luo Y, Chen X, Chen J, Song C, Shen J, Xiao H, et al. Preoperative Prediction of Pancreatic Neuroendocrine Neoplasms Grading Based on Enhanced Computed Tomography Imaging: Validation of Deep Learning With a Convolutional Neural Network. *Neuroendocrinology* (2020) 110(5):338–50. doi: 10.1159/000503291
38. Gao X, Wang X. Deep Learning for World Health Organization Grades of Pancreatic Neuroendocrine Tumors on Contrast-Enhanced Magnetic Resonance Images: A Preliminary Study. *Int J Comput Assist Radiol Surg* (2019) 14(11):1981–91. doi: 10.1007/s11548-019-02070-5
39. Bian Y, Jiang H, Ma C, Wang L, Zheng J, Jin G, et al. Ct-Based Radiomics Score for Distinguishing Between Grade 1 and Grade 2 Nonfunctioning Pancreatic Neuroendocrine Tumors. *AJR Am J Roentgenol* (2020), 1–12. doi: 10.2214/AJR.19.22123
40. Yao X, Sun K, Bu X, Zhao C, Jin Y. Classification of White Blood Cells Using Weighted Optimized Deformable Convolutional Neural Networks. *Artif Cells Nanomed Biotechnol* (2021) 49(1):147–55. doi: 10.1080/21691401.2021.1879823
41. Sanner MF, Dieguez L, Forli S. Improving Docking Power for Short Peptides Using Random Forest. *J Chem Inf Model* (2021) 61(6):3074–90. doi: 10.1021/acs.jcim.1c00573
42. Kimura K, Yoshida S. Usefulness of Texture Features of Apparent Diffusion Coefficient Maps in Predicting Chemoradiotherapy Response in Muscle-Invasive Bladder Cancer. *Eur Radiol* (2021). doi: 10.1007/s00330-021-08110-6
43. Yu H, Huang Z, Li M, Wei Y, Zhang L, Yang C, et al. Differential Diagnosis of Nonhypervascular Pancreatic Neuroendocrine Neoplasms From Pancreatic Ductal Adenocarcinomas, Based on Computed Tomography Radiological Features and Texture Analysis. *Acad Radiol* (2020) 27(3):332–41. doi: 10.1016/j.acra.2019.06.012
44. Shi YJ, Zhu HT, Liu YL, Wei YY, Qin XB, Zhang XY, et al. Radiomics Analysis Based on Diffusion Kurtosis Imaging and T2 Weighted Imaging for Differentiation of Pancreatic Neuroendocrine Tumors From Solid Pseudopapillary Tumors. *Front Oncol* (2020) 10:1624. doi: 10.3389/fonc.2020.01624

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroinformatics

ORIGINAL RESEARCH
published: 30 July 2021
doi: 10.3389/fninf.2021.667008

Deep Neuro-Vision Embedded Architecture for Safety Assessment in Perceptive Advanced Driver Assistance Systems: The Pedestrian Tracking System Use-Case

Francesco Rundo[1]*, *Sabrina Conoci*[2], *Concetto Spampinato*[3], *Roberto Leotta*[4], *Francesca Trenta*[4] and *Sebastiano Battiato*[4]

[1] STMicroelectronics, ADG Central R&D Division, Catania, Italy, [2] Department of Chemical, Biological, Pharmaceutical and Environmental Sciences, University of Messina, Messina, Italy, [3] PerCeiVe Lab, University of Catania, DIEEI, Catania, Italy, [4] IPLAB, Department of Mathematics and Computer Science, University of Catania, Catania, Italy

In recent years, the automotive field has been changed by the accelerated rise of new technologies. Specifically, autonomous driving has revolutionized the car manufacturer's approach to design the advanced systems compliant to vehicle environments. As a result, there is a growing demand for the development of intelligent technology in order to make modern vehicles safer and smarter. The impact of such technologies has led to the development of the so-called Advanced Driver Assistance Systems (ADAS), suitable to maintain control of the vehicle in order to avoid potentially dangerous situations while driving. Several studies confirmed that an inadequate driver's physiological condition could compromise the ability to drive safely. For this reason, assessing the car driver's physiological status has become one of the primary targets of the automotive research and development. Although a large number of efforts has been made by researchers to design safety-assessment applications based on the detection of physiological signals, embedding them into a car environment represents a challenging task. These mentioned implications triggered the development of this study in which we proposed an innovative pipeline, that through a combined less invasive Neuro-Visual approach, is able to reconstruct the car driver's physiological status. Specifically, the proposed contribution refers to the sampling and processing of the driver PhotoPlethysmoGraphic (PPG) signal. A parallel enhanced low frame-rate motion magnification algorithm is used to reconstruct such features of the driver's PhotoPlethysmoGraphic (PPG) data when that signal is no longer available from the native embedded sensor platform. A parallel monitoring of the driver's blood pressure levels from the PPG signal as well as the driver's eyes dynamics completes the reconstruction of the driver's physiological status. The proposed pipeline has been tested in one of the major investigated automotive scenarios i.e., the detection and monitoring of pedestrians while driving (pedestrian tracking). The collected performance results confirmed the effectiveness of the proposed approach.

Keywords: driver drowsiness monitoring, deep learning, pedestrian tracking, adas, photoplethysmographic

OPEN ACCESS

Edited by:
Antonio Fernández-Caballero, University of Castilla-La Mancha, Spain

Reviewed by:
Parampreet Kaur, Lovely Professional University, India
Stavros I. Dimitriadis, Greek Association of Alzheimer's Disease and Related Disorders, Greece

***Correspondence:**
Francesco Rundo
francesco.rundo@st.com

Received: *11 February 2021*
Accepted: *29 June 2021*
Published: *30 July 2021*

Citation:
Rundo F, Conoci S, Spampinato C, Leotta R, Trenta F and Battiato S (2021) Deep Neuro-Vision Embedded Architecture for Safety Assessment in Perceptive Advanced Driver Assistance Systems: The Pedestrian Tracking System Use-Case. Front. Neuroinform. 15:667008. doi: 10.3389/fninf.2021.667008

1. INTRODUCTION

Drowsiness is a symptom related to a lack of awareness which affects concentration, reaction time, and most seriously, safety (Schmidt, 1989; Rundo et al., 2019c). In the last few years, researchers discovered the mechanisms for which the level of attention is strongly correlated to the cardiac activity (Schmidt, 1989; Kurian et al., 2014; Rundo et al., 2019c).

The main functions of the heart are regulated by the Autonomic Nervous System (ANS). Specifically, the sympathetic and the parasympathetic nervous system, the two branches of ANS, are responsible for regulating many cardiac mechanisms, which are reflected in the attentional state of a subject (Schmidt, 1989). As stated previously, physiological signals represent a relevant data source to assess a subject's physiological condition (Kurian et al., 2014; Dastjerdi et al., 2017; Rundo et al., 2019d). The study of physiological signals have received much attention from the scientific community of the automotive industry (Kurian et al., 2014; Dastjerdi et al., 2017; Rundo et al., 2019d). Specifically, the growing proliferation of non-invasive medical devices to collect physiological parameters has led to the development of advanced new tools to be integrated into vehicle-environment. In this context, PhotoPlethysmoGraphic (PPG) signal has been proposed as a valid solution to analyze a subject's physiological status (Kurian et al., 2014; Dastjerdi et al., 2017; Rundo et al., 2019c,d). With the recent advances in safety awareness systems, the car manufacturers have spent a lot of efforts to develop innovative ADAS architectures based on PPG signal processing (Kurian et al., 2014; Rundo et al., 2019c). PPG is a convenient and simple physiological signal that provides information about the cardiac activity of a subject (Rundo et al., 2019d) and, therefore, the drowsiness status of a subject as well as pathologies which may indirectly have an impact on the subject's guidance[1]. In this work, we also focused on the use of the PPG signal for monitoring the subject's blood pressure. Several studies have pointed out that a robust driving risk assessment system leverages the so-called "driver fatigue condition" (which combines drowsiness and blood pressure monitoring) to perform a robust risk estimation (Husodo et al., 2018).

In this regard, the authors investigated promising solutions based on the use of a sensor framework to determine the level of driving safety through the driver's drowsiness assessment as well as the correlated blood pressure analysis (Littler et al., 1973; Husodo et al., 2018; Hui and Kan, 2019). Current solutions propose the use of wearable sensors or embedded devices equipped with such sensors in order to detect the first signs of fatigue[2] (i.e., a significant and progressive lowering of the body temperature, the heart rate, etc.). Therefore, they require the use of invasive methodologies for the car driver that are often not feasible in automotive applications (Littler et al., 1973; Husodo et al., 2018; Hui and Kan, 2019).

Several reports have highlighted that cardiovascular diseases could have a major impact on the health of a subject, affecting also the level of attention (Wu et al., 2015). In order to gather information about users' health conditions (and specifically the "car driver" users), major interventions have been made for monitoring and analyzing the systolic and diastolic pressure of both healthy and hypertensive subjects through physiological signals. Preliminary studies have confirmed that there is a strong correlation between cardiovascular risk, drowsiness estimation, pressure level quantification, and physiological PPG signal (Littler et al., 1973; Schmidt, 1989; Kurian et al., 2014; Wu et al., 2015; Dastjerdi et al., 2017; Husodo et al., 2018; Hui and Kan, 2019; Rundo et al., 2019c,d). Inspired by the recent research studies, we defined the proposed work, providing an effective solution to overcome the limitations of the PPG acquisition. Due to the high sensibility of the PPG signal to Motion Artifacts (MA) generated with the body movements, *ad-hoc* sensing device combined with an innovative processing workflow was used (Dastjerdi et al., 2017; Rundo et al., 2019d). Therefore, it is not always easy to use methods based on PPG sampling and processing. In automotive field, the sensors of the PPG signal are usually embedded in the driver's steering in order to detect the needed physiological data (from the car driver hand placed on the steering) for drowsiness monitoring.

In Kurian et al. (2014) and Rundo et al. (2019c), the authors of these contribution have developed and patented an application that allows to detect the PPG waveforms of the driver from his/her hand placed in the car steering. Anyway, it is needed that the car-driver applies the hand in the PPG sensing devices arranged in the steering wheel; otherwise, the PPG signal cannot be collected (Kurian et al., 2014; Rundo et al., 2019c,d). Similarly, it would be advantageous to have a non-invasive system for monitoring the driver's blood pressure from PPG signal without having to wear such medical devices. In order to address the aforementioned issues, the authors propose an innovative pipeline for estimating some features of the PPG signal of any subject without the use of invasive devices. This paper presents an efficient Deep Learning pipeline designed to perform a non-invasive PPG features reconstruction using an innovative low frame-rate Motion Magnification algorithm (Wu et al., 2012). Moreover, we developed a proper Application System Framework (ASF) to assess both the driver drowsiness trough the usage of PPG signal features and the correlated blood pressure level, providing a robust evaluation of the driving safety.

The authors have a lot investigated the mentioned issue related to the car driver drowsiness monitoring as reported in Rundo et al. (2020a), Rundo et al. (2020b), Battiato et al. (2020), and Rundo et al. (2020c). They performed in-depth studies in relation to the robustness and performance of the delivered solutions. This paper is arranged into five sections. In section 2, the PPG based theory was introduced with a special focus to the physical principle that characterizes its formation. In section 3, we presented the used PPG sensing device. In section 4, the main scientific literature contributions are reported and discussed. In section 5, the whole proposed pipeline is described while in

[1]https://www.who.int/gho/ncd/risk_factors/blood_pressure_prevalence_text/en/
[2]https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/ersosynthesis2015-fatigue25_en.pdf

section 6 both the experiments and validation results are reported and discussed.

2. THE PPG SIGNAL: UNDERLYING PHYSICAL PHENOMENA

PhotoPlethysmoGraphy is a non-invasive method for measuring blood flow in the cardiovascular circulatory system (Dastjerdi et al., 2017). More specifically, PPG is a physiological signal generally obtained through a combined optical-to-electric sensing system able to transduce the blood flow dynamic into electrical waveforms (Dastjerdi et al., 2017; Vinciguerra et al., 2019).

Through the mentioned optical sampling methods, we are able to collect the PPG signal by illuminating the region of interest of the subject's skin with a Light Emitting Diode (LED). The used LED is coupled with a photo-sensing device that is able to capture the part of emitted light (photons), which is not absorbed by the subject blood flow (Back-Scattered signal; Vinciguerra et al., 2019). In particular, the photo-sensing device captures the part of reflected LEDs emitted light which is not absorbed mainly by the oxygenated hemoglobin (HbO2) and in residual form by the non-oxygenated (Hb) present in the subject's blood (Conoci et al., 2018; Vinciguerra et al., 2019). An electrical transduction circuit of the optical signal will complete the typical PPG signal acquisition pipeline. In **Figure 1**, the overall representation of the PPG pattern formation is reported. It shows the physical phenomenon underlying the formation of the PPG signal. Through the action of the ANS the heart rate of the subject is regulated (Schmidt, 1989). This regulation produces an impact in the arterial blood flow of the i.e., in the back-scattering data linked to the dynamics of HbO2 and Hb (Schmidt, 1989; Kurian et al., 2014; Rundo et al., 2019c,d). This collected data will be properly electrically transduced and processed by means of an automotive grade microprocessor, described in the next paragraphs.

3. THE PROPOSED PPG SENSING SYSTEM

Different hardware architectures have been proposed in order to sample the PPG signal (Agrò et al., 2014; Liu et al., 2016; Conoci et al., 2018; Vinciguerra et al., 2019). The authors proposed the use of an hardware architecture based on the employment of a photo-multiplier silicon device called SiPM (SiliconPhotoMultiplier) (Mazzillo et al., 2009, 2018; Liu et al., 2016; Vinciguerra et al., 2017; Rundo et al., 2018a).

The designed sensing device is composed by two OSRAM LED emitters (SMD package) emitting at 850 nm. These leds are then used as optical light sources with a SiPM device (detector) which has a total area of $4.0 \times 4.5\ mm^2$ and 4,871 square microcells with 60 μm of the pitch. The proposed SiPM device has a geometrical fill factor of 67.4% and is packaged in a surface mount housing (SMD) of $5.1 \times 5.1\ mm^2$ total area. More details about the used hardware in Conoci et al. (2018), Agrò et al. (2014), Liu et al. (2016), and Mazzillo et al. (2009). Furthermore, to sample the PPG optical signal [through the embedded 12-bit Analogic to Digital Converter (ADC)] and to handle the implemented filtering and stabilization algorithms, the SPC5x 32-bit Chorus microcontroller was used (Vinciguerra et al., 2017; Mazzillo et al., 2018; Rundo et al., 2018a). **Figure 2** shows the proposed designed PPG sensing hardware platform. The so sampled PPG raw signal comprises a pulsatile ("AC") physiological signal, which is correlated to cardiac-synchronous changes in the blood volume, superimposed with a slowly varying ("DC") component containing lower frequency sub-signals, which is correlated to respiration, thermoregulation, and so on (Vinciguerra et al., 2017; Mazzillo et al., 2018; Rundo et al., 2018a). Usually, the sampled raw PPG signal contains various types of noise: electronic, motion artifacts, micro-movements due to breathing and so on (Mazzillo et al., 2009, 2018; Liu et al., 2016; Vinciguerra et al., 2017; Rundo et al., 2018a). For this reason, it is needed to filter the raw PPG signal to obtain only the "AC" part of our interest (as it is directly correlated with the subject's heart activity) de-noised from the various kinds of noise. **Figure 3A** shows an instance of raw noised PPG signal sampled from the hand of a recruited subject. To achieve this purpose, a raw PPG signal filtering pipeline was implemented by authors. More in detail, in the aforementioned pipeline a Butterworth bandpass filter in the 0.5–10 Hz range was used. The authors also implemented an innovative, robust stabilization, and de-noising pipeline (motion artifacts, micro-breathing movements, etc.) called PPG-PRS (Choi et al., 2016; Rundo et al., 2018a, 2021). In **Figure 3B**, we reported a detail of the so filtered PPG waveforms. As shown in **Figure 3B**, a filtered steady-stable PPG signal is obtained by the deep pipeline proposed by the authors and called PPG-PRS (PPG Pattern Recognition System) (Choi et al., 2016; Rundo et al., 2018a, 2021). The so generated signal contains PPG waveforms which are compliant with the standard for this type of signal (Vinciguerra et al., 2017; Mazzillo et al., 2018).

The PPG-PRS technique is based on the use of a bio-inspired "reaction-diffusion" mathematical model that characterizes the two phases of cardiac activity (specifically it has been hypothesized that the diastolic phase is combined with a "reaction" dynamic and systolic to a phenomenon of "diffusion"). Through a cross-correlation analysis between the sampled PPG signal (filtered in the range 0.5–10 Hz) and the standard compliant PPG signal generated by the aforementioned reaction-diffusion mathematical model, the authors are able to recognize the waveforms of the PPG signal which are compliant with the standard and consequently discard those affected by noise or artifacts. Further implementation details of the PPG-PRS stabilization and de-noising pipeline can be found in Rundo et al. (2018a, 2021).

As introduced in the previous sections, in order to make a non-invasive sampling of the driver's PPG signal, the PPG sensing devices were embedded in the car steering in different positions, specifically, those where statistically the driver places the hands while driving. In this way, each time the driver also places a single hand in the steering part in which the PPG sensing device was embedded, the described signal formation

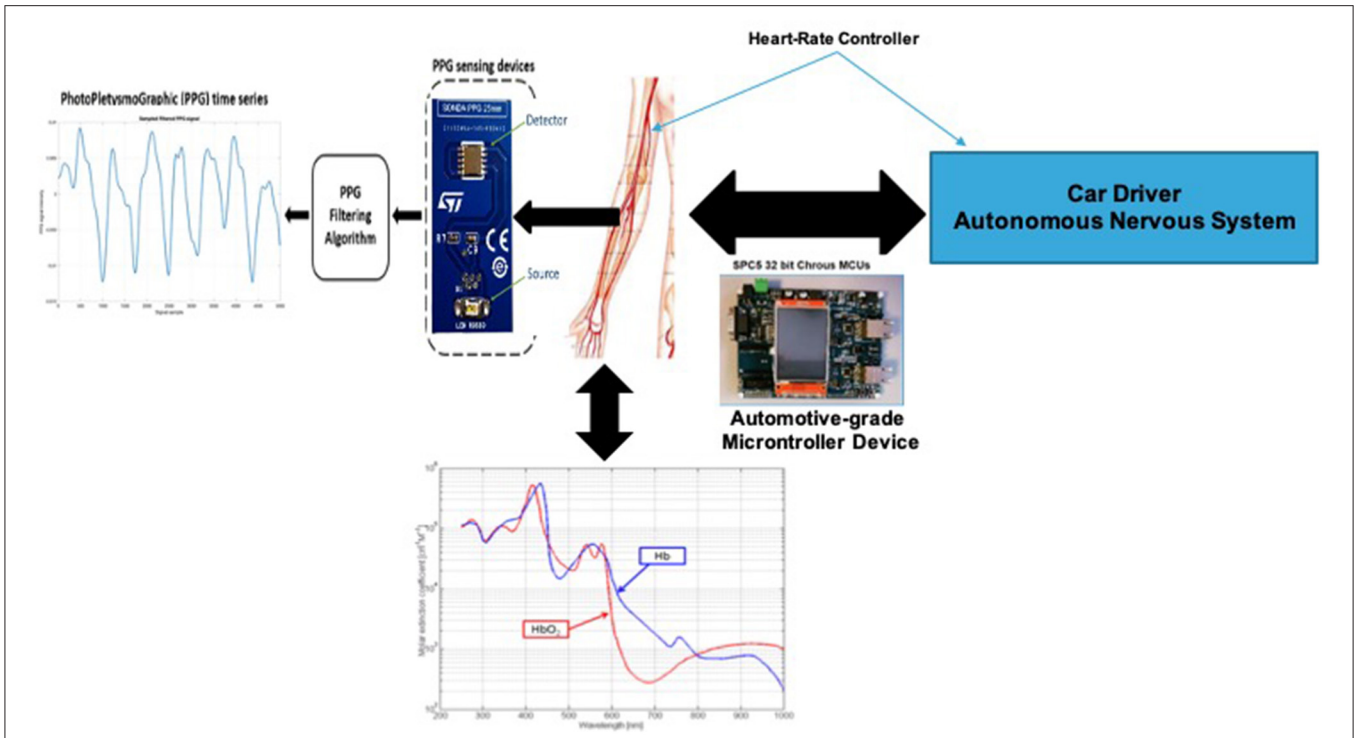


FIGURE 1 | The PPG signal formation diagram.

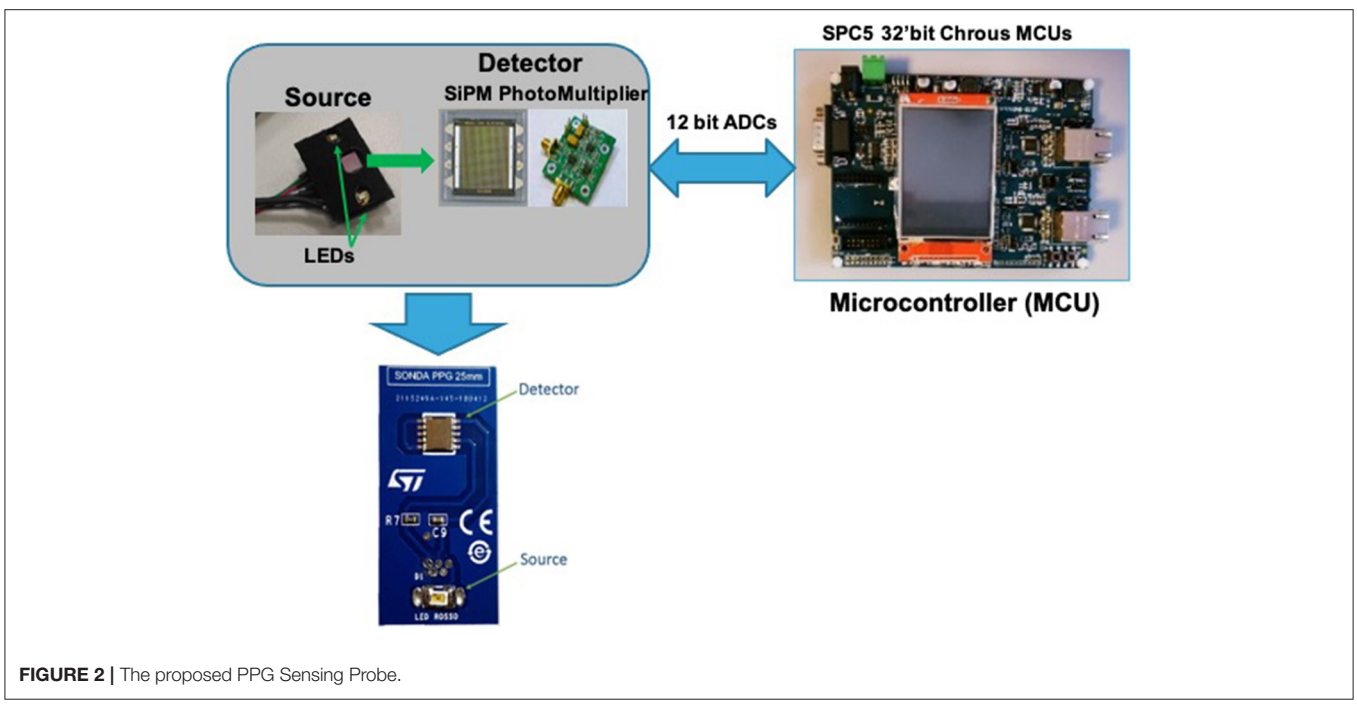


FIGURE 2 | The proposed PPG Sensing Probe.

mechanism will start working and the raw data will be sampled and then subsequently processed accordingly. In **Figure 4**, the overall scheme of the proposed PPG sampling pipeline is shown.

4. RELATED WORKS

Over the last decade, several researchers have investigated the issue that concern the evaluation of the driver's

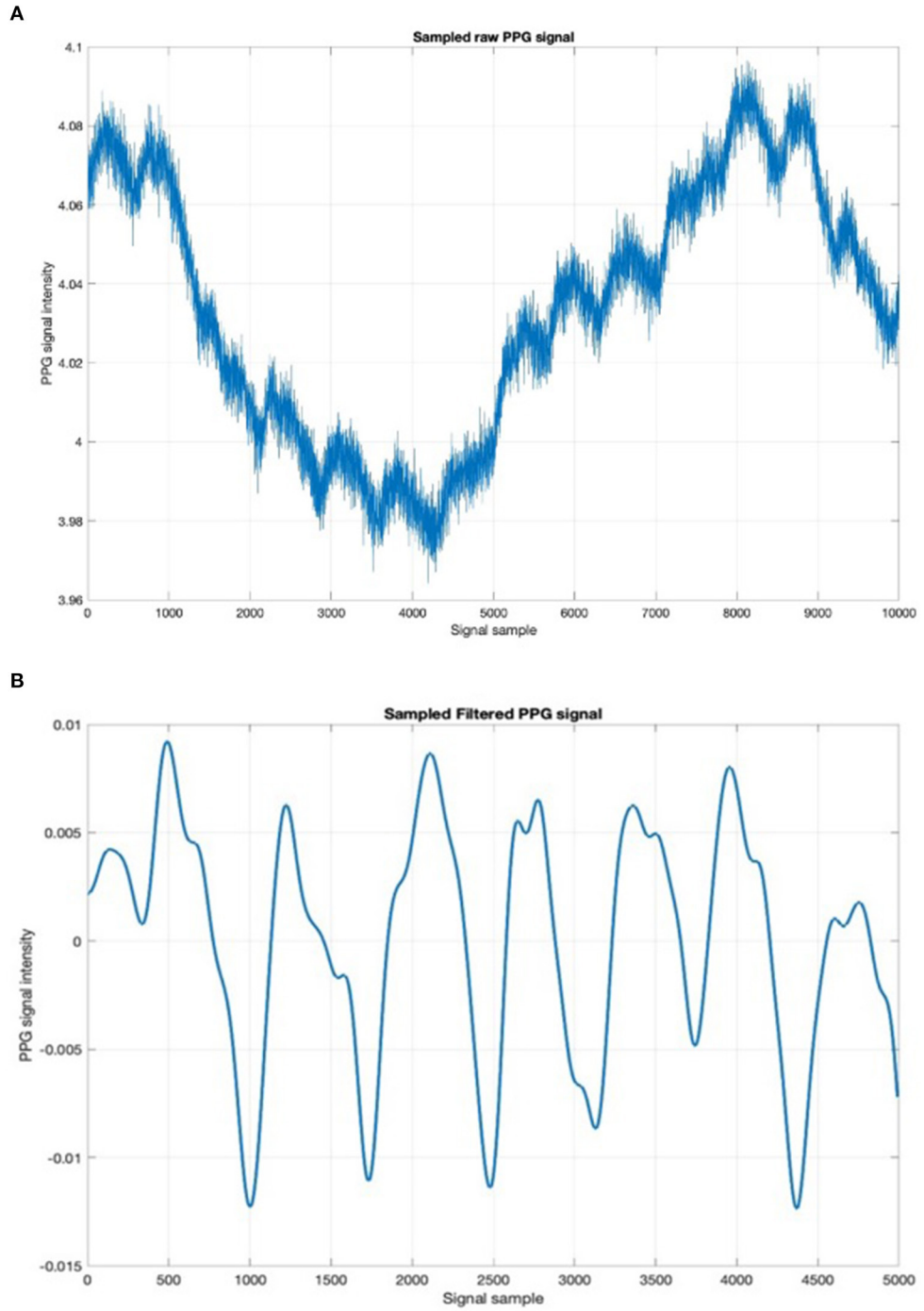


FIGURE 3 | (A) Sampled raw PPG time-series. **(B)** Detail of the filtered PPG signal.

status with the physiological signals (Choi et al., 2016). Typically, these solutions are computationally expensive and invasive (Vavrinskỳ et al., 2010). Moreover, the continuous monitoring of physiological parameters during car driving requires high-performance systems (both hardware and software) to ensure high levels of safety while driving (Vavrinskỳ et al., 2010; Choi et al., 2016; Rundo et al., 2018a, 2021).

In this context, the PPG signal is the easiest to acquire among all the physiological signals that can be sampled by the driver (Vavrinskỳ et al., 2010; Choi et al., 2016). By simply allocating sensing devices (LED + SiPM) in the car steering wheel, the raw

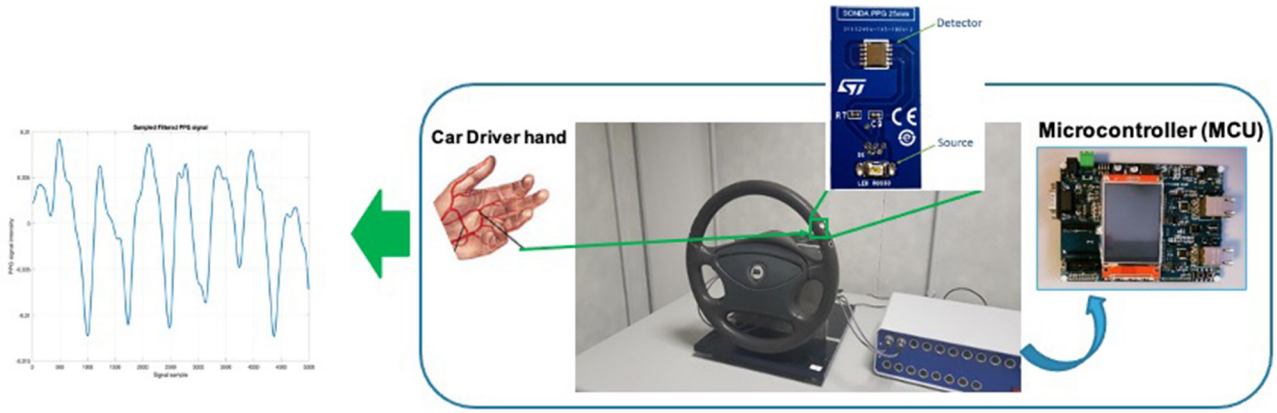


FIGURE 4 | The overall scheme of the proposed PPG sensing device embedded in the car steering.

PPG signal of the driver can be easily sampled using the physical mechanism described in the previous paragraphs. The sampled PPG signal is post-processed by the microcontroller devices placed on the car (ADAS framework) to monitor drowsiness (Vavrinskỳ et al., 2010; Choi et al., 2016; Koh et al., 2017). In the other hand, to obtain the ElectroCardioGraphic (ECG) signal of the car driver, at least three contact-points would be required (Einthoven triangle; Rundo et al., 2018a). Therefore, the use of the PPG signal is preferred.

Another reliable source of data (which can be sampled in a non-invasive way) is the visual information referring to the driver's face and related facial expressions that seems to be correlated to the level of attention (Vural et al., 2007). For these reasons, many visio-based and physio-based solutions have been designed and implemented. In Koh et al. (2017), the authors analyzed the pattern of Heart Rate Variability (HRV) to monitor the drowsiness of the car driver. The HRV is a simple indicator suitable to measure the variation in the time interval between consecutive heartbeats (in milliseconds; Vavrinskỳ et al., 2010; Choi et al., 2016; Koh et al., 2017; Rundo et al., 2021). Several studies confirmed the correlation between a subject's attention level and the dynamic of related HRV, specifically with the analysis of such indicators extracted from the HRV called HF (High-Frequency power), LF (Low-Frequency power), and the LF/ HF ratio (Vural et al., 2007; Lee et al., 2011, 2019; Choi et al., 2017; Koh et al., 2017; Deng and Wu, 2019).

In Lee et al. (2019), a Convolutional Neural Network was used to classify drowsy/wakeful status. The CNN was fed by analyzing three types of recurrence plots (RPs) derived from the R-R intervals. The authors found that the simultaneous use of ECG and PPG signal (needed to collect the HRV) inevitably introduced a non negligible noise and artifacts. Another solution related to assess car-driver fatigue is reported in Choi et al. (2017) in which the authors proposed a system to measure the car driver's emotional and physiological status coming from a wearable device placed on the wrist. As well as previous work, a pre-processing pipeline was developed in order to extract such valid time-series from the acquired signals. The so extracted features were classified by using a Support Vector Machine method. The results confirmed the effectiveness of the proposed solution in detecting and distinguishing the driver's physiological status. In Lee et al. (2011), the researchers proposed an innovative method to detect driver drowsiness combining Computer Vision and Image Processing approaches. Specifically, they evaluated the PPG signal waveform in order to record changes from awake to a drowsy state. At this stage, they detect the eye region through the use of template matching in combination with a Genetic algorithm to analyze the eye's behavior. In order to derive the final classification, PPG drowsy signals are evaluated with eye motion behavior to provide more robust and effective results. Deng and Wu (2019) proposed a novel approach, called DriCare, to estimate driver drowsiness by using a face-tracking algorithm. The authors designed a new method to individuate 68 key points in facial regions with the aim of evaluating drivers' fatigue status. Another promising work that investigates the problem of assessing car driver's fatigue state is Jabbar et al. (2020) in which the authors developed a Convolutional Neural Networks to classify drowsiness status. Specifically, the authors used facial landmarks as input data for their proposed model. The main contribution of their work is the development of a deep learning-based system that can be easily integrated into a car environment. As introduced in this scientific contribution, several studies investigated the driving safety assessment, in addition to the detection of the level of car driver attention as well as to monitor driver blood pressure. In the scientific literature many approaches deals with the arterial stiffness or the estimation of blood pressure with the advantage of Deep Learning (DL) methods and PPG signal based analysis. In Monte-Moreno (2011), the authors used a photoplethysmography sensor in order to estimate the diastolic and systolic blood pressure in a non-invasive way. In particular, the PPG waveform was used to gather the features used as input data in a various machine learning algorithm to be able to estimate the systolic (SBP) and diastolic (DBP) blood pressure and the blood glucose level (BGL). The results confirmed that Random Forest achieved better prediction estimation confirming the relationship between

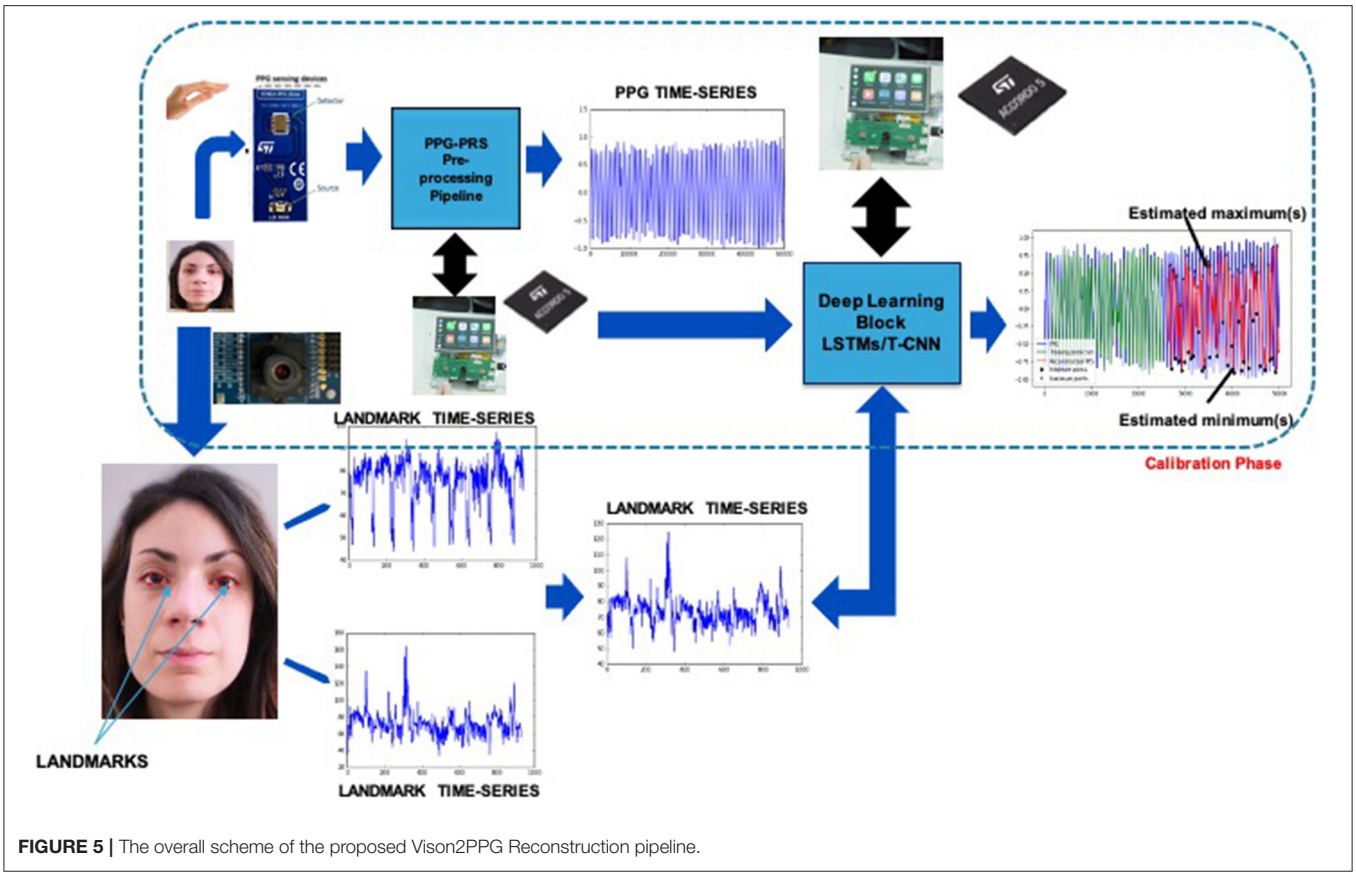


FIGURE 5 | The overall scheme of the proposed Vison2PPG Reconstruction pipeline.

the shape of PPG waveform and blood pressure level. The effectiveness of Machine Learning (ML)-based techniques have been furtherly confirmed by experimental results that showed the relationship there is among the PPG waveform, blood pressure and glucose levels. Slapničar et al. (2019) investigated the problem of detecting Blood Pressure (BP) using an ML-based architecture. The authors fed a novel spectro-temporal Deep Neural Network (DNN) with the PPG and its first and second derivative to be able to overcome limitations that concern the cuff-based devices. The ability to compute dependency between PPG waveforms and blood pressure and the effectiveness of the proposed model was confirmed by means of leave-one subject-out experiments. In Alty et al. (2007), a pipeline to predict arterial stiffness (an indicator correlated to subject blood pressure) was proposed by the authors. Moreover, with the purpose of examining cardiovascular disease they performed classification of subjects into high and low aortic pulse wave velocity (PWV) classes. The collected results confirmed the effectiveness of the proposed Support Vector Machine (SVM) based solution. In Rundo et al. (2018b), a novel approach was described by the authors in order to estimate cardiovascular disease risk by means blood pressure. The method reported in Rundo et al. (2018b) measures the subject blood pressure by analyzing the correlated PPG signal. In Huynh et al. (2018), the authors used the averaging Impedance Plethysmography (IPG) for the detection of Pulse Transit Time (PTT) in order to estimate the blood pressure. The tests showed that the estimation of blood pressure (BP) achieved interesting results (RMSE: 8.47 ± 0.91 and 5.02 ± 0.73 mmHg for systolic and diastolic levels, respectively). On the other hand, the previous approaches needs the use of invasive medical devices and sometimes require the need to sample the ECG signal in addition to the PPG and therefore impracticable in the automotive application.

5. METHODS AND MATERIALS

5.1. The Driving Safety Assessment Through Physiological Driver Analysis

Recent studies have highlighted the need to assess the car driver's physiological state in order to create highly safety automotive-grade applications. As previously reported, the PPG is a less-invasive signal suitable to provide useful information about the physiological condition of the driver. In fact, the main limitation of the existing solutions consists in the integration of sensors inside the vehicle to acquire the physiological signals. Most of the solutions propose a PPG signal sampling methodology based on the usage of such sensors placed on the car steering wheel. However, this implies that the driver has to maintain an unnatural behavior i.e., the hands constantly over the sensors. Moreover, if these sensors no longer work while driving, the classic pipelines would not be able to collect the signal, representing a serious risk during the real-time safety assessment

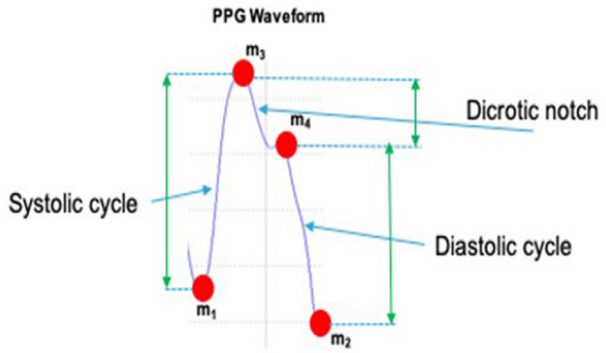


FIGURE 6 | A filtered compliant PPG waveform with a detail of such extreme points m1, m2, m3, m4.

(Kurian et al., 2014; Rundo et al., 2018b, 2021). In **Figure 4**, we schematized the PPG sampling pipeline embedding the sensors placed on the steering wheel.

With this regard, we designed an innovative pipeline which combines an enhanced version of the Motion Magnification analysis and Deep Learning approach for a non-invasive processing of the car driver PPG signal. As a result, the proposed approach is able to overcome the aforementioned critical issues. More specifically, we developed an innovative module called Vision2PPG Reconstruction Pipeline that reconstructs the features of the driver's PPG signal from a visual data. Through the use of a video-camera placed on the car dashboard, we recorded the driver's face, tracking the facial movements over the frames. Finally, by *ad-hoc* processing of these visual data we developed *ad-hoc* algorithm to reconstruct such features of the car driver's PPG signal even in absence of native sensing data. The use of Deep Architectures properly trained will complete the proposed pipeline.

5.2. The Vision2PPG Reconstruction Pipeline

In this section, we go through the details of the developed Vision2PPG pipeline. The proposed pipeline consists of a PPG sensing device designed to perform a preliminary system calibration and a camera device that captures the dynamic of the driver's frontal face. The collected data have been used to extract facial descriptors (landmarks). The designed module is based on Computer Vision techniques which cover specific requirements regarding the automotive certification ASIL-x (Dastjerdi et al., 2017; Vinciguerra et al., 2019). A deep learning algorithm (embedded in an ASIL-B microcontroller as firmware) completes the proposed pipeline by correlating the subject's face descriptors with the corresponding PPG waveforms. The overall flowchart of the proposed Vision2PPG pipeline is shown in **Figure 5**. A detailed description of the pipeline will be reported in the following paragraphs. As PPG sensors, we implemented the sensing system design described in the previous section. The designed four PPG sensing probes will be embedded in the driver's steering wheel equidistant from each other. The designed LEDs will emit at the wavelength of 850 nm (near infrared). The first phase of the proposed pipeline is the training-calibration task. The training-calibration phase includes a learning stage in which the pipeline determines the correlation between visual time-dynamic of the segmented face descriptors (landmarks) with the associated PPG features. This module was designed to collect enough data to characterize the car driver drowsiness as well as the correlated blood pressure level. At this stage, the raw PPG signal is firstly sampled by using the designed coupled LED-SiPM sensing probes (Mazzillo et al., 2018; Rundo et al., 2018a). We also applied the PPG-PRS (means PPG Pattern Recognition System) algorithm to the so collected raw PPG signal in order to obtain a compliant filtered PPG time-series (Rundo et al., 2018a, 2021).

The PPG-PRS preliminary filters the raw PPG signal cutting off artifacts and noise through the use of a Butterworth bandpass filter in the range 0.5–10 Hz. Moreover, a study of the first and second derivatives of the filtered PPG signal was implemented in the PPG-PRS algorithm. This analysis allows to detect the minimum and maximum extreme points of each selected compliant PPG waveform. Finally, by means of a second order dynamic and Reaction-Diffusion system, which emulates the physiological phenomenon of the PPG signal formation, the PPG-PRS algorithm is able to stabilize the sampled PPG signal. The configuration of the PPG-PRS algorithm is exactly the one reported in the scientific contribution (Rundo et al., 2018a). In **Figure 6**, we depicted an instance of a filtered PPG compliant waveform, identifying the corresponding extreme points m_1, m_2, m_3, m_4. The second part of the proposed Vision2Physio reconstruction pipeline is composed by the enhanced motion magnification module. More in detail, we performed the PPG sampling simultaneously with the recording of a video sequence of the subject's frontal face by using a low frame-rate camera device under normal light conditions. Specifically, we used a device with a max resolution of 2.3 Mpx and 50 fps as framerate. Several studies have demonstrated that the face of a subject performs visual micro-movements closely related to the cardiac pumping activity (systolic and diastolic phase; Oh et al., 2018). The PPG signal (as a cardiac-related signal) is strongly correlated to the aforementioned micro-movements (Balakrishnan et al., 2013). In order to make these micro-movements visible at naked eyes, some authors have designed innovative motion magnification techniques which require a video-camera devices with high frame-rate (on average, frame rate ranging from 10 Kfps up; Rubinstein et al., 2013). Motion Magnification refers to amplifying facial micro-movements in order to reveal the flow of blood (Balakrishnan et al., 2013; Rubinstein et al., 2013; Oh et al., 2018). However, the method originally proposed for motion magnification showed an evident issue in relation to automotive applications as it requires, as mentioned, the need for a high frame-rate video device (of the order of Kfps). This constraint is not easily covered in automotive field mainly for reasons of costs and sustainability of the underlying hardware. For this reason, by extending a previous version already implemented, the authors propose in this contribution a different motion magnification method which addresses the mentioned issues. The preliminar version

of the proposed enhanced motion magnification algorithm was reported by the authors in Trenta et al. (2019). With this regard, we firstly developed a method to process video frames depicting face sequences of the car driver in order to preliminary identify significant landmarks or descriptors on the subject's face. The proposed pipeline is shown in **Figure 7**.

In order to detect the aforementioned landmarks, the approach implemented in Kazemi and Sullivan (2014) was used. This algorithm locates the (x, y) coordinates of 68 facial points (landmarks) describing the facial structures (i.e., mouth, eyes, nose, etc.) on a subject's face image. The approach is based on the usage of a pre-trained learning model (Kazemi and Sullivan, 2014). The main advantage of this algorithm consists in obtaining near real-time high-quality landmarks recognition and tracking even with a low frame-rates video. More details in Kazemi and Sullivan (2014). In our previous solution reported in Trenta et al. (2019). both the hardware PPG signal sensing system and the landmarks detection algorithm were quite different. The PPG sensing system used green LEDs. Furthermore, although Kazami and Sullivan's algorithm was used to retrieve the landmarks of the driver's face, the number of used landmarks were significantly lower than the total number (68 descriptors), specifically, there were only two (landmarks adjacent to the eyes). For the reconstruction of the PPG features from the visual subject's face frames, a deep classifier based on LSTM was used. However, this model only reconstructed the minimum points of each PPG waveform. The setup reported in Trenta et al. (2019) allowed us to prove that the motion magnification model we proposed was able to achieve optimal performance in the automotive field. However, the pipeline described in Trenta et al. (2019) was implemented in a prototype system by National Instruments that required data buffering and therefore the overall system response time was significantly slowed down. Furthermore, considering that the pipeline used only two car driver landmarks adjacent to the eyes, in certain scenarios this visual data could be no longer available (for instance a scenario in which the driver wearing sunglasses) and therefore the pipeline would be no longer applicable. These limitations have been largely overcome in the motion magnification pipeline herein proposed. Formally, we have defined the reconstruction of landmarks dynamics by means of Kazemi and Sullivan based function $\psi_{KS}(.)$. If we indicate with $I_{t_k}(x, y)$ the captured $M \times N$ gray-level (or luminance gray-level channel in case of color camera device) video frame of the car's driver at the instant t_k, the $i-th$ dynamic landmark $\ell_i(t_k, x_l, y_l)$ is reconstructed as follow:

$$\ell_i(t_k, x_{l_i}, y_{l_i}) = \psi_{KS}(I_{t_k}(x, y)); \; k = 1..N_f; \; i = 1..N_L \qquad (1)$$

where $\ell_i(t_k, x_{li}, y_{li})$ represents the pixel intensity variation of the i-th landmark identified at the space position (x_{li}, y_{li}) on the frame while N_f represents the number of captured frames and N_L represents the number of identified landmarks (i.e., 68 as per Kazemi et al. based algorithm). Therefore, in the proposed pipeline the whole set of landmarks was used making the proposed method more robust than the version reported in Trenta et al. (2019). We did a test in a scenario where the driver is wearing sunglasses. While the Trenta et al. (2019) method is no longer applicable, our approach continues to work by presenting overlapping performances (see experimental results session for more details). Furthermore, it must be said that the Trenta et al. (2019) method is not applicable to the reconstruction of the driver's pressure level (see next paragraphs) due to the reduced number of landmarks. Moreover, the usage of a PPG sensing device emitting at near-infrared spectrum allowed to have a native PPG signal more detailed than that obtained using green light (as in Trenta et al., 2019) and that is for some specific characteristics implicit in the physiological process of the signal formation (Schmidt, 1989; Kurian et al., 2014; Dastjerdi et al., 2017; Conoci et al., 2018; Rundo et al., 2019d; Vinciguerra et al., 2019).

Once the representative landmark dynamics of the driver have been identified, we proceed analysing the so collected descriptors time-series $\ell_i(t_k, x_{li}, y_{li})$ in order to correlate their intensity temporal dynamics with the underlying cardiac activity. The proposed method does not require high frame-rate camera devices as for the method to which it is inspired (Littler et al., 1973). It requires a normal commercial vision device having a framerate in the range $\geq$40 fps. To model the aforementioned relationship between face landmarks and the cardiac activity of the analyzed driver, the authors propose two deep learning frameworks: one based on the usage of Deep Long Short-Term Memory (D-LSTM) architectures and the other one based on the usage of 1D Temporal Deep Dilated Convolution Neural Network (1D-TDCNN) network. Moreover, in order to improve the robustness of the proposed safety assessment pipeline, we introduce a classical Deep Convolutional Neural Network (D-CNN) for car driver's eyes tracking to be correlated with level of attention or drowsiness. In the following sections, some details of the proposed framework will be outlined.

5.3. The Deep Learning Framework

As described in the previous section, the proposed Deep Learning framework consisted of two parts: (i) a deep architecture (D-LSTM or 1D-TDCNN) employed to classify driver's PPG signal or in case it is no longer available, the driver landmarks dynamics $\ell_i(t_k, x_{li}, y_{li})$ retrieved from the Vision2PPG Reconstruction pipeline, (ii) a Depp CNN used to perform eyes-based visual classification of the car-driver drowsiness. In the following subsections, the authors proceed to the scientific description of each block of the proposed Deep Learning pipeline.

5.3.1. The Car Driver Landmarks Deep Classifier

As introduced, we propose two deep learning basic solutions to correctly classify the driver visual landmarks and correlate them with such PPG features (specifically: the extreme points of the PPG signal). The first setup uses *ad-hoc* Deep Long Short-Term Memory (D-LSTM) framework. In particular, our D-LSTM network is based on Vanilla architecture, firstly proposed by Hochreiter and Schmidhuber (1997). As mentioned, the authors have already used such LSTM-based architecture to address similar automotive application with respect to what it is herein described (Trenta et al., 2019). Moreover, vanilla D-LSTM architectures have been largely employed in the automotive

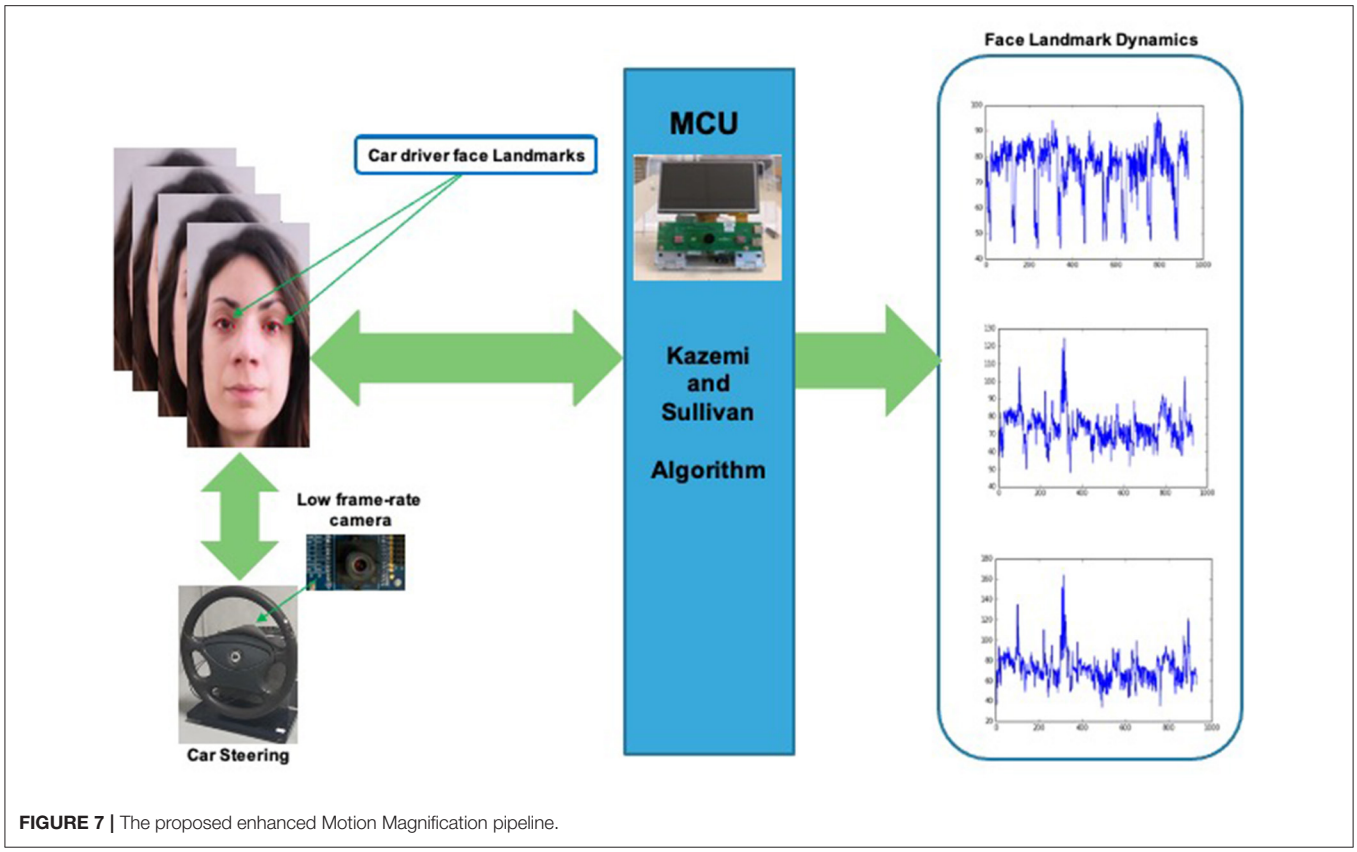


FIGURE 7 | The proposed enhanced Motion Magnification pipeline.

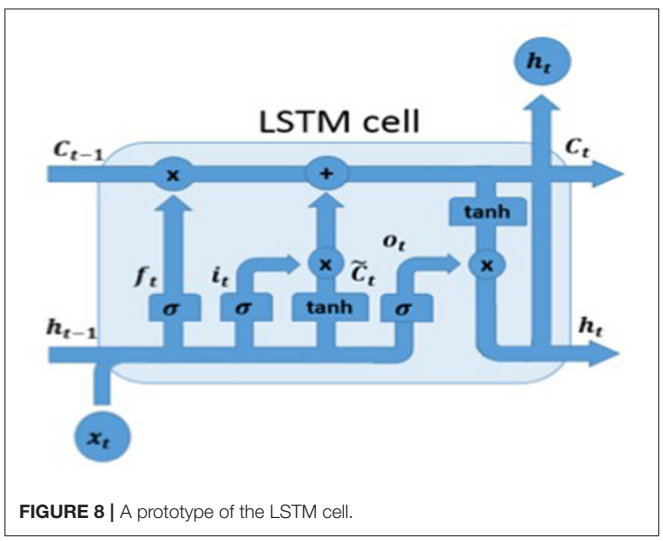


FIGURE 8 | A prototype of the LSTM cell.

field (Monte-Moreno, 2011; Koh et al., 2017; Vinciguerra et al., 2017). We remark that in this contribution, the authors significantly improved the approach previously described in Trenta et al. (2019). Specifically, the proposed approach allows to better reconstruct the PPG signal (more extreme points) and consequently improves the performance in terms of drowsiness classification. Furthermore, the proposed approach allows to estimate the corresponding level of blood pressure. In the next paragraphs, more details about the performance and benchmarking of the proposed approach. Let's introduce the designed deep platform. The proposed D-LSTM architecture is composed of one input layer, two hidden layers and one output layer. In order to significantly improve the classification performance, the input layer was designed with 64 input units and, consequently, the two hidden layers with 64 and 128 cells, respectively. Finally, the output layer, that consists in 1 cell, will provide the predicted PPG samples that will be used to determine the extreme points m_1, m_2, m_3, m_4. In **Figure 8**, we reported the basic unit structure of the used D-LSTM. The mathematical model which represents the learning model of the cell is reported as follow:

$$f_t = \sigma(W_f \bullet [h_{t-1}, x_t] + b_f \tag{2}$$

$$i_t = \sigma(W_i \bullet [h_{t-1}, x_t] + b_i \tag{3}$$

$$\widetilde{C}_t = \tanh(W_C \bullet [h_{t-1}, x_t] + b_C) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{5}$$

$$o_t = \sigma(W_o \bullet [h_{t-1}, x_t] + b_o \tag{6}$$

$$h_t = o_t * \tanh(C_t) \tag{7}$$

From **Figure 8**, it is clear that given x_t as input vector, the previous output cell h_{t-1} as well as the previous cell memory C_{t-1}, the current cell output h_t and the current cell memory C_t, are suitable to determine what information the D-LSTM has to be stored. Equations (4)–(7) allows to model the behavior of the proposed D-LSTM. In order to improve the overall performance, a batch normalization and dropout layer have been added to the output of each LSTM layer. Our model was trained with an initial learning rate of 10^{-3}, a batch size set to 512 and with the maximum number of training epochs set to 200. During the training-calibration phase of each recruited driver subject, the suggested Deep LSTM is able to learn the correlation there is among the selected facial landmark time-series $\ell_i(t_k, x_{li}, y_{li})$ and the corresponding sampled PPG signal. The designed Deep LSTM pipeline will produce in output the predicted PPG waveforms from which the extreme points m_1, m_2, m_3, m_4 will be detected as reported in **Figure 6**. After several tests we found that not all the identified landmarks are correlated with the subject's cardiac activity and then in order to obtain acceptable performances, in terms of reconstruction of the PPG signal features, only a subset of landmarks can be processed. Furthermore, thanks to the implemented D-LSTM architecture we have verified that a more simple composite landmarks signal can be used for addressing the needed PPG reconstruction task. Specifically, we perform the following computation:

$$\mu_\ell(t_k) = \frac{1}{N_L} \sum_{j=1}^{N_L} \ell_j \left(t_k, x_{l_j}, y_{l_j} \right) \quad \forall\, t_k \tag{8}$$

The so computed (properly normalized) signal $\mu_\ell(t_k)$ will be given as input of the designed D-LSTM. Consequently, the architecture will be trained to find the correlation between the input signal $\mu_\ell(t_k)$ and the corresponding PPG signal. **Figure 9** shows the overall scheme of the proposed D-LSTM. Basically, the extreme points m_1, m_2, m_3, m_4 for each predicted PPG waveforms (see **Figure 6** for more details) are computed by performing the same analysis applied by the mentioned PPG-PRS algorithm (Rundo et al., 2018b) i.e., by analyzing the first and second derivative of the D-LSTM estimated PPG signal.

The second analyzed deep architecture that we have implemented to correlate the visual features and the corresponding PPG signal, is based on the use of a temporal deep architecture. Specifically, *ad-hoc* 1D Temporal Deep dilated Convolutional Neural Network (1D-TDCNN) has been developed (Zhao et al., 2019). The main building block consists of a dilated causal convolution layer that operates over the time steps of each sequence (Zhao et al., 2019). The proposed 1D-TDCNN includes multiple residual blocks, each containing two sets of dilated causal convolution layers with the same dilation factor, followed by normalization, ReLU activation, and spatial dropout layers. Furthermore, a 1×1 convolution is applied to adapt the number of channels between the input and output. Specifically, we implemented a 1D-TDCNN composed of 25 blocks with a downstream SoftMax layer. Each of the deep blocks comprise a dilated convolution layer with 3×3 kernel filters, a spatial dropout layer, another dilated convolution layer, a ReLU layer, and a final spatial dropout. The dilation factor size starts of a factor equal to 2 and it will be increased for each block till the value of 16. As for LSTM based solutions, the so-designed 1D-TDCNN output represents predicted extreme points of the PPG signal. The following **Figure 10** shows the proposed 1D-TDCNN based solution.

To validate the effectiveness of the proposed Deep pipelines in reconstructing such features of the original PPG signal, we computed the Fast Fourier Transform (FFT) of a signal obtained through the difference between the PPG reconstructed minimum points (Trenta et al., 2019). As described in the introductory part of this contribution, the HRV monitoring is one of the classic structured physiological method for determining the drowsiness level of a subject. As widely confirmed in scientific literature, the HRV can be computed performing a proper Fast Fourier Transform (FFT) of a differential physiological signal obtained from ECG i.e., by means of the distance R-Peak to R-Peak as well as from PPG through the distance of the minimum points of two consecutives waveforms (Lee et al., 2011; Rundo et al., 2018b; Trenta et al., 2019). Therefore, a well robust measurement of HRV of a subject (car driver in our application) can be obtained by computing the FFT spectrum of the PPG based physiological differential signal. In case of the PPG signal is unavailable the predicted PPG minimum points reconstructed by Vision2PPG block will be used. Once the HRV is computed, by means of classical analysis based on the usage of the HF (High-Frequency power), LF (Low-Frequency power) and the LF/ HF ratio, the drowsiness of a subject is easily detected and monitored (Lee et al., 2011, 2019; Choi et al., 2017; Deng and Wu, 2019). More details about the performance validation of the proposed deep pipelines will be reported in the experimental results section.

As soon as the proposed Deep architecture -both D-LSTM either 1D-TDCNN- has learned the correlation there is among the driver's facial landmarks and extreme points of the relative PPG signal, the training-calibration phase will be dropped, and therefore the system will work feed-forward. The calibration phase of the whole described pipeline requires a 15/20 s of PPG signal (the proposed sensing device is able to execute an acquisition at 1 KHz) with the relative visual frames (acquired at 50 fps) and it will be needed to perform only at once. Obviously, this system will be enabled by the driver hardware control unit whenever the physiological signal of the driver will not be detected for some reasons by the PPG sensors embedded in the car's steering. Therefore, compared to the previous approach proposed by the authors in the contributions (Rundo et al., 2018a,b, 2019d; Trenta et al., 2019), the method herein described allows to reconstruct more discriminating PPG features (the four extremal points of the PPG waveform against only the minimums of the previous version). Furthermore, the Vision2PPG pipeline with the downstream classifier is much more robust than in scenarios where some visual landmarks are no longer available. By using a 1D TDCNN, the long-range temporal dependencies of the PPG signal are better treated allowing an effective changes detection in the state of attention induced by the autonomic nervous system of the driver. In addition, the proposed pipeline

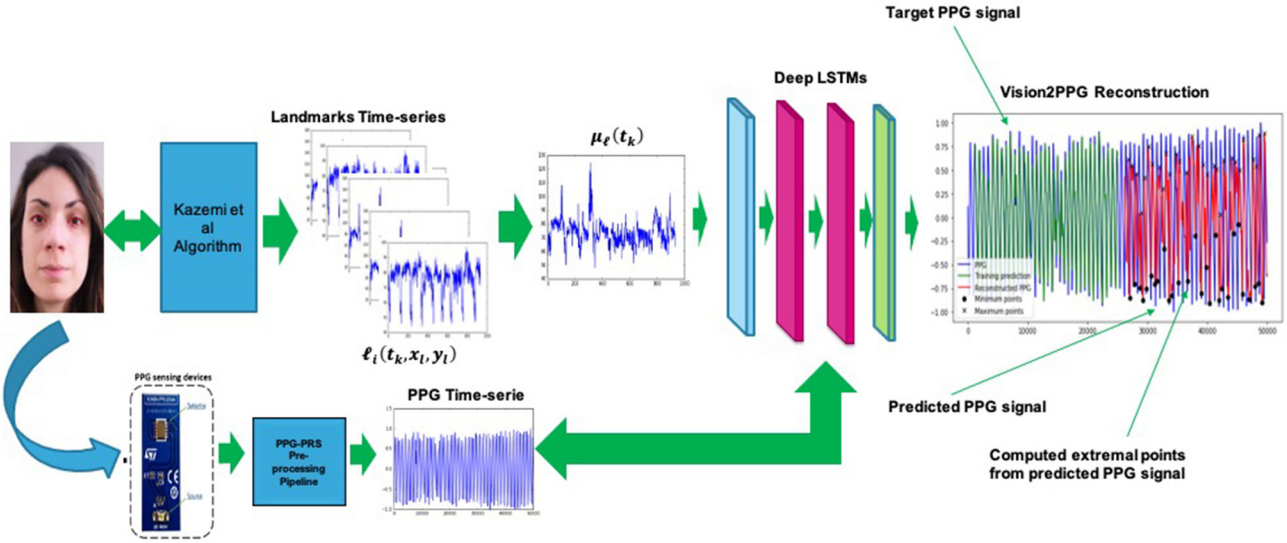


FIGURE 9 | The proposed D-LSTM pipeline for PPG features estimation. The corresponding extreme points are highlighted with black circle and "x" symbol in the predicted PPG waveforms (red signal).

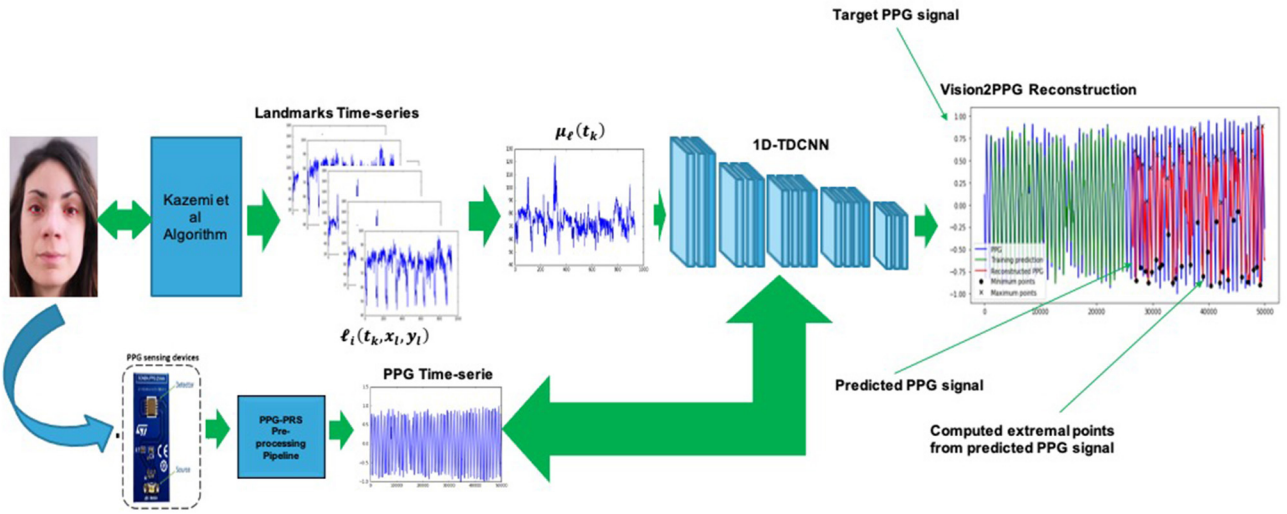


FIGURE 10 | The proposed 1D-TDCNN pipeline for PPG features estimation. The corresponding extreme points are highlighted with black circle and "x" symbol in the predicted PPG waveforms (red signal).

provides features that allow to also obtain a less invasive and cuff-less assessment of the driver's pressure level not possible in the solutions previously proposed in Trenta et al. (2019).

5.3.2. The Car Driver Eye's Tracking Trough Deep CNN Pipeline

In order to have a simultaneous safety monitoring pipeline without technology overlap with the previously described blocks, the authors implemented a further assisted pipeline based on the usage of *ad-hoc* Deep Convolutional Neural Network (D-CNN). The proposed 2D D-CNN network is quite simple and includes three convolutional layers each of which has a ReLU activation layer (with batch-normalization), such 2×2 max pooling layers except for the last convolutional layer. The first convolutional layer performs 32 operations with 3×3 kernel filters, where the second and the third shows 64 and 128 kernel filters of 3×3, respectively. A stack of densely connected layers and Softmax complete the proposed D-CNN pipeline in order to perform two classes classification of the input visual data i.e., drowsy/wakeful status of the analyzed driver. Fine-tuning is done for 100 epochs using Adam optimizer and cross entropy as loss function. To carry out the experiments, we also set the

learning rate to 0.001 and the batch size to 32. In **Figure 11**, the proposed D-CNN architecture. The input visual frame of the D-CNN is a patch depicting a single eye (77 × 77 resolution), which we segment from the driver's face using the algorithm proposed in Viola and Jones (2001). This algorithm does not require an annotated images dataset to work and can it be easily carried on embedded hardware architectures while maintaining remarkable performance both in segmentation and in execution speed. More details in Viola and Jones (2001). The proposed full D-CNN has been designed in order to get it portable to such embedded hardware solution ASIL-B certified i.e., STA1295 Accordo5 MCU[3] as described in this work. Specifically, the D-CNN (as well as the previous D-LSTM/1D-TDCNN pipelines) will be hosted in the STA1295 Accordo5 embedded automotive grade DUAL ARM A7 which includes a 3D-GFX accelerator cell suitable for this kind of processing (Rundo et al., 2019a). In **Figure 12**, we reported a scheme of the overall platform. All input data (PPG samples and visual frames) will be properly stored in such buffers allocated in the EMI memory of the STA1295 MCU platform.

5.4. The Car-Driver Blood Pressure Estimation Pipeline

The application described in this section regards the monitoring of the driver blood pressure as strongly correlated to the driving safety and driver attention level. We propose a novel solution for measuring blood pressure from such features of the PPG signal as well as from the corresponding facial video frames of the subject by using the proposed Vision2PPG reconstruction pipeline (in case the PPG signal unavailability). We remark that our proposed blood pressure estimation pipeline works both in the scenario in which the PPG is unavailable and in the case in which the physiological signal was correctly sampled by the sensing probe. As described, by means of the Vision2PPG pipeline, we will obtain the extreme points of the subject's PPG signal. From these predicted data, by means of a suitably configured Shallow Neural Network (Rundo et al., 2019c), the authors are able to discriminate normal blood pressure subjects from those who had pressure values beyond the norm. The blood pressure estimation pipeline has been designed in order to be ported as firmware running in the STA1295 Accordo5 embedded automotive processor. Once the set of characteristic extreme points of the PPG waveforms have been collected (by means of the Vision2PPG reconstruction pipeline or from native PPG signal), we are able to characterize the subject's cardiac activity (systole, dicrotic, and diastole phases) on which the level depends the blood pressure (Rundo et al., 2018b). Let's formalize this application. In **Figure 6**, we have reported an instance of compliant PPG waveform with highlighted extreme points m_x. For each pair of PPG waveforms PPG^j, PPG^{j+1} we define the following indicators:

$$\varphi = [m_1^j, m_2^j, m_3^j, m_4^j, dx_i^j, dy_i^j, mAI^J] \forall\, j = 1..(N^{PPG} - 1) \; ; i = 1, 2, 3, 4 \quad (9)$$

$$m_i^j = (x_{m_i^j}, y_{m_i^j}) \,, \forall\, j = 1..(N^{PPG} - 1) \; ; i = 1, 2, 3, 4 \quad (10)$$

$$dx_i^j = x_{m_i^{j+1}} - x_{m_i^j}, \;\; \forall\, j = 1..(N^{PPG} - 1) \; ; i = 1, 2, 3, 4 \quad (11)$$

$$dy_i^j = y_{m_i^{j+1}} - y_{m_i^j} \forall\, j = 1..(N^{PPG} - 1) \; ; i = 1, 2, 3, 4 \quad (12)$$

$$mAI^j = ((y_{m_3^j} - y_{m_1^j}) - y_{m_4^j})/(y_{m_3^j} - y_{m_1^j}) \quad (13)$$

where mAI^J is a modified version of the so-called Augmentation Index, usually computed for measuring the arterial stiffness (Vavrinskỳ et al., 2010) while N^{PPG} represents the number of estimated PPG waveforms. The other indicators reported in the Equations (9)–(12) are able to characterize cardiac cycles and, therefore, the relative blood pressure level. The input of the above-mentioned Shallow Neural Network (SNN) is represented by the elements of the vector φ. The SNN is a Fully Connected Multi-Layers Network with two hidden layers of 500 and 300 neurons and a binary output. It was designed with the target of learning the correlation there is among the so computed input elements in φ and the associated diastolic and systolic blood pressure values. Furthermore, the aforementioned SNN was trained with the Scaled Conjugate Gradient backpropagation (SCG) algorithm (Rundo et al., 2019c). As output, the SNN framework will produce a value that may be treated like a discriminating flag able to differentiate the subject showing normal pressure values (0) with respect to the hypertensive or hypotensive subject (1). Specifically, we calibrate the system in order to detect if the subject has a pressure level that is on average 15% higher or lower than a reference value. The set 120/80, which indicates 120 mmHg for systolic pressure and 80 mmHg for diastolic pressure, has been considered normal blood pressure values (as confirmed by the team of physiologists who assisted us in this study). On the other hand, higher or significantly lower (15%) values are considered anomalous. It should be noted that the proposed system is able to monitor and discriminate even different pressure levels (with respect to the classic 120/80 mmHg) requiring a different and adequate calibration. This is to cover the cases of moderate hypotension/hypertension affected subjects who physiologically present a normal blood pressure level a little different from 120/80 mmHg. For these subjects, the pipeline calibration phase will refer to different reference pressure values. In any case, during the calibration phase of the Vision2PPG recognition system, the training of the SNN block will be performed in order to correlate the blood pressure (reference value and current value) and PPG levels of the subject preparing to drive. The proposed pipeline has been tested and validated in one of the most interesting automotive scenario: the pedestrian tracking system.

5.5. The Deep Network With Criss-Cross Attention for Pedestrian Tracking System

The tracking of pedestrians while driving is certainly one of the most important aspects in the field of safety automotive requirements. The detection and subsequent monitoring of

[3]https://www.st.com/en/automotive-infotainment-and-telematics/sta1295.html

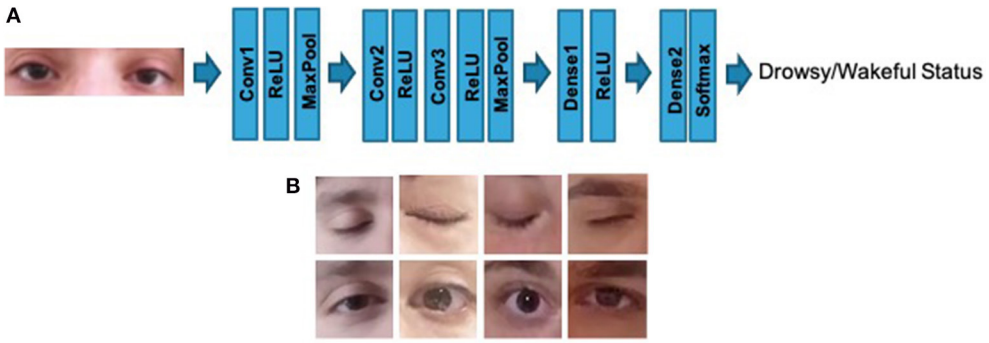


FIGURE 11 | (A) The proposed D-CNN architecture. **(B)** Some instances of D-CNN input visual data.

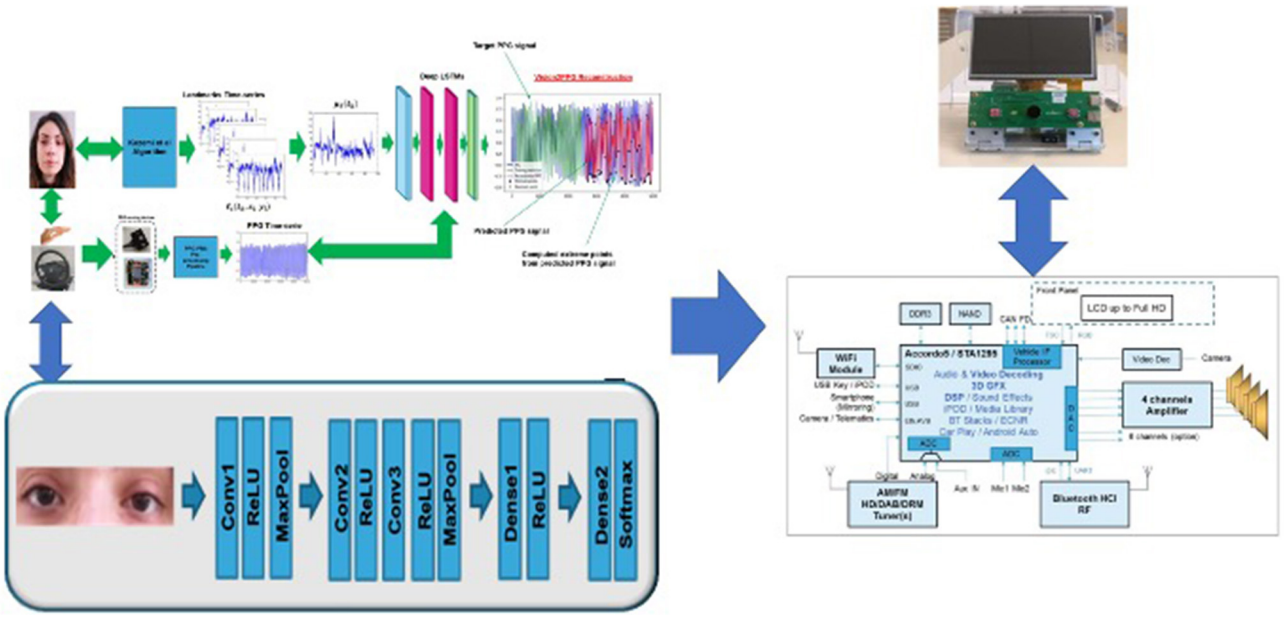


FIGURE 12 | The STA1295A Accordo5 platform hosting the D-LSTM Vison2PPG Reconstruction pipeline as well as D-CNN architecture.

pedestrians in the driving scene allows the automatic driver assistance system to continuously validate if the driving dynamics and the level of attention are compatible with the presence of pedestrians in the scene. Many authors have investigated this relevant issue by analyzing the advantages inherent in the use of deep learning architectures (Tian et al., 2015; Song et al., 2018; Jeon et al., 2019; Bhola et al., 2020). The authors investigated several interesting object detection and tracking architecture backbones to be adapted to pedestrian tracking. Specifically, the researchers found that deep learning systems embedding attention mechanisms significantly increase the performance in classification and segmentation of the underlying backbones. For these reasons, we found it useful to implement an innovative network that included the recent self-attention context through the use of Criss-Cross layers (Huang et al., 2019). More in detail, for each source image/feature pixel, an innovative Criss-Cross attention module computes the contextual information of all the correlated pixels on its Criss-Cross path. This attention pre-processing combined with further recurrent operations allow the Criss-Cross method to leverage the full-image dependencies during the learning session of the deep network (Huang et al., 2019). Let us formalize the attention processing embedded in the Criss-Cross layer. Given a local feature map $H \in R^{C\times W\times H}$ where C is the original number of channels while $W \times H$ represents the spatial size of the generated feature map trough a Deep Convolutional Network. The Criss-Cross layer applies two preliminary 1×1 convolutional layers on H in order to generate two feature maps F_1 and F_2, which belong to $R^{C'\times W\times H}$ and in which C' represents the reduced number of channels due to dimension reduction with respect to original (C). Let define an *Affinity* function suitable to generate the Attention-Map $A_M \in R^{(H+W-1)\times(W\times H)}$. The affinity operation is so defined. For each

position u in the spatial dimension of F_1, we extract a vector $F_{1,u} \in R^C$. Similarly, we define the set $\Omega_u \in R^{(H+W-1)\times C}$ by extracting feature vectors from F_2 at the same position u, so that, $\Omega_{i,u} \in R^{C'}$ is the i-th element of Ω_u. Taking into account the above operations, we can define the introduced Affinity operation as follows:

$$\delta^A_{i,u} = F_{1,u}\Omega^T_{i,u} \tag{14}$$

where $\delta^A_{i,u} \in D$ is the affinity potential i.e. the degree of correlation between features $F_{1,u}$ and $\Omega_{i,u}$, for each i = [1,...,H + W - 1], and $D \in R^{(H+W-1)\times(W\times H)}$. Then, we apply a softmax layer on D over the channel dimension to calculate the attention map A_M. Finally, another convolutional layer with 1×1 kernel will be applied on the feature map H to generate the re-mapped feature $\vartheta \in R^{C\times W\times H}$ to be used for spatial adaptation. At each position u in the spatial dimension of ϑ, we can define a vector $\vartheta_u \in R^C$ and a set $\Phi_u \in R^{(H+W-1)\times C}$. The set Φ_u is a collection of feature vectors in ϑ having the same row or column with position u. At the end, the final contextual information will be obtained by an *Aggregation* operation defined as follows:

$$H'_u = \sum_{i=0}^{H+W-1} A_M^{i,u}\Phi_{i,u} + H_u \tag{15}$$

where H'_u is a feature vector in $H' \in R^{C\times W\times H}$ at position u while $A_M^{i,u}$ is a scalar value at channel i and position u in A_M. The so defined contextual information H'_u is then added to the given local feature H to augment the pixel-wise representation and aggregating context information according to the spatial attention map A_M. These feature representations achieve mutual gains and are more robust for semantic segmentation. Anyway, the criss-cross attention module is able to capture contextual information in horizontal and vertical directions but the connections between one pixel and its around is not processed. To overcome this issue, the authors firstly introduced the Criss-Cross methodology proposed a Recurrent Criss-Cross processing in which classic Criss-Cross operations can be unrolled into R loops. We defined $R = 2$ for our purpose as suggested by the original description (Huang et al., 2019).

The proposed Criss-Cross layer has been embedded in the Mask-R-CNN architecture used as pedestrian detection, segmentation and tracking. The architecture of Mask-R-CNN has been descripted in He et al. (2017). The Mask-R-CNN architecture extends previous detection and segmentation similar solutions such as Fast-R-CNN or Faster R-CNN by adding a branch-pipeline for predicting an object mask in parallel with the existing pipeline for bounding box recognition (He et al., 2017). One of the main parts of the Mask-R-CNN architecture is the Deep convolutional network used for the feature maps extraction. For this purpose, different backbones have been tested in the scientific literature (He et al., 2017, 2020). We denote the used backbone architecture using the nomenclature *network-depth-features* (He et al., 2017, 2020). In this proposed work, we used the Mask-R-CNN with a backbone based on a 2D ResNet-101 (He et al., 2017, 2020) in which we embedded, in the last block, a layer of self attention based on the aforementioned Criss-Cross methodology. The following **Figure 13** shows the overall scheme of the proposed Criss-Cross enhanced Mask-R-CNN and some instances of the detected and segmented pedestrians both with red bounding-boxes and without. This architecture performed very well as we reached a test-set performance mIoU of 0.695 over CamVid dataset which is in line with the performance of other more complex architectures (He et al., 2017, 2020). Obviously, it is reiterated that the target of the work herein described in this manuscript is not to propose an architecture that outperforms the others in relation to the detection, tracking and segmentation of pedestrians but rather a system that detects the level of driving safety in risky scenarios. This proposed enhanced Criss-Cross architecture has been used as it presents an excellent trade-off between segmentation performance and complexity for an embedded system (as mentioned, this pipeline is being ported over the STA1295 MCU system). Furthermore, the Mask-R-CNN also allows us to obtain the bounding-box of the pedestrian which we will need to determine the distance from the driver's car. Quite simply, the height and width of the segmentation bounding box of each segmented pedestrian will be determined. Only bounding-boxes that have at least one of the two dimensions greater than two heuristically fixed thresholds (L_1 and L_2, respectively for length and width) will be considered salient pedestrians, i.e., pedestrians that must be considered by the driver when choosing the driving dynamics. The other pedestrians will be considered non salient and therefore not involved in safety level assessment. This so computed distance assessment will be used in the next block of the proposed pipeline.

5.6. The Driving Safety Monitoring System

We proposed an innovative pipeline for monitoring the car driving safety by means of visual and physiological data analysis. Specifically, the designed pipeline is able to combine the driver's physiological and visual data sampled in *ad-hoc* implemented sensing framework embedded in the car. We developed a system that was non-invasive for the driver, addressing the classic critical issues based on the usage of such physiological bio-signal difficulty to sample in a vehicle (ECG or EEG) or which require such sensors to be worn by the driver. For these reasons we have implemented a solution based on the usage of less-invasive car driver's PPG signal processing. When that PPG signal is no longer available for some reason, a parallel pipeline based on the usage of car driver visual data will be able to reconstruct specific features of the missed driver's PPG signal. From these estimated features we can reconstruct the level of driving safety by monitoring the driver's fatigue level i.e., a degree of attention through the analysis of Heart Rate Variability (HRV) and tracing the correlated blood pressure dynamics. Moreover, in order to increase the robustness of the proposed approach, a further visual driver face processing has been implemented trough *ad-hoc* D-CNN which will provide a classification of the car driver attention.

The proposed application use-case is mainly aimed at driving in risky conditions, for instance, in scenarios in which one or more pedestrians are moving in the driving scene. By means of a modified Mask-R-CNN network with an attention layer through Criss-Cross methodology, we are able to detect and segment

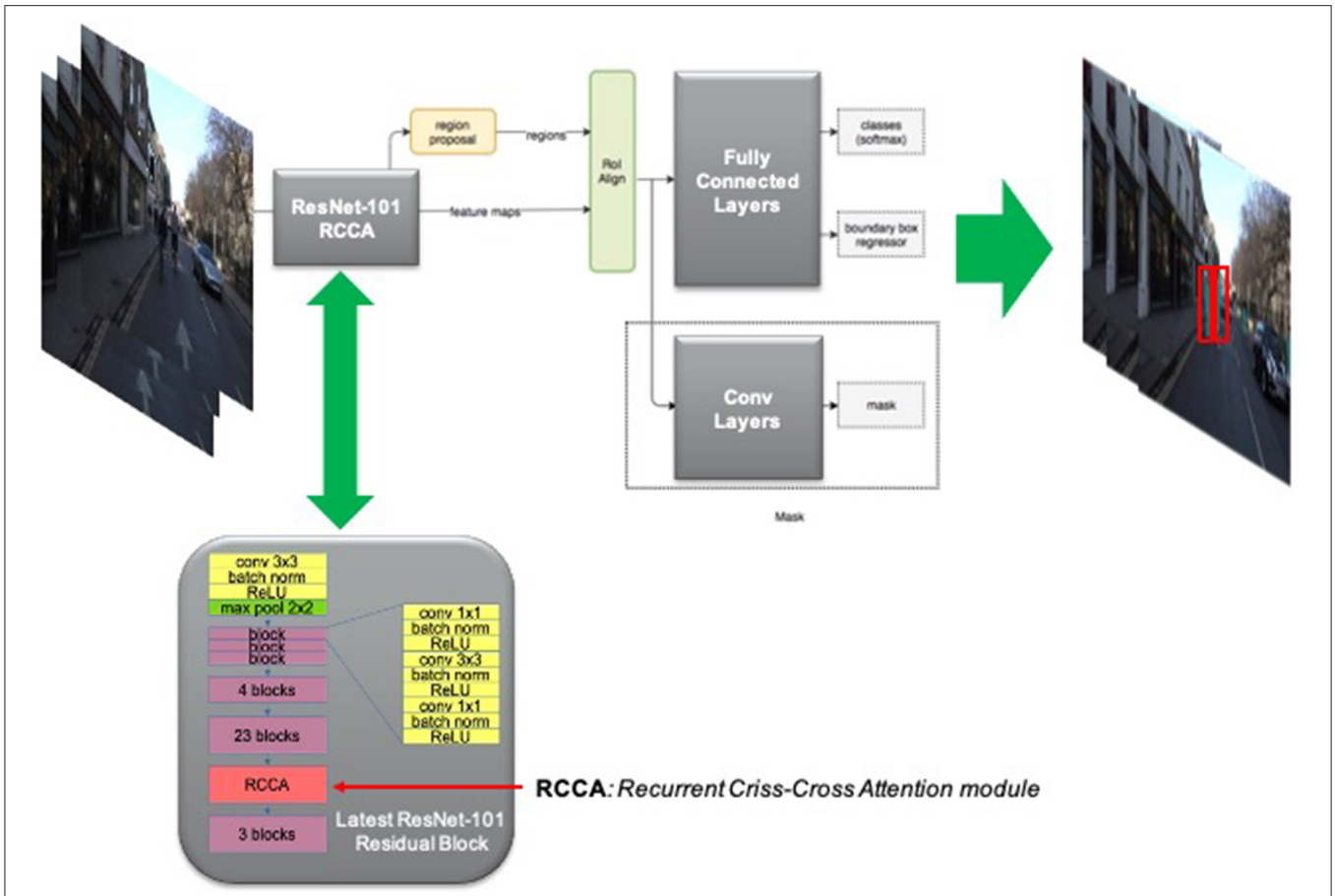


FIGURE 13 | The proposed Pedestrian Tracking System based on Criss-Cross enhanced Mask-R-CNN.

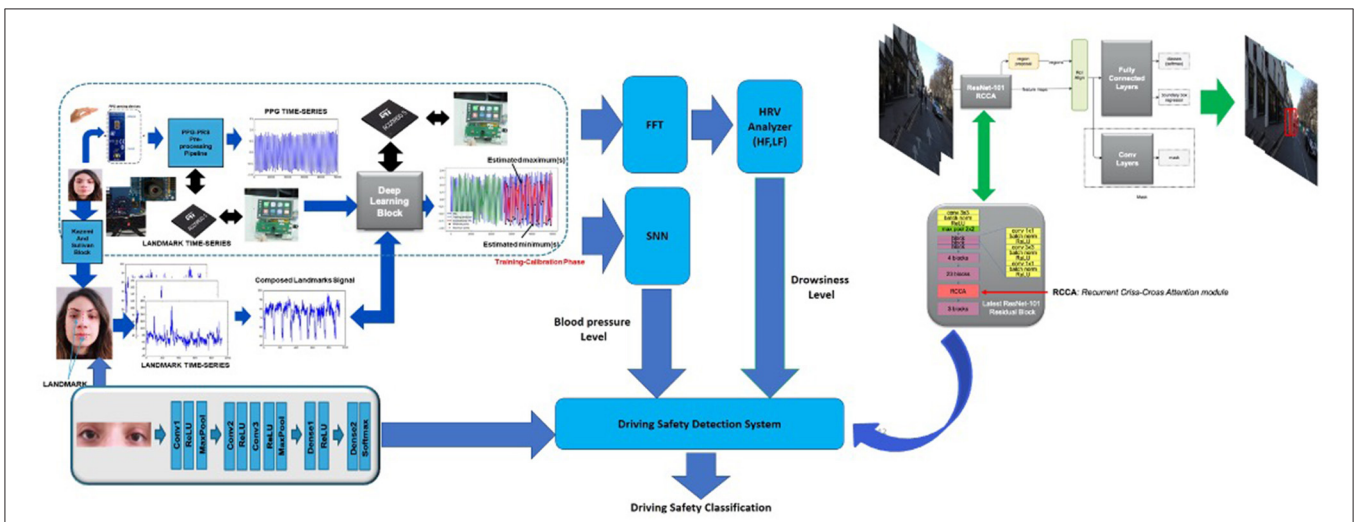


FIGURE 14 | The overall scheme of the proposed solution with a detail of the designed Driving Safety Detection System.

pedestrians in different configurations. Furthermore, we are able to obtain bounding-box segmentation and tracking which will allow us to estimate the distance of the pedestrian from the car. A comprehensive scheme of the proposed driving safety estimation architecture is showed in **Figure 14**. In this scheme is highlighted a block called Driving Safety Detection System (DSDS) which will process the outputs obtained by each of the previously described pipelines, specifically, the assessment of the driver's attention level (HRV analysis), the driver's blood pressure level (SNN output), the classification of the driver attention level performed by the implemented D-CNN and the distance of the detected and segmented pedestrians from the car. In detail, the DSDS will trigger an acoustic alert-signal, with different intensity according to the risk level, if one of the following setup will come true:

- High Risk Level (Alert-Signal with High intensity) HRV shows low attention level *AND* (the SNN output shows abnormal blood pressure *OR* D-CNN shows low attention level) *AND* the Mask-R-CNN identifies such salient pedestrians;
- Medium-Low Risk Level (Alert-Signal with Medium-Low intensity) HRV shows low attention level *OR* D-CNN shows low attention level *AND* the Mask-R-CNN identifies such salient pedestrians;

The acoustic signal is managed by the STA1295A Accordo5 Audio sub-system which will host the DSDS software implementation (see **Figure 12**). The system will therefore provide an assessment of the driving safety in the analyzed use-case: pedestrian tracking. All the developed sub-systems of the proposed pipeline are ongoing to be ported over the STA1295 Accordo5 Dual ARM Cortex MCUs platform.

5.7. Dataset

Under the supervision of physiologists, a dataset of selected and monitored subjects was collected. More in details, for each recruited subject we performed an acquisition of the PPG signal simultaneously with systolic and diastolic blood pressure measurements and contextually with a session of face video-recording. All data collections were conducted by inducing in both subjects such states of high attention and states of low attention that the physiologists who supervised the clinical study properly induced. More in details, for each recruited subject the EGG signal was sampled as well as the EEG time-series from which the physiologist analyze the dynamic: alpha waves representative of low attention status or beta waves representative of medium-high attention level (Guo and Markoni, 2019; Rundo et al., 2019b). Participants were recruited after signing the informed consent form provided by the Ethical Committee CT1 (authorization n.113/2018/PO). The dataset (further recruitments with respect to the work described in Rundo et al., 2018b, 2019b) is composed of 71 subjects (males and females, min age: 20 years, max age: 75 years) splitted into 45 subjects having a less or equal to 120/80 mmHg and 26 subjects having a blood pressure higher than 120/80 mmHg both in physiological and pathological state.

To carry out our experiments, we paid attention to the subjects with an arterial pressure greater or lower than 15% in average with respect to normal values (configured during the training-calibration phase of the Vision2PPG reconstruction pipeline for each subject). In case of hypotensive or hypertensive subjects, we have adjusted the reference blood pressure accordingly. The overall study was carried out in accordance with the protocol of the Declaration of Helsinki. For each subject, the blood pressure measurements were certified by means of a medical sphygmomanometer device. The PPG was sampled by the proposed sensing device at 1 kHz frequency. For retrieving visual data, we have used a commercial color camera device having max resolution of 2.3 Mpx and 50 fps as frame-rate. The collected minimum pressure value is around 98/70, while the maximum pressure value is 158/90. Each measurement session lasts 10 min, 5 of which in a state of high attention and 5 in a state of low attention. The level of adequate attention has been induced by performing mathematical calculations, reading anecdotes or by viewing representative driving scenarios where high driver attention is required (car overtaking, changing lane, braking, etc.). During this phase the subjects' EEG signal was sampled, confirming the presence of beta waves (Rundo et al., 2019b). Similarly, for low attention measurement sessions in which subjects were asked to relax, close their eyes for a few moments, listen slow music that induces states of relaxation, thus inducing a decrease in heart rate and the simultaneous presence of typical visual expressions showing drop in attention such as the decrease in the frequency of eye blinking. Only when the EEG showed the dynamic of alpha waves, the data (PPG and visual) were acquired so as to have certainty of the low attention state of the analyzed subject. Same approach for high attention states corresponding to beta waves. We divided the dataset as follows: 70% of the data has been used for the training while the remaining 30% for validation (15%) and testing (15%). We have run our experiments as well as training and testing of the proposed deep learning architectures in MATLAB full toolboxes rel. 2019b environment running in a server having an Intel 16-Cores and NVIDIA GeForce RTX 2080 GPU.

6. RESULTS

In order to validate the proposed composed approach we have tested each of the proposed blocks i.e., the PPG-based HRV driver drowsiness detection as well as the driver blood pressure level estimation and finally the visual drowsiness estimation based on D-CNN processing and the pedestrian detection, segmentation and tracking. About the proposed PPG based HRV based drowsiness detection pipeline, the following **Table 1** reports the overall testing accuracy of the approach as well as accuracy for each classified class.

A description of the hardware and software system implemented to obtain these performances is reported. Four PPG sensing probes have been embedded in the car steering with the characteristics described in the introductory section of this paper having near infrared emission LEDs at 850 nm. In addition, a camera with characteristics of 50 fps and a maximum resolution of 2.3 Mpx has been mounted in the base of the steering wheel and directed to sample the driver's face. As

TABLE 1 | Car driver HRV drowsiness monitoring performance.

Method	Overall accuracy (%)	Class 1 drowsy driver accuracy (%)	Class 2 driver accuracy (%)
Proposed LSTM	95.07	95.77	94.36
Proposed 1D-TDCNN	95.67	96.33	94.78
Trenta et al. (2019) with LSTM	94.43	95.10	93.77
Rundo et al. (2019c)	93.66	95.77	91.54
MLP	87.94	87.55	88.33
SVM	88.05	86.98	89.11
ResNet-50	93.90	93.85	93.95

TABLE 2 | Car driver's face Landmarks detection methods: Robustness comparison.

Method	Overall accuracy car driver NS (%)	Overall accuracy car driver WS (%)	Class 2 accuracy car Driver WE (%)
Trenta et al. (2019) with LSTM	94.43	0.00	70.16
Proposed 1D-TCNN	95.67	92.31	93.98

(NS, normal scenario; WS, wearing sunglasses; WE, wearing eyeglasses).

TABLE 3 | Blood pressure (BP) performance.

Method	Overall accuracy (%)	Class 1 normal BP accuracy (%)	Class 2 abnormal BP accuracy (%)
Proposed	90.14	88.88	92.30
Rundo et al. (2019a)	88.73	91.11	84.61
Trenta et al. (2019) with LSTM	<50	<50	< 50
MLP	89.47	87.62	91.33
SVM	82.24	81.98	82.05
ResNet-50	90,01	89.02	91.01

described in the Vision2PPG reconstruction section, the vision system has been calibrated to reconstruct the PPG features processed as per Equations (9)–(13). The driver's PPG signal data sampled by the sensing probes will be then converted by the analog-to-digital converters (ADCs) embedded in the SPC5x Chorus MCU used to acquire and pre-process the raw PPG signal from the sensors. Furthermore, the so collected raw PPG signal will be stabilized and filtered in the range 0.5–10 Hz by means of the previously introduced PPG-PRS pipeline running as firmware in the mentioned embedded MCU. The visual frames of the Vision2PPG block will instead be stored in the DRAM of the embedded system based on STA1295A Accordo5 MCU in which the landmarks detection system based on the Kazemi and Sullivan approach is running. Therefore, the results reported in **Table 1** were obtained by creating an application use-case lasting 45 minutes of acquisition of different scenarios, specifically ones in which the PPG signal was available (the driver placed the hand over the PPG probes embedding on the steering) and another ones in which this PPG signal was no longer available and therefore the Vision2PPG reconstruction block start the reconstruction of the PPG features. The PPG signal is sampled at a frequency of 1 Khz. The used ADCs embedded in the SPC5x CHORUS have a resolution of 12 bits. The set of PPG features have been reconstructed by Vision2PPG block through the usage of both D-LSTM (three layers of 64,64,128 cells trained with an initial learning rate of 10-3, a batch size set to 512 and with the maximum number of training epochs set to 200) and 1D-TCNN architectures (25 blocks -convolutions, normalization layer, spatial dropout, ReLU, residual block-, kernel 3×3, dilated convolutional with factor starting from 2 to 16, mini Batch Size of 1; initial learning rate equal to 0.001; and a Dropout Factor rated of 0.1). A comparison with similar approaches based on different methodologies have been reported in **Table 1**. More specifically, we have compared our architecture with our previous solution described in Rundo et al. (2019c) and Trenta et al. (2019) as well with an approach based on the usage of Support Vector Machine (SVM), Multi-layer Perceptron (MLP having an hidden layer of 600 neurons and trained with a more performer Levenberg-Marquardt algorithm) and finally with a ResNet-50, arranging the input landmarks in a 224×224 matrix. As highlighted from the reported benchmark comparison results, the method we propose and based on the

use of deep 1D-TCNN network outperforms the others not only in terms of accuracy but rather in robustness as can also be seen from **Table 2**. Specifically, the method we propose presents a noteworthy advantages in terms of robustness compared to the previous version (Trenta et al., 2019) in real scenarios in which the driver wears sunglasses or eyeglasses (highly probable scenarios). Furthermore, in terms of computational complexity and therefore implementation costs, our solution based on 1D-TDCNN networks shows a lower complexity than the architectures that show comparable performances: ResNet-50 has a size of 98 Mb against the 42 Mb of our 1D-TDCNN. This aspect is very important considering that this solution must be applied in the automotive field over the embedded platforms mentioned in the previous sections, confirming an undoubted advantage in terms of sustainability costs with greater performance in terms of accuracy. The accuracy reported in **Table 1** was obtained by analyzing the HRV associated with the features of the PPG signal thus processed, by analyzing the HF, LF, and HF/LF ratio indicators (see section 5.3.1). We remark that most significant advantage of the proposed method is non-invasiveness in that it does not require the driver to wear any sensor nor does it requires the driver necessarily places the hand on the part of the steering in which the PPG sensors are housed. Similarly, in the same dataset, we validated the reconstruction of the driver's blood pressure level through the SNN network. For this testing session, we used the same aforementioned hardware setup of the PPG sensing probes. The collected PPG features (both native from sensors and reconstructed from Vision2PPG block) will be fed to the implemented SNN Fully Connected Multi-Layers Network having 64 neurons as input layer joined with two hidden layers of 500 and 300 neurons and a binary output. The network was trained with Scaled Conjugate Gradient backpropagation (SCG) algorithm and compared with other methodologies such as MLP, SVM, ResNet-50, and previous method described in Trenta et al. (2019). The validation scenario consists in constantly monitoring the subject's blood pressure level (by means of a digital blood pressure detection device embedded in the driver's arm) and detecting the classification reconstructed by our SNN output system (output in the range 0–0.5 means pressure within the norm while values above 0.5 indicate abnormal blood pressure of the driver). Once again, an examination of the benchmark comparison values reported in **Table 3** confirms the evident advantages of the proposed pipeline as it not only outperforms in accuracy our previous solution reported in Rundo et al. (2019a) and the other methodologies tested but, in terms of complexity is significantly lower than a solution based on deep learning (ResNet-50). Furthermore, the pipeline herein described allows to overcome the raised limit of the previous version reported in Trenta et al. (2019) which had no significant accuracy in the classification of the subject's blood pressure level.

A comparison with another approach (Rundo et al., 2019c) is reported. As evident from comparison data reported in **Table 1**, the proposed method performs very well (both by using the D-LSTM and 1D-TDCNN backbone) in that it shows overall accuracy and accuracy for each classified class slightly greater to the compared methods. The significant advantage

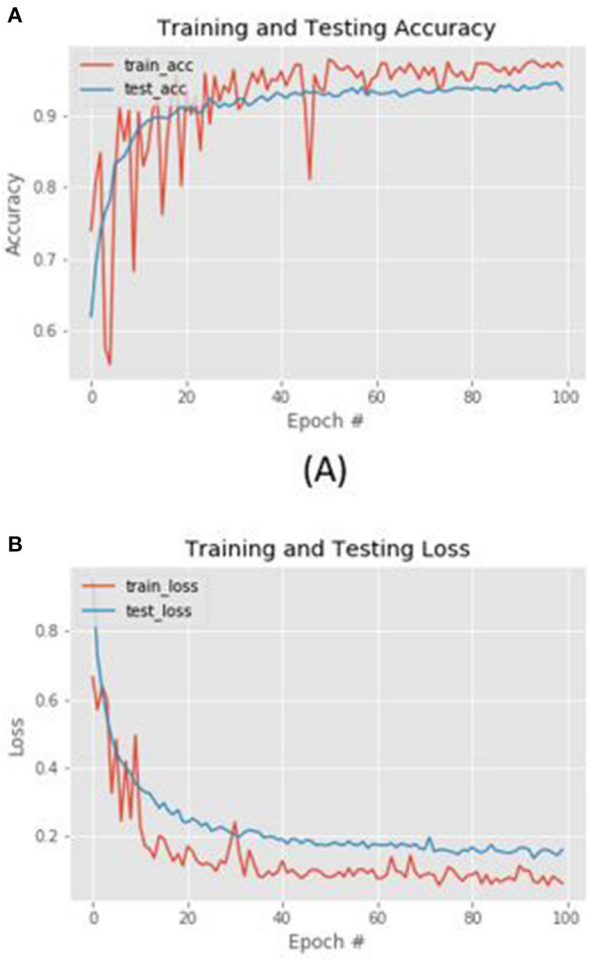


FIGURE 15 | For the proposed D-CNN performance diagrams: **(A)** Training and testing accuracy **(B)** training and testing loss.

of the proposed method is non-invasiveness in that it does not require the driver to wear any sensor nor does it requires the driver necessarily places his/her hand on the part of the steering in which the PPG signal sensors are housed. Similarly, in the same dataset, we validated the reconstruction of the driver's blood pressure level through the SNN network trained with the features extracted from PPG signal (or from the Vision2PPG reconstruction block). Also, in this case we have compared our method with others reported in the literature (Rundo et al., 2019a). From **Table 3**, the performance of the proposed method is very promising as we are able to classify normal-pressure subjects (with an accuracy of 88.88% with respect to 91.11% showed by similar PPG based pipeline; Rundo et al., 2019a) from those who instead have a pressure level out of range (accuracy of 92.30%) with an overall accuracy of 90.14% relatively outperforming with respect to the similar methods based on PPG signal sampling. We tested the D-CNN based pipeline for tracking and classifying the visual

features extracted from the driver's face during the measurement sessions. Also in this case we have obtained an accuracy about of 90% in the testing dataset confirming the robustness of the proposed approach. In **Figure 15**, we show the loss and accuracy related to the D-CNN learning and testing phase. Although the target of this proposal is not the detection, tracking and segmentation of pedestrians in the driving scene, we have still validated the deep architecture we have developed by including a Self Attention layer based on Criss-Cross processing in the Mask-R-CNN network with ResNet-101 as the backbone. As explained, that enhanced Mask-R-CNN was used as simple application use-case for checking driving safety by mean of the proposed pipeline. As introduced in the previous section, we obtained in the testing phase a performance mIoU of 0.695 over CamVid dataset which outperforms other similarly complex architectures described in the literature (He et al., 2017, 2020).

7. CONCLUSIONS AND DISCUSSION

In this study, we addressed the mentioned issue regarding the estimation of driving safety by using a non-invasive and robust methodology. Contrary to existing approaches, our proposed method does not require contact with the driver or the necessity to wear PPG sensors to collect the physiological signal. The advantages of non-invasiveness are however accompanied by overlapping (and sometimes even higher) performances than those obtained by the classic methods which, however, require the use of invasive sensing devices. The proposed multi-modal approach that involves the use of visual and physiological data and to correlate to each other (by means of Vision2PPG reconstruction system) allows to obtain high fault tolerance performance. The experimental results allow to be confident about the applicability of the proposed Vison2Physio approach in different scientific applications. The implemented pipeline is ongoing to be ported to SPC5 Chorus based platform (PPG sensing and processing) while the deep learning architectures will be ported in the STA1295A Accordo5 embedded automotive platform in which two ARM Cortex A7 and 3D GFX accelerator hardware are able to host the developed software as firmware. The embedded Operating System used for both applications is *ad-hoc* YOCTO Linux distribution released for this specific kind of application. The use-case analyzed in this proposal concerns the intelligent tracking of pedestrians in a safe driving scenario. However, the implemented approach can be successfully used in several other automotive use cases. For instance, we have implemented a safe driving application in which, through a deep architecture based on Fully Convolutional Neural Network with Self Attention, we are able to classify the level of risk of the driving scenario and at the same time the driver drowsiness through the PPG-based monitoring pipeline herein described (Rundo et al., 2020a). Another automotive issue we have addressed is the robust identification of the car driver. Through *ad-hoc* intelligent pipeline based on the usage of PPG signal and deep network, we have designed the so called "physiological fingerprint" of the driver used for an effective identification (Rundo et al., 2020d). Therefore the proposed method can be generalized and applied in various automotive scenarios in which it is necessary to characterize the level of attention of the driver or in all the use-cases that deal with driving safety. As future development, we plan to collect more data with the aim of improving the effectiveness of the proposed approach. Specifically, we will address further application in the automotive field with special focus to autonomous driving and ADAS systems during low-light driving scenarios both inside and outside the car.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of confidentiality restrictions related to registered patents. Requests to access the datasets should be directed to francesco.rundo@st.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Committee CT1 (authorization n.113/2018/PO). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

FR and FT worked on data collection and data analysis. FR, RL, FT, and CS worked on writing of the manuscript. SC and SB supervised the research and worked on data collection. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by from the National Funded Program 2014–2020 under Grant Agreement Nr. 1733, (ADAS + Project). The reported information is covered by the following registered patents: IT Patent Nr. 102017000120714, 24 October 2017. IT Patent Nr. 102019000005868, 16 April 2018; IT Patent Nr. 102019000000133, 07 January 2019.

ACKNOWLEDGMENTS

The authors thank the physiologists belonging to the Department of Biomedical and Biotechnological Sciences (BIOMETEC) of the University of Catania, who collaborated in this work in the context of the clinical study covered by Ethical Committee CT1 authorization n.113/2018/PO.

REFERENCES

Agró, D., Canicattí, R., Tomasino, A., Giordano, A., Adamo, G., Parisi, A., et al. (2014). "PPG embedded system for blood pressure monitoring," in *2014 AEIT Annual Conference-From Research to Industry: The Need for a More Effective Technology Transfer (AEIT)*, 1–6. doi: 10.1109/AEIT.2014.7002012

Alty, S. R., Angarita-Jaimes, N., Millasseau, S. C., and Chowienczyk, P. J. (2007). Predicting arterial stiffness from the digital volume pulse waveform. *IEEE Trans. Biomed. Eng.* 54, 2268–2275. doi: 10.1109/TBME.2007.897805

Balakrishnan, G., Durand, F., and Guttag, J. (2013). "Detecting pulse from head motions in video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3430–3437. doi: 10.1109/CVPR.2013.440

Battiato, S., Conoci, S., Leotta, R., Ortis, A., Rundo, F., and Trenta, F. (2020). "Benchmarking of computer vision algorithms for driver monitoring on automotive-grade devices," in *2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, 1–6. doi: 10.23919/AEITAUTOMOTIVE50086.2020.9307437

Bhola, G., Kathuria, A., Kumar, D., and Das, C. (2020). "Real-time pedestrian tracking based on deep features," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1101–1106. doi: 10.1109/ICICCS48265.2020.9121061

Choi, M., Koo, G., Seo, M., and Kim, S. W. (2017). Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness. *IEEE Trans. Instrument. Measure.* 67, 634–645. doi: 10.1109/TIM.2017.2779329

Choi, Y., Han, S. I., Kong, S.-H., and Ko, H. (2016). Driver status monitoring systems for smart vehicles using physiological sensors: a safety enhancement system from automobile manufacturers. *IEEE Signal Process. Mag.* 33, 22–34. doi: 10.1109/MSP.2016.2602095

Conoci, S., Rundo, F., Fallica, G., Lena, D., Buraioli, I., and Demarchi, D. (2018). "Live demonstration of portable systems based on silicon sensors for the monitoring of physiological parameters of driver drowsiness and pulse wave velocity," in *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 1–3. doi: 10.1109/BIOCAS.2018.8584709

Dastjerdi, A. E., Kachuee, M., and Shabany, M. (2017). "Non-invasive blood pressure estimation using phonocardiogram," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–4. doi: 10.1109/ISCAS.2017.8050240

Deng, W., and Wu, R. (2019). Real-time driver-drowsiness detection system using facial features. *IEEE Access* 7, 118727–118738. doi: 10.1109/ACCESS.2019.2936663

Guo, J.-M., and Markoni, H. (2019). Driver drowsiness detection using hybrid convolutional neural network and long short-term memory. *Multimed. Tools Appl.* 78, 29059–29087. doi: 10.1007/s11042-018-6378-6

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969. doi: 10.1109/ICCV.2017.322

He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2020). Mask r-cnn. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 386–397. doi: 10.1109/TPAMI.2018.2844175

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., and Liu, W. (2019). "CCNET: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 603–612. doi: 10.1109/ICCV.2019.00069

Hui, X., and Kan, E. C. (2019). "Seat integration of RF vital-sign monitoring," in *2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC)*, *Vol. 1*, 1–3. doi: 10.1109/IMBIOC.2019.8777742

Husodo, A. Y., Hermawan, I., Jatmiko, W., Wiweko, B., Boediman, A., and Pradekso, B. K. (2018). "Multi-parameter measurement tool of heart rate and blood pressure to detect Indonesian car drivers drowsiness," in *2018 3rd International Seminar on Sensors, Instrumentation, Measurement and Metrology (ISSIMM)*, 68–73. doi: 10.1109/ISSIMM.2018.8727729

Huynh, T. H., Jafari, R., and Chung, W.-Y. (2018). Noninvasive cuffless blood pressure estimation using pulse transit time and impedance plethysmography. *IEEE Trans. Biomed. Eng.* 66, 967–976. doi: 10.1109/TBME.2018.2865751

Jabbar, R., Shinoy, M., Kharbeche, M., Al-Khalifa, K., Krichen, M., and Barkaoui, K. (2020). "Driver drowsiness detection model using convolutional neural networks techniques for android application," in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, 237–242. doi: 10.1109/ICIoT48696.2020.9089484

Jeon, H.-M., Nguyen, V. D., and Jeon, J. W. (2019). "Pedestrian detection based on deep learning," in *IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society, Vol. 1*, 144–151. doi: 10.1109/IECON.2019.8927417

Kazemi, V., and Sullivan, J. (2014). "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1867–1874. doi: 10.1109/CVPR.2014.241

Koh, S., Cho, B. R., Lee, J.-I., Kwon, S.-O., Lee, S., Lim, J. B., et al. (2017). "Driver drowsiness detection via ppg biosignals by using multimodal head support," in *2017 4th International Conference on Control, Decision and Information Technologies (CoDIT)*, 383–388. doi: 10.1109/CoDIT.2017.8102622

Kurian, D., PL, J. J., Radhakrishnan, K., and Balakrishnan, A. A. (2014). "Drowsiness detection using photoplethysmography signal," in *2014 Fourth International Conference on Advances in Computing and Communications*, 73–76. doi: 10.1109/ICACC.2014.23

Lee, B.-G., Jung, S.-J., and Chung, W.-Y. (2011). Real-time physiological and vision monitoring of vehicle driver for non-intrusive drowsiness detection. *IET commun.* 5, 2461–2469. doi: 10.1049/iet-com.2010.0925

Lee, H., Lee, J., and Shin, M. (2019). Using wearable ECG/PPG sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots. *Electronics* 8:192. doi: 10.3390/electronics8020192

Littler, W. A., Honour, A., and Sleight, P. (1973). Direct arterial pressure and electrocardiogram during motor car driving. *Br. Med. J.* 2, 273–277. doi: 10.1136/bmj.2.5861.273

Liu, J., Yan, B. P.-Y., Dai, W.-X., Ding, X.-R., Zhang, Y.-T., and Zhao, N. (2016). Multi-wavelength photoplethysmography method for skin arterial pulse extraction. *Biomed. Opt. Exp.* 7, 4313–4326. doi: 10.1364/BOE.7.004313

Mazzillo, M., Condorelli, G., Sanfilippo, D., Valvo, G., Carbone, B., Fallica, G., et al. (2009). Silicon photomultiplier technology at stmicroelectronics. *IEEE Trans. Nucl. Sci.* 56, 2434–2442. doi: 10.1109/TNS.2009.2024418

Mazzillo, M., Maddiona, L., Rundo, F., Sciuto, A., Libertino, S., Lombardo, S., et al. (2018). Characterization of SiPMs with NIR long-pass interferential and plastic filters. *IEEE Photon. J.* 10, 1–12. doi: 10.1109/JPHOT.2018.2834738

Monte-Moreno, E. (2011). Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. *Artif. Intell. Med.* 53, 127–138. doi: 10.1016/j.artmed.2011.05.001

Oh, T.-H., Jaroensri, R., Kim, C., Elgharib, M., Durand, F., Freeman, W. T., et al. (2018). "Learning-based video motion magnification," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 633–648. doi: 10.1007/978-3-030-01225-0_39

Rubinstein, M., Wadhwa, N., Durand, F., Freeman, W. T., and Wu, H.-Y. (2013). Revealing invisible changes in the world. *Science* 339:519.

Rundo F., Petralia S., Fallica G., and Conoci S. (2019d). "A nonlinear pattern recognition pipeline for PPG/ECG medical assessments," in *Sensors. CNS 2018. Lecture Notes in Electrical Engineering, Vol. 539,* eds Andò B. et al. (Cham: Springer). doi: 10.1007/978-3-030-04324-7_57

Rundo, F., Conoci, S., Battiato, S., Trenta, F., and Spampinato, C. (2020a). "Innovative saliency based deep driving scene understanding system for automatic safety assessment in next-generation cars," in *2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, 1–6. doi: 10.23919/AEITAUTOMOTIVE50086.2020.9307425

Rundo, F., Conoci, S., and Fallica, P. G. (2021). *Method of Processing Electrophysiological Signals and Corresponding System, Vehicle, and Computer Program Product.* USA Patent Nr. 10987007.

Rundo, F., Conoci, S., Ortis, A., and Battiato, S. (2018a). An advanced bio-inspired photoplethysmography (PPG) and ECG pattern recognition system for medical assessment. *Sensors* 18:405. doi: 10.3390/s18020405

Rundo, F., Ortis, A., Battiato, S., and Conoci, S. (2018b). Advanced bio-inspired system for noninvasive cuff-less blood pressure estimation from physiological signal analysis. *Computation* 6:46. doi: 10.3390/computation6030046

Rundo, F., Ortis, A., Battiato, S., and Conoci, S. (2019a). "Advanced multi-neural system for cuff-less blood pressure estimation through nonlinear HC-features," in *ICETE (1)*, 327–331. doi: 10.5220/0007909403210325

Rundo, F., Rinella, S., Massimino, S., Coco, M., Fallica, G., Parenti, R., et al. (2019b). An innovative deep learning algorithm for drowsiness detection from EEG signal. *Computation* 7:13. doi: 10.3390/computation7010013

Rundo, F., Spampinato, C., Battiato, S., Trenta, F., and Conoci, S. (2020b). "Advanced 1d temporal deep dilated convolutional embedded perceptual system for fast car-driver drowsiness monitoring," in *2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, 1–6. doi: 10.23919/AEITAUTOMOTIVE50086.2020.9307400

Rundo, F., Spampinato, C., and Conoci, S. (2019c). *Ad-hoc* shallow neural network to learn hyper filtered photoplethysmographic (PPG) signal for efficient car-driver drowsiness monitoring. *Electronics* 8:890. doi: 10.3390/electronics8080890

Rundo, F., Spampinato, C., Conoci, S., Trenta, F., and Battiato, S. (2020c). "Deep bio-sensing embedded system for a robust car-driving safety assessment," in *2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE)*, 1–6. doi: 10.23919/AEITAUTOMOTIVE50086.2020.9307409

Rundo, F., Trenta, F., Leotta, R., Spampinato, C., Piuri, V., Conoci, S., et al. (2020d). "Advanced temporal dilated convolutional neural network for a robust car driver identification," in *ICPR Workshops*, 184–199.

Schmidt, Robert, F., and Thews, G. (1989). "Autonomic nervous system," in *Human Physiology, 2nd Edn*, ed W. Janig, 333–370. doi: 10.1007/978-3-642-73831-9_16

Slapničar, G., Mlakar, N., and Luštrek, M. (2019). Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. *Sensors* 19:3420. doi: 10.3390/s19153420

Song, H., Choi, I. K., Ko, M. S., Bae, J., Kwak, S., and Yoo, J. (2018). "Vulnerable pedestrian detection and tracking using deep learning," in *2018 International Conference on Electronics, Information, and Communication (ICEIC)* (Honolulu, HI), 1–2. doi: 10.23919/ELINFOCOM.2018.8330547

Tian, Y., Luo, P., Wang, X., and Tang, X. (2015). "Deep learning strong parts for pedestrian detection," in *Proceedings of the IEEE International Conference on Computer Vision* (Santiago), 1904–1912. doi: 10.1109/ICCV.2015.221

Trenta, F., Conoci, S., Rundo, F., and Battiato, S. (2019). "Advanced motion-tracking system with multi-layers deep learning framework for innovative car-driver drowsiness monitoring," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)* (Lille), 1–5. doi: 10.1109/FG.2019.8756566

Vavrinský, E., Tvarožek, V., Stopjaková, V., Soláriková, P., and Brezina, I. (2010). "Monitoring of car driver physiological parameters," in *The Eighth International Conference on Advanced Semiconductor Devices and Microsystems* (Smolenice), 227–230. doi: 10.1109/ASDAM.2010.5667021

Vinciguerra, V., Ambra, E., Maddiona, L., Oliveri, S., Romeo, M. F., Mazzillo, M., et al. (2017). "Progresses towards a processing pipeline in photoplethysmogram (PPG) based on SiPMs," in *2017 European Conference on Circuit Theory and Design (ECCTD)* (Catania), 1–5. doi: 10.1109/ECCTD.2017.8093327

Vinciguerra, V., Ambra, E., Maddiona, L., Romeo, M., Mazzillo, M., Rundo, F., et al. (2019). "PPG/ECG multisite combo system based on SiPM technology," in *Sensors. CNS 2018. Lecture Notes in Electrical Engineering, Vol. 539, eds* Andò B. et al. (Cham: Springer). doi: 10.1007/978-3-030-04324-7_44

Viola, P., and Jones, M. (2001). "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Vol. 1* (Kauai, HI). doi: 10.1109/CVPR.2001.990517

Vural, E., Cetin, M., Ercil, A., Littlewort, G., Bartlett, M., and Movellan, J. (2007). "Drowsy driver detection through facial movement analysis," in *International Workshop on Human-Computer Interaction* (Rio de Janeiro: Springer), 6–18. doi: 10.1007/978-3-540-75773-3_2

Wu, C.-Y., Hu, H.-Y., Chou, Y.-J., Huang, N., Chou, Y.-C., and Li, C.-P. (2015). High blood pressure and all-cause and cardiovascular disease mortalities in community-dwelling older adults. *Medicine* 94:47. doi: 10.1097/MD.0000000000002160

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.* 31, 1–8. doi: 10.1145/2185520.2185561

Zhao, W., Gao, Y., Ji, T., Wan, X., Ye, F., and Bai, G. (2019). Deep temporal convolutional networks for short-term traffic flow forecasting. *IEEE Access* 7, 114496–114507. doi: 10.1109/ACCESS.2019.2935504

Conflict of Interest: FR was employed by the company STMicroelectronics.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroinformatics

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fninf.2021.605729

Multi-Dimensional Enhanced Seizure Prediction Framework Based on Graph Convolutional Network

Xin Chen[1,2], *Yuanjie Zheng*[1,3,4,5]*, *Changxu Dong*[1] and *Sutao Song*[1,3,4,5]*

[1] School of Information Science and Engineering at Shandong Normal University, Jinan, China, [2] School of Computer Science and Engineering at Southeast University, Nanjing, China, [3] Key Lab of Intelligent Computing and Information Security in Universities of Shandong, Shandong Normal University, Jinan, China, [4] Shandong Provincial Key Laboratory for Novel Distributed Computer Software Technology, Shandong Normal University, Jinan, China, [5] Institute of Biomedical Sciences, Shandong Normal University, Jinan, China

In terms of seizure prediction, how to fully mine relational data information among multiple channels of epileptic EEG? This is a scientific research subject worthy of further exploration. Recently, we propose a multi-dimensional enhanced seizure prediction framework, which mainly includes information reconstruction space, graph state encoder, and space-time predictor. It takes multi-channel spatial relationship as breakthrough point. At the same time, it reconstructs data unit from frequency band level, updates graph coding representation, and explores space-time relationship. Through experiments on CHB-MIT dataset, sensitivity of the model reaches 98.61%, which proves effectiveness of the proposed model.

Keywords: epilepsy EEG signal, seizures prediction, multichannel relationship, graph convolutional network, space-time prediction

OPEN ACCESS

Edited by:
Giuseppe Luigi Banna, United Lincolnshire Hospitals NHS Trust, United Kingdom

Reviewed by:
Xizi Song, Tianjin University, China
Dong Wen, Yanshan University, China
Fangzhou Xu, Qilu University of Technology, China

***Correspondence:**
Yuanjie Zheng
zhengyuanjie@gmail.com
Sutao Song
sutao.song@sdnu.edu.cn

Received: *13 September 2020*
Accepted: *26 July 2021*
Published: *19 August 2021*

Citation:
Chen X, Zheng Y, Dong C and Song S (2021) Multi-Dimensional Enhanced Seizure Prediction Framework Based on Graph Convolutional Network. Front. Neuroinform. 15:605729. doi: 10.3389/fninf.2021.605729

1. INTRODUCTION

Epilepsy is a chronic disease caused by brain dysfunction, which is characterized by sudden and transient (Jia et al., 2004). In the study of seizure symptoms, EEG plays an important role. It is an important auxiliary technology for epilepsy diagnosis (Yuan et al., 2012). In traditional diagnosis of epilepsy symptoms, EEG data is often analyzed by experienced doctors. This process takes the doctors too much time and energy. Besides, doctors often work for long time, their judgments are also likely to be negatively affected by their body fatigue. To solve the problem, automatic detection technology of epilepsy EEG is born.

At present, there are abundant researches of automatic detection based on epilepsy EEG. However, only a few literatures have focused on the analysis of seizure prediction. In real life, it is meaningful to predict seizures. For patients, uncertainty of seizures may cause unpredictable accidents, which may seriously affect life and work (Holmes, 1984; Ahmed, 2005). The effective prediction of seizure can help patients to solve the problems in time, thus reducing the loss of patients to a minimum. In addition, for doctors and researchers, the effective prediction of seizures not only helps them to explore the basic mechanisms of epileptic seizures, but also provides important support for building accurate and stable epilepsy auxiliary diagnostic tools. In this paper, we will focus on seizure prediction.

The process of seizure prediction generally consists of EEG signal acquisition, data preprocessing, feature extraction, and classification. In particular, mining rich and effective features from native EEG data is essential to improve classification accuracy. According to variation of the

domains that EEG features are extracted from, features may come from time domain, frequency domain, and the time-frequency domain.

In time domain, Tessy et al. (2016) focused on extracting two time-domain features of line length and energy to obtain high classification results on k-nearest neighbors (KNN) classifier. The algorithm proposed by Shanir et al. (2015) was based on average and minimum of each segment. In addition, the team of Zhang et al. (2018) divided EEG into several clusters and extracted a set of time-domain features from each cluster. Each group of features was regarded as a node of complex network. Then, average weighted degree was calculated from network as classification feature.

In frequency-domain, through experiments on Freiburg and CHB-MIT databases, Zhou et al. (2018) proved that classification accuracy of frequency-domain signals is significantly better than that of time-domain signals. Al Ghayab et al. (2019) extracted statistical features from sub windows by Fourier transform, and sorted features by using information gain technique to select the most appropriate ones.

In time-frequency-domain, more and more researches focused on extracting EEG features from time-frequency distribution (TFDs). Guerrero-Mosquera et al. (2010) extracted length, frequency, and energy from the smoothed Wigner-Ville distribution (SWVD) by using trajectory estimated from McAulay-Quatieri sinusoidal model. A new method (Wavelet-Chaos) of wavelet transformation was proposed by analyzing δ, θ, α, and β subbands of EEG, they found that significant differences could be captured by combining subband information (Adeli et al., 2007). Sharma et al. (2014) used cubical threshold denoising methods based on wavelet to analyze EEG signals before extracting statistical features from frequency bands ($0 \sim 8$, $8 \sim 16$, $16 \sim 32$, and $0 \sim 32$ Hz). In addition, local binary pattern (LBP) was extended to analysis of EEG signals, because of its outstanding advantages such as rotation invariance and gray invariance. For example, Shanir et al. (2018) proposed a morphological feature extraction method based on LBP operator.

Considering comprehensively above these points, we propose multi-dimensional enhanced seizure prediction framework based on graph convolutional network (MESPF). The contributions of this research are as follows:

- It is very importance to improve accuracy of seizure prediction. To provide the prediction model with more powerful and abundant data, we enhance overall consideration of dimensions about epilepsy EEG signals. In particular, we take multichannel spatial relationship as breakthrough point, update representation of graph relational data at frequency band level, and explore the space-time relationships.
- We propose a multi-dimensional enhanced seizure prediction framework, which mainly includes information reconstruction space, graph state encoder, and space-time predictor. We combine technical advantages of wavelet packet decomposition, graph convolutional network and gated recurrent neural network in construction of the framework. From aspects of frequency, channel, and time, we intend to mine richer and more effective data information than previous studies.
- Finally, we apply the framework to data set (CHB-MIT) for verification. In terms of experimental results, it has surpassed or approached many existing algorithms. In short, our framework provides a novel way for peers to study principle of seizure prediction.

The organizational structure of our article is as follows, the second section introduces prediction principle of the framework (MESPF), the third section briefly describes experimental results and analysis, and the fourth section is summary.

2. METHODOLOGIES

For seizure prediction, we propose a multi-dimensional enhanced seizure prediction framework (MESPF). It mainly includes information reconstruction space, graph state encoder, and space-time predictor. The main process of the model is shown in **Figure 1**. The following contents give specific explanations in turn.

2.1. Information Reconstruction Space

Patients vary in physiological mechanisms, and their pre-ictal signals may occur in different frequency band range. To capture these subtle difference, we designs an information reconstruction space. It is mainly used to highlight specificity and explore the effect of feature enhancement on graph encoding.

Since epileptic EEG signals are random, non-stationary and non-linear, we actively introduce wavelet packet decomposition (WPD) in information reconstruction space to decompose EEG signals. It should be emphasized that since range of wavelet transform is mainly low frequency part of signals, it is difficult to characterize a large amount of detailed information. However, wavelet packet decomposition can orthogonally decompose signal in full frequency range, and resolution of high frequency part is better than the former. It is a more precise analysis method than wavelet transform. So that it has gradually become one of main methods for analyzing non-stationary signals (Hyvarinen et al., 2001). Through wavelet packet decomposition, we can analyze EEG signals from multiple frequency bands.

The principle of information reconstruction space in MESPF is shown in **Figure 2**. By decomposing epileptic EEG signals and calculating characteristics of energy value, it can update representation of graph data. First of all, $C_1 - C_n$ represents 18 channels. In the third step, data of each channel is processed by wavelet packet decomposition technique. Then we calculate energy values of corresponding sub-bands and reconstruct vectors characterized by energy values (step 4–5). After completion of channel data reconstruction, we use Pearson correlation coefficient calculation method to calculate correlation between multiple channels (step 6). We update graph relationship representation as one of direct inputs to graph encoder (equivalent to edges of graph). Finally, related relational data and each channel data itself (equivalent to vertex data of graph) are used as input to graph encoder, as shown in step 7.

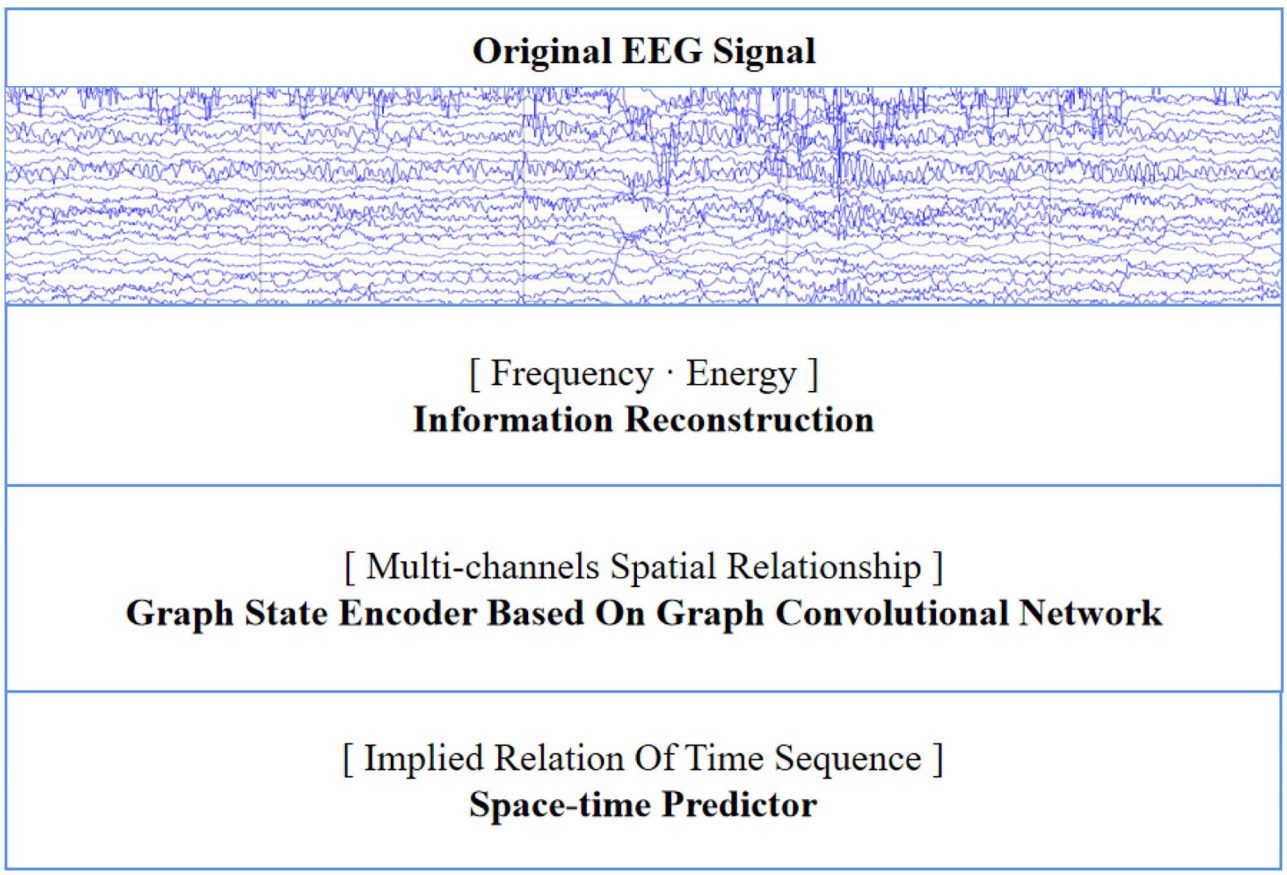


FIGURE 1 | Our model mainly includes information reconstruction space, graph state encoder, and space-time predictor.

Specifically, decomposition principle of information reconstruction space is shown in **Figure 3**. For a signal, wavelet coefficients (A and D) are obtained by wavelet packet decomposition of the first layer. A represents low frequency part of signal, and D represents high frequency part. Each node in graph represents a data sequence. Then, energy values of each sub-band are calculated, respectively. In the experiment, four-layer wavelet packet decomposition is adopted to obtain wavelet coefficients of 16 frequency bands. In this way, each signal unit can be represented by energy values. It should be noted that since epilepsy EEG signal is continuous waveform, our experiment uses a relatively smooth Daubechies (Db) wavelet base.

2.2. Graph State Encoder

After signal unit is processed by information reconstruction space, the graph data are given into graph encoder for further feature processing.

Seizure is a synergistic result of multiple brain regions. For EEG, each channel records activities of different brain regions, and there must be a certain relationship between channels. Therefore, we build a graph state encoder to explore relationship by extracting spatial features between multiple channels. The specific structure and parameters about graph encoder are shown in **Figure 4**, which mainly includes input layer, graph convolution layer-1, graph convolution layer-2, fully connected layer, and graph status code layer. The activation layer in network uses Rectified Linear Unit (ReLU), which can speed up convergence speed. In general, after extracting features of graph space through graph convolution layer, relevant data are weighted by fully connected layer. Then, graph status code is generated. As for status code, it is composed of 18 eigenvalues, which, respectively, represent 18 channels of graph data. Each status code represents data characteristics of corresponding time period. Finally, we feed status codes into space-time predictor in chronological order.

As for graph state encoder, we refer to method proposed by Kipf and Welling (2016) and build sub-module based on graph convolutional network. In particular, we need to emphasize application of filter. Firstly, we define convolution operation on graph as:

$$Conv = X_{MESPF} * K_e \tag{1}$$

X_{MESPF} represents direct input (signal) of graph state encoder, K_e represents convolution kernel. Furthermore, in previous studies, Defferrard et al. (2016) introduced Chebyshev polynomials,

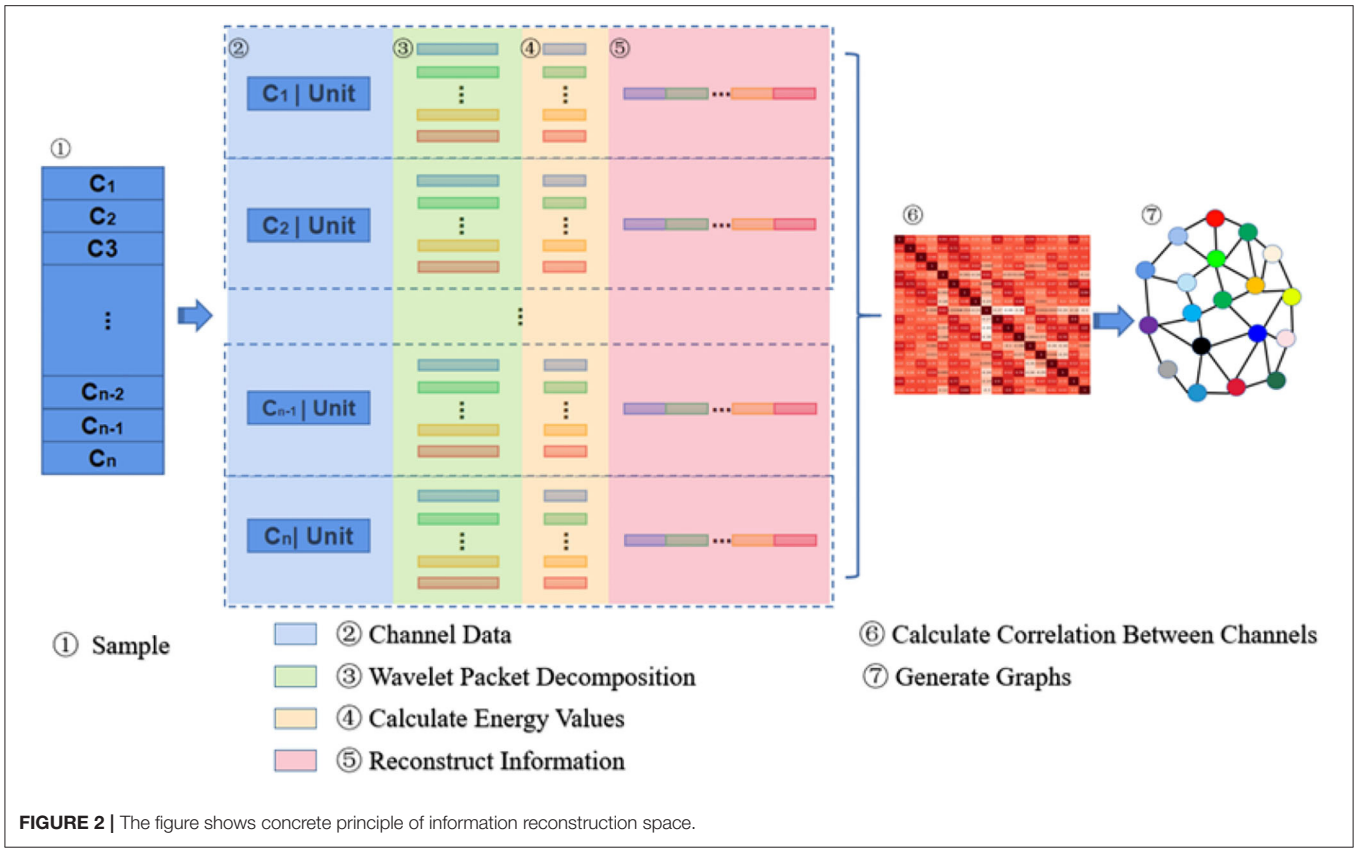


FIGURE 2 | The figure shows concrete principle of information reconstruction space.

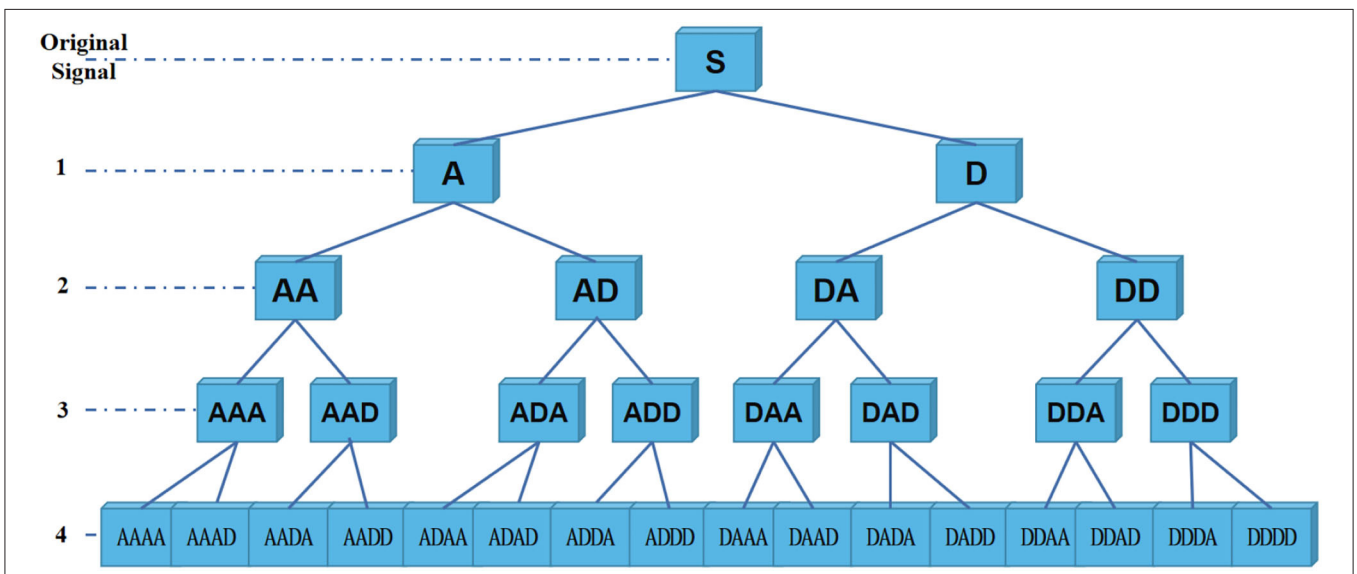


FIGURE 3 | EEG signals are divided into several sub-bands by wavelet packet decomposition.

and they provided an algorithm for constructing fast local filters in spectral domain, which can learn local, static and combined features on graph. Subsequently, Kipf and Welling (2016) optimized convolutional network structure through the first-order local approximation of spectrogram convolution. In process of graph convolution, due to existence of graph Fourier transform, computational complexity of model is relatively high. Chebyshev polynomials have numerical stability and computational efficiency in field of polynomial function approximation. Therefore, we actively introduce Chebyshev

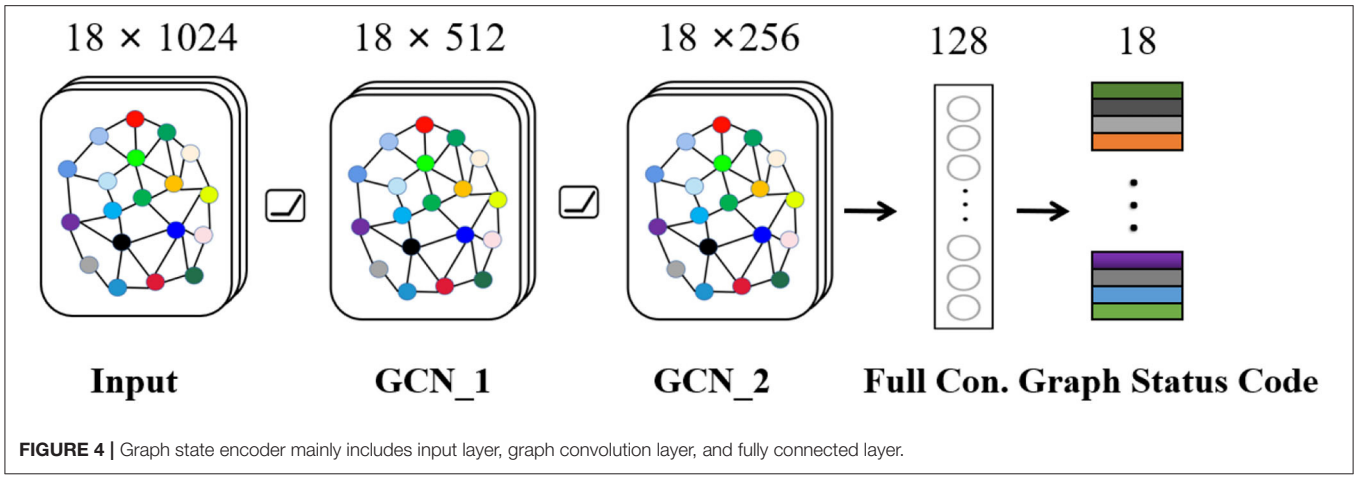


FIGURE 4 | Graph state encoder mainly includes input layer, graph convolution layer, and fully connected layer.

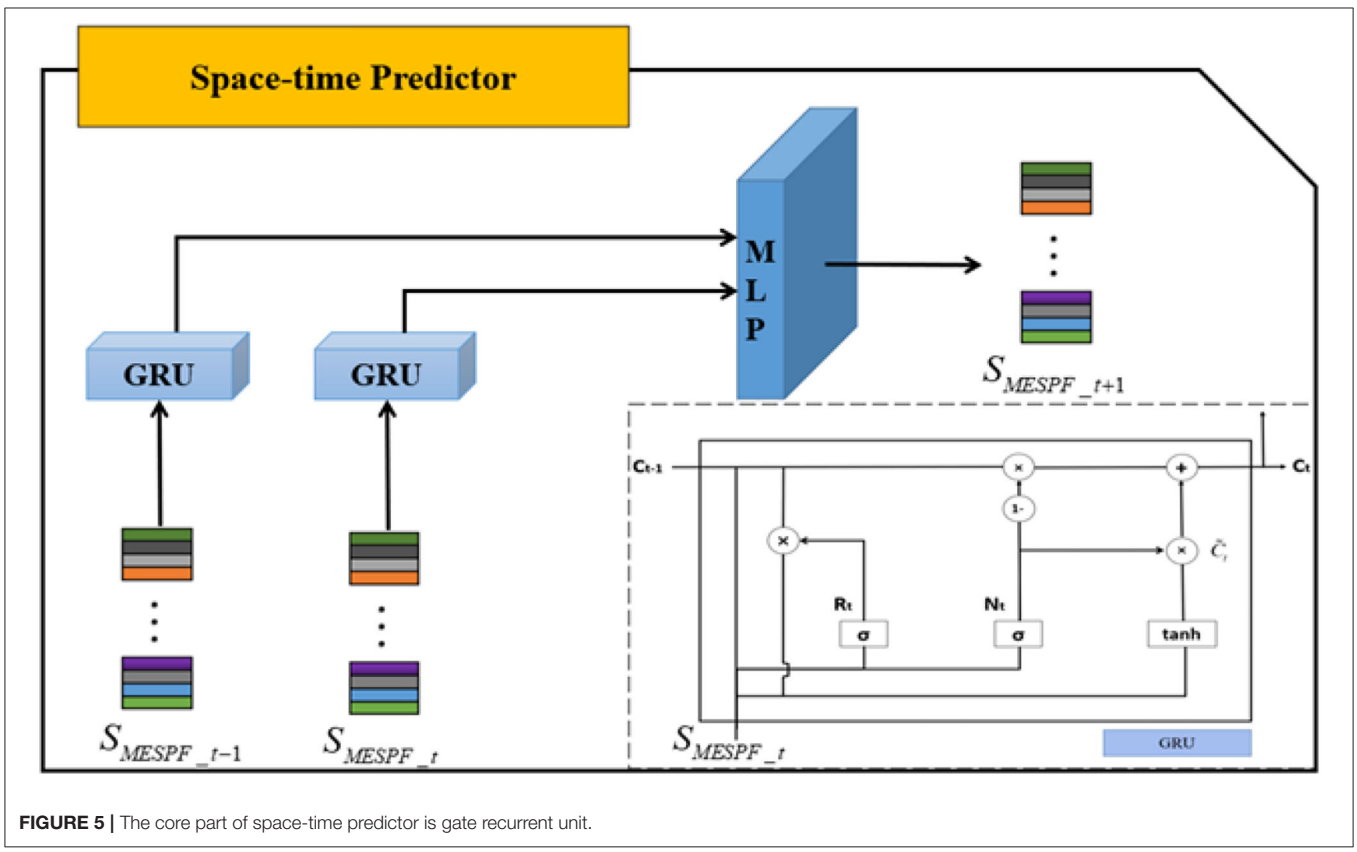


FIGURE 5 | The core part of space-time predictor is gate recurrent unit.

polynomials into the model. The relevant formula is as follows:

$$T_k(X_{MESPF}) = 2X_{MESPF}T_{k-1}(X_{MESPF}) - T_{k-2}(X_{MESPF}) \quad (2)$$

$$K_e(\Lambda) \approx \sum_{n=0}^{N} \theta_n T_n(\widetilde{\Lambda}) \quad (3)$$

θ represents vector of Chebyshev coefficients, $\widetilde{\Lambda} = \frac{2}{\lambda_{max}}\Lambda - I$, Λ represents diagonal matrix of eigenvalues, λ_{max} represents maximum eigenvalue of regularized feature matrix, I represents identity matrix, convolution on graph is finally expressed as:

$$O_{MESPF} = X_{MESPF} * K_e \approx \sum_{n=0}^{N} \theta_n T_n(\widetilde{\Lambda}) X_{MESPF} \quad (4)$$

O_{MESPF} represents the output of graph convolution layer. Then O_{MESPF} is processed through ReLU activation function, this is shown in Formula 5.

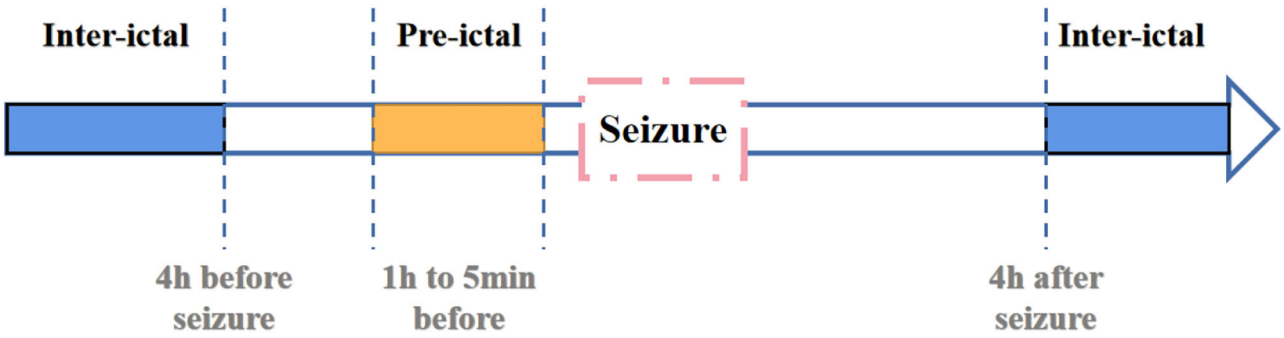


FIGURE 6 | According to actual seizure process, EEG is divided into different stages.

$$O^*_{MESPF} = ReLU(O_{MESPF}) = max(0, O_{MESPF}) \quad (5)$$

Finally, through fully connected layer, we synthesize several graph space features to generate status code S_{MESPF}.

$$S_{MESPF} = FC(O^*_{MESPF})_{18} \quad (6)$$

2.3. Space-Time Predictor

In the model, space-time predictor is constructed. After exploring multi-channel spatial relationship, we further explore change rule at time series level.

The internal structure of space-time predictor is shown in **Figure 5**. S_{MESPF_t-1} and S_{MESPF_t}, respectively, represent graph status codes at time $t-1$ and time t. Specifically, as complexity of neural network architecture increases or training time of experimental data becomes longer, phenomenon of gradient disappearance or explosion is easy to occur. It is difficult to master law of EEG timing signal. The emergence of gate recurrent network solves related problems well. So, it is mainly built on the basis of gate recurrent unit. And direct input of space-time predictor is output of graph state encoder. It should be emphasized that relevant input data should be entered in chronological order. By mining implicit relationship in terms of timing, prediction results are finally output by multilayer perceptron (MLP).

Focus on gated recurrent unit, Cho et al. (2014) integrate forget gate and input gate into an update gate. Specifically, GRU includes update gate and reset gate. The former controls extent to which state information at previous moment is brought into the current state. And the latter controls extent to which state information at previous moment is ignored. The GRU structure is shown in lower right corner of **Figure 6**. Then, we combine diagram and formulas to further explain.

The calculations for reset gate (R_t) is shown in Formula (5). Both W^r and U^r are weight matrices. And σ represents sigmoid function, which can map results between 0 and 1. The calculation principle of update gate (N_t) is similar to that of reset gate.

$$R_t = \sigma(W^r S_{MESPF\_t} + U^r C_{t-1}) \quad (7)$$

$$N_t = \sigma(W^z S_{MESPF\_t} + U^z C_{t-1}) \quad (8)$$

The calculation of candidate hidden layer ($\widetilde{C}_t$) is as follows. Tanh represents tanh function, and its mapping range is −1 to +1.

$$\widetilde{C}_t = \tanh(W S_{MESPF\_t} + R_t U C_{t-1}) \quad (9)$$

The final output hidden layer information (C_t) calculation formula is as follows:

$$C_t = (1 - N_t) * C_{t-1} + N_t * \widetilde{C}_t \quad (10)$$

3. RESULTS ANALYSIS

3.1. Data Set Description

In order to prove effectiveness of the framework, we further apply it to CHB-MIT data set. The download link for complete database is https://physionet.org/content/chbmit/1.0.0/. It consist of 23 data samples from 22 subjects(5 males, ages 3–22; and 17 females, ages 1.5–19). All samples are stored in EDF format. And all signals are sampled at 256 samples per second. It should be noted that this is a verification experiment of validity. In order to achieve effective prediction of seizures, we uses the framework to classify pre-ictal and inter-ictal.

In previous studies, Litt et al. (2001) have demonstrated that seizure-like EEG signals become more frequent at 2 h before actual seizure. They propose that accumulated energy will increase within 50 min before initial state. And Affes et al. (2019) propose that pre-seizure phase is 1 h before seizure. Based on researches mentioned above, we defines pre-ictal period as data between 1 h and 5 min before seizure. And the definition of inter-ictal is shown in **Figure 6**.

It should be noted that in the process of data collection, data of inter-ictal period is much more than data of pre-ictal period. To make the number of samples in these two periods equal, an overlapped window is applied in pre-ictal period for data segmentation, and the window overlap rate is set as 50%.

3.2. Experimental Indicators

The evaluation indicators of our model include Accuracy, Sensitivity, Specificity, False Positive Rate, and F1-Score. The calculation formulas for these indicators are as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (11)$$

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad (12)$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \quad (13)$$

$$False\ Positive\ Rate = \frac{FP}{FP+TN} \times 100 \quad (14)$$

$$Positive\ Predictive\ Value\ (PPV) = \frac{TP}{TP+FP} \times 100 \quad (15)$$

$$Negative\ Predictive\ Value\ (NPV) = \frac{TN}{TN+FN} \times 100 \quad (16)$$

$$F1 - Score = \frac{2TP}{2TP+FP+FN} \quad (17)$$

TP indicates that it is actually a positive example and the predicted result is a positive example. FP is actually a negative example, and the predicted result is a positive example. TN indicates that it is actually a negative example, and predicted result is a negative example. FN indicates that it is actually a positive example, and predicted result is a negative example.

3.3. Experimental Environment Configuration

As for experimental environment, information reconstruction space of MESPF model is completed on the Windows10 system with Intel(R) Core(TM) I7-6500U CPU @ 2.50GHz 2.59GHz. Construction of graph encoder and space-time predictor is completed on the Ubuntu system. Relevant model is built based on TensorFlow framework with Python. Adam optimizer is also used in related experiments, and initial learning rate is set at 0.01.

In addition, the loss function of the model includes cross entropy term and L2 regularization term, as shown in Formula (18). $H(p_{MESPF\_O}, q_{MESPF\_R})$ represents the loss function of MESPF model for seizure prediction. p_{MESPF_O} represents related predicted value of space-time predictor. q_{MESPF_R} represents true label for each data unit.

$$H(p_{MESPF\_O}, q_{MESPF\_R}) = \sum_x P_{MESPF\_O}(x) \cdot \log\left(\frac{1}{q_{MESPF\_R}(x)}\right) + \frac{\lambda}{2}\sum w^2 \quad (18)$$

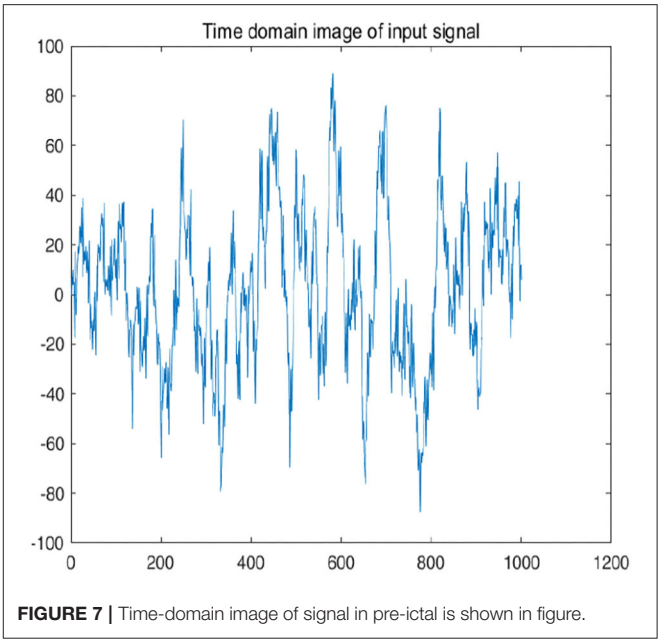


FIGURE 7 | Time-domain image of signal in pre-ictal is shown in figure.

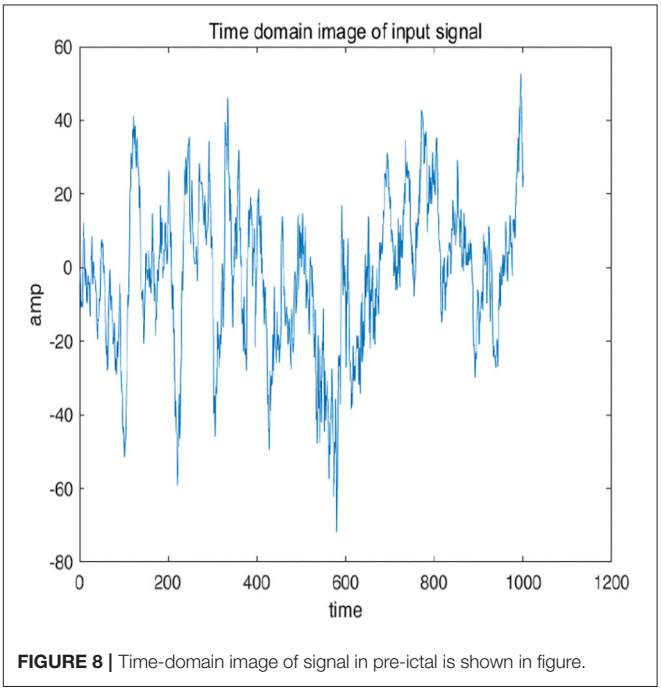


FIGURE 8 | Time-domain image of signal in pre-ictal is shown in figure.

3.4. Analysis of Experimental Results

To further explain principle of information reconstruction space, we set up a control group to visualize the process of reconstruction. One group is pre-ictal samples, and the other group is inter-ictal samples. **Figures 7**, **8** show time domain information. It can be seen that there are more high frequency signals before seizure. These are extremely important features for predicting seizure in advance. **Figures 9**, **10** are images after transforming from time domain to frequency domain. It can also

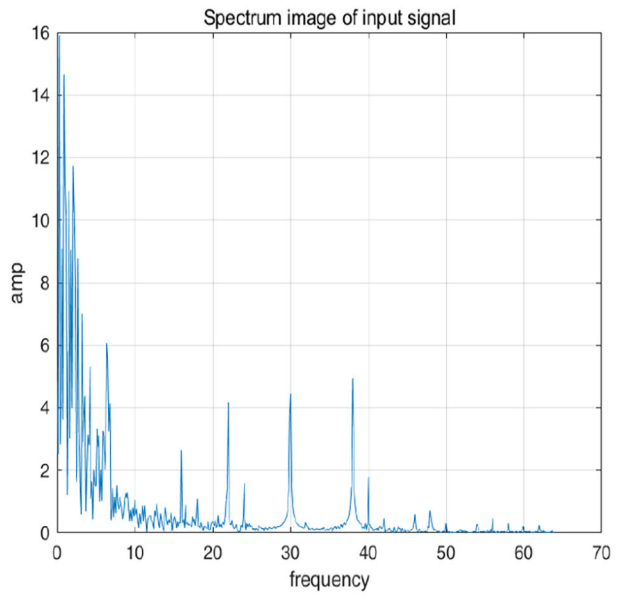


FIGURE 9 | Frequency-domain image of signal in inter-ictal is shown in figure.

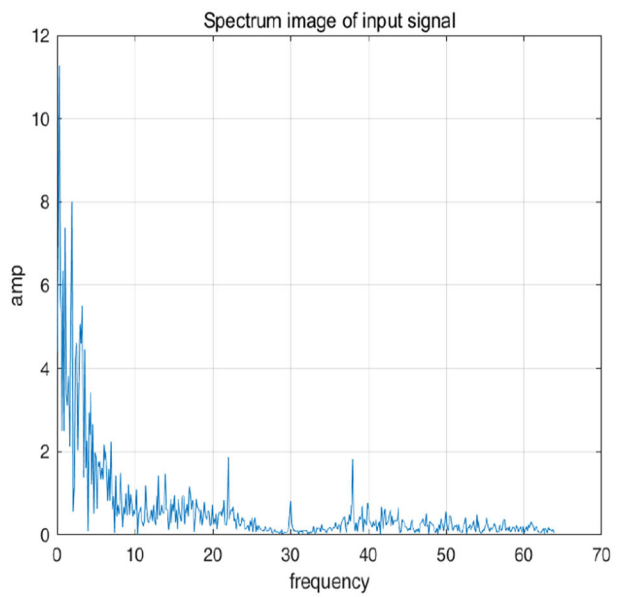


FIGURE 10 | Frequency-domain image of signal in inter-ictal is shown in figure.

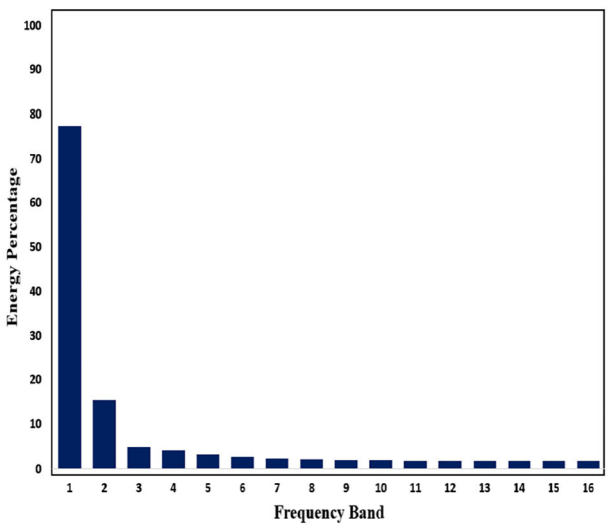


FIGURE 11 | Energy percentage of each frequency band (pre-ictal) in the fourth layer of WPD is shown in figure.

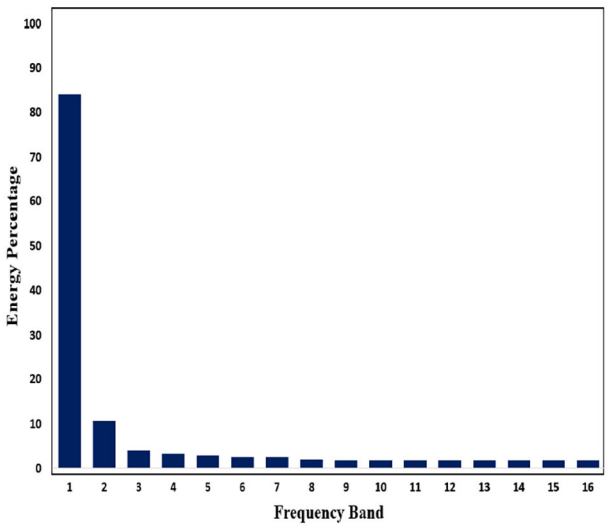


FIGURE 12 | Energy percentage of each frequency band (inter-ictal) in the fourth layer of WPD is shown in figure.

be seen from the former that times of high frequency increases significantly in pre-ictal period.

Subsequently, in order to mine the law of different frequency band in signals, we use WPD to decompose signals. Through WPD, we extract wavelet packet coefficients of nodes. Then energy of wavelet packet coefficients is used as eigenvalue to construct eigenvector. **Figures 11**, **12** show energy values of the fourth layer after four-layer WPD.

In our experiment, each epileptic patient is assigned a unique ID number, from case 1 to case 24. The comprehensive index analysis for each patient is shown in **Table 1**. These indicators include specificity, sensitivity, PPV, NPV, FPR, and accuracy. More than half of the experiments achieve sensitivity of 100%. The average level of false positive rate is only 0.0106. Experimental data shows that the lower false positive rate, the better model's performance. For details of other experimental indicators, please refer to **Table 1**.

As a comprehensive index, F1-Score balances effects of precision and recall, and can systematically evaluate a classifier. The value of F1-Score ranges from 0 to 100%, and the larger

TABLE 1 | Specific indicators of samples are shown in the table.

ID	Specificity	Sensitivity	PPV	NPV	FPR	Accuracy
1	98.85	100.00	98.87	100.00	0.0115	99.43
2	100.00	100.00	100.00	100.00	0.0000	100.00
3	99.67	100.00	99.67	100.00	0.0033	99.84
4	97.82	97.85	97.81	97.85	0.0219	97.83
5	100.00	100.00	100.00	100.00	0.0000	100.00
6	92.85	95.38	93.02	95.26	0.0715	94.11
7	98.65	100.00	98.66	100.00	0.0135	99.32
8	100.00	99.34	100.00	99.34	0.0000	99.67
9	100.00	100.00	100.00	100.00	0.0000	100.00
10	100.00	100.00	100.00	100.00	0.0000	100.00
11	99.24	94.56	99.20	94.80	0.0076	96.90
12	98.46	94.50	98.39	94.71	0.0155	96.48
13	96.98	96.24	96.96	96.27	0.0302	96.61
14	99.02	100.00	99.04	100.00	0.0097	99.51
15	98.63	97.05	98.61	97.10	0.0137	97.84
16	97.76	100.00	97.81	100.00	0.0224	98.88
17	99.12	100.00	99.12	100.00	0.0089	99.56
18	99.42	97.04	99.41	97.11	0.0057	98.23
19	100.00	100.00	100.00	100.00	0.0000	100.00
20	100.00	97.18	100.00	97.26	0.0000	98.59
21	100.00	100.00	100.00	100.00	0.0000	100.00
22	100.00	100.00	100.00	100.00	0.0000	100.00
23	98.37	98.31	98.37	98.31	0.0163	98.34
24	99.85	99.26	99.84	99.26	0.0016	99.55
Average	98.95	98.61	98.95	98.64	0.0106	98.78

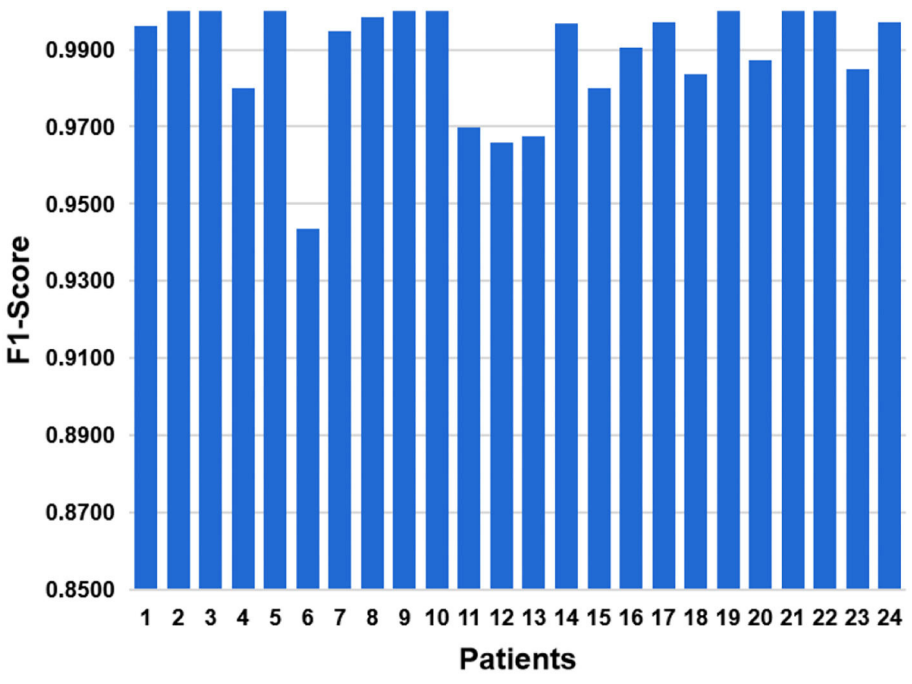


FIGURE 13 | In order to balance influences of precision and recall, we have added F1-Score indicator.

TABLE 2 | In classification tasks of pre-ictal and inter-ictal, the framework is compared with other advanced methods.

References	Years	Method	Specificity	Sensitivity	Accuracy
Khan et al. (2012)	2012	LDA	100.00	83.00	91.80
Kiranyaz et al. (2014)	2014	Automated patient-specific	94.71	89.00	-
Pramod et al. (2014)	2015	NN	99.29	98.06	-
Alotaiby et al. (2015)	2015	CSP	98.61	86.84	92.72
Yuan et al. (2018)	2018	Multi-view DL	-	-	94.37
Solaija et al. (2018)	2018	Dynamic mode decomposition	98.93	87.00	-
Dash et al. (2020)	2019	IFD and HMM	99.85	96.78	99.60
Our method	2020	MESPF	98.95	98.61	98.78

the F1 value, the better the model performance. The result of F1-Score is shown in **Figure 13**. The average level of 24 cases is 0.9877. The best F1-Score value reach 100% for patients with ID 2, 3, and 5. And the lowest F1-score is 0.9419 for patient with ID 6.

Finally, we compare the performance of the proposed model with some existing algorithms, from traditional ones to that based on deep learning. The results are shown in **Table 2**. In terms of our method, we have analyzed epileptic EEG signals from multiple dimensions, including energy characteristics of frequency bands, spatial characteristics of channels, and timing characteristics of signals. In the process, we make full use of advantages of wavelet packet decomposition, graph convolutional network and gate recurrent unit network in their professional fields.

As we all know, sensitivity is the proportion of people who are actually sick that are correctly judged as true positives. It is a meaningful indicator in medical clinical diagnosis. **Table 2** shows an excellent performance of our model in sensitivity (98.61%). Besides, the accuracy is also consistent with the high level of other algorithms, and the specificity also reaches the average level of the other algorithms.

4. CONCLUSIONS

In conclusion, we propose the MESPF framework to explore the law of epilepsy EEG signals. From frequency, channel and time, we, respectively, build information reconstruction space based on wavelet packet decomposition, graph state encoder based on graph convolution network, and space-time predictor based on gated recurrent unit. We make full use of advantages of different methods to build an efficient seizure prediction framework (MESPF). MESPF has achieved better results in classification of pre-ictal and inter-ictal than existing methods. This integrated method of multidimensional and multi-method epileptic EEG provides a more novel idea for peers to study biomedical signals.

As far as the research is concerned, we take frequency, space (channel), and time into consideration. However, there is still a lot of work to be done. In our next step, we will consider more features, such as multiple spikes, approximate entropy, information entropy, fuzzy entropy, etc. Finally, we hope to build a more accurate and stable intelligent framework for seizure prediction by continuously mining surface meaning and internal correlation of epileptic EEG.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.physionet.org/content/chbmit/1.0.0/.

AUTHOR CONTRIBUTIONS

XC: conceptualization, methodology, software, and writing—reviewing and editing. YZ: conceptualization, methodology, supervision, and funding acquisition. CD: validation, formal analysis, and data curation. SS: conceptualization, methodology, and writing—reviewing and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 61572300, 81871508, and 61773246); Taishan Scholar Program of Shandong Province of China (No. TSHW201502038); Major Program of Shandong Province Natural Science Foundation (No. ZR2018ZB0419).

REFERENCES

Adeli, H., Ghosh-Dastidar, S., and Dadmehr, N. (2007). A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. *IEEE Trans. Biomed. Eng.* 54, 205–211. doi: 10.1109/TBME.2006.886855

Affes, A., Mdhaffar, A., Triki, C., Jmaiel, M., and Freisleben, B. (2019). "A convolutional gated recurrent neural network for epileptic seizure prediction," in *International Conference on Smart Homes and Health Telematics* (Cham: Springer), 85–96. doi: 10.1007/978-3-030-32785-9_8

Ahmed, S. N. (2005). Epileptic seizures and epilepsy. *Epilepsia* 46, 1700–1701. doi: 10.1111/j.1528-1167.2005.00273_3.x

Al Ghayab, H. R., Li, Y., Siuly, S., and Abdulla, S. (2019). Epileptic seizures detection in EEGs blending frequency domain with information gain technique. *Soft Comput.* 23, 227–239. doi: 10.1007/s00500-018-3487-0

Alotaiby, T. N., Abd El-Samie, F. E., Alshebeili, S. A., Aljibreen, K. H., and Alkhanen, E. (2015). "Seizure detection with common spatial pattern and support vector machines," in *2015 International Conference on Information and Communication Technology Research (ICTRC)* (Abu Dhabi). doi: 10.1109/ICTRC.2015.7156444

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., et al. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*. doi: 10.3115/v1/D14-1179

Dash, D. P., Kolekar, M. H., and Jha, K. (2020). Multi-channel EEG based automatic epileptic seizure detection using iterative filtering decomposition and hidden markov model. *Comput. Biol. Med.* 116:103571. doi: 10.1016/j.compbiomed.2019.103571

Defferrard, M., Bresson, X., and Vandergheynst, P. (eds.). (2016). "Convolutional neural networks on graphs with fast localized spectral filtering," in Advances in Neural Information Processing Systems (Barcelona, New York, NY: Curran Associates), 3844–3852.

Guerrero-Mosquera, C., Trigueros, A. M., Franco, J. I., and Navia-Vázquez, A. (2010). New feature extraction approach for epileptic EEG signal detection using time-frequency distributions. *Med. Biol. Eng. Comput*. 48, 321–330. doi: 10.1007/s11517-010-0590-5

Holmes, G. L. (1984). *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*.

Hyvarinen, A., Oja, E., and Karhunen, J. (2001). *Independent Component Analysis*. New York, NY: Wiley.

Jia, W., Gao, S., Gao, X., and Yang, F. (2004). The progress in epileptic seizure prediction. *J. Biomed. Eng.* 21:325. doi: 10.3321/j.issn:1001-5515.2004.02.038

Khan, Y. U., Rafiuddin, N., and Farooq, O. (2012). "Automated seizure detection in scalp EEG using multiple wavelet scales," in *2012 IEEE International Conference on Signal Processing, Computing and Control* (IEEE, Waknaghat), 1–5. doi: 10.1109/ISPCC.2012.6224361

Kipf, T. N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. 5th International Conference on Learning Representations, ICLR 2017, Toulon, France. *arXiv preprint arXiv:1609.02907*.

Kiranyaz, S., Ince, T., Zabihi, M., and Ince, D. (2014). Automated patient-specific classification of long-term electroencephalography. *J. Biomed. Inform*. 49, 16–31. doi: 10.1016/j.jbi.2014.02.005

Litt, B., Esteller, R., Echauz, J., D'Alessandro, M., Shor, R., Henry, T., et al. (2001). Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron* 30, 51–64. doi: 10.1016/S0896-6273(01)00262-8

Pramod, S., Page, A., Mohsenin, T., and Oates, T. (2014). Detecting epileptic seizures from EEG data using neural networks. *arXiv preprint arXiv: 1412.6502*.

Shanir, P., Khan, Y. U., and Farooq, O. (2015). "Time domain analysis of EEG for automatic seizure detection," in *National Conference on Emerging Trends in Electrical and Electronics Engineering (ETEEE-2015)*. (Jamia Millia Islamia, New Delhi).

Shanir, P. M., Khan, K. A., Khan, Y. U., Farooq, O., and Adeli, H. (2018). Automatic seizure detection based on morphological features using one-dimensional local binary pattern on long-term EEG. *Clin. EEG Neurosci*. 49, 351–362. doi: 10.1177/1550059417744890

Sharma, P., Khan, Y. U., Farooq, O., Tripathi, M., and Adeli, H. (2014). A wavelet-statistical features approach for nonconvulsive seizure detection. *Clin. EEG Neurosci*. 45, 274–284. doi: 10.1177/1550059414535465

Solaija, M. S. J., Saleem, S., Khurshid, K., Hassan, S. A., and Kamboh, A. M. (2018). Dynamic mode decomposition based epileptic seizure detection from scalp EEG. *IEEE Access* 6, 38683–38692. doi: 10.1109/ACCESS.2018.2853125

Tessy, E., Shanir, P. M., and Manafuddin, S. (2016). "Time domain analysis of epileptic EEG for seizure detection," in *2016 International Conference on Next Generation Intelligent Systems (ICNGIS)* (IEEE, Kottayam), 1–4. doi: 10.1109/ICNGIS.2016.7854034

Yuan, Q., Zhou, W., Liu, Y., and Wang, J. (2012). Epileptic seizure detection with linear and nonlinear features. *Epilepsy Behav*. 24, 415–421. doi: 10.1016/j.yebeh.2012.05.009

Yuan, Y., Xun, G., Jia, K., and Zhang, A. (2018). A multi-view deep learning framework for EEG seizure detection. *IEEE J. Biomed. Health Inform*. 23, 83–94. doi: 10.1109/JBHI.2018.2871678

Zhang, H., Meng, Q., Meng, B., Liu, M., and Li, Y. (2018). "Epileptic seizure detection based on time domain features and weighted complex network," in *International Conference on Intelligent Computing* (Springer, Wuhan), 483–492. doi: 10.1007/978-3-319-95933-7_57

Zhou, M., Tian, C., Cao, R., Wang, B., Niu, Y., Hu, T., et al. (2018). Epileptic seizure detection based on EEG signals and cnn. *Front. Neuroinform*. 12:95. doi: 10.3389/fninf.2018.00095

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 27 August 2021
doi: 10.3389/fonc.2021.555277

Bioelectric Impedance Analysis Test Improves the Detection of Prostate Cancer in Biopsy Candidates: A Multifeature Decision Support System

Riccardo Bartoletti[1]*, *Alberto Greco*[2], *Tommaso Di Vico*[1], *Jacopo Durante*[1], *Vincenzo Ficarra*[3], *Enzo Pasquale Scilingo*[2] and *Gaetano Valenza*[2]

[1] Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy, [2] Department of Information Engineering & Research Center "E. Piaggio", University of Pisa, Pisa, Italy, [3] Department of Human and Paediatric Pathology "Gaetano Barresi" University of Messina, Messina, Italy

OPEN ACCESS

Edited by:
Giuseppe Luigi Banna, United Lincolnshire Hospitals NHS Trust, United Kingdom

Reviewed by:
M. Carmen Martinez-Bisbal, University of Valencia, Spain
Radosław Mlak, Medical University of Lublin, Poland

***Correspondence:**
Riccardo Bartoletti
riccardo.bartoletti@hotmail.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 27 April 2020
Accepted: 28 July 2021
Published: 27 August 2021

Citation:
Bartoletti R, Greco A, Di Vico T, Durante J, Ficarra V, Scilingo EP and Valenza G (2021) Bioelectric Impedance Analysis Test Improves the Detection of Prostate Cancer in Biopsy Candidates: A Multifeature Decision Support System. Front. Oncol. 11:555277. doi: 10.3389/fonc.2021.555277

Prostate cancer (PCa) gold-standard diagnosis relies on prostate biopsy, which is currently overly recommended since other available noninvasive tools such as prostate-specific antigen (PSA) multiparametric MRI (mMRI) showed low diagnostic accuracy or high costs, respectively. The aim of the study was to determine the accuracy of a novel Bioelectric Impedance Analysis (BIA) test endorectal probe for the selection of patients candidate to prostate biopsy and in particular the clinical value of three different parameters such as resistance (R), reactance (Xc), and phase angle (PA) degree. One-hundred twenty-three consecutive candidates to prostate biopsy and 40 healthy volunteers were enrolled. PSA and PSA density (PSAD) determinations, Digital Rectal Examination (DRE), and the novel BIA test were analyzed in patients and controls. A 16-core prostate biopsy was performed after a mMRI test. The study endpoints were to determine accuracy of BIA test in comparison with PSA, PSAD levels, and mMRI and obtain prostate cancer (PCa) prediction by BIA test. The Mann-Whitney *U*-test, the Wilkoxon rank test, and the Holm-Bonferroni's method were adopted for statistical analyses, and a computational approach was also applied to differentiate patients with PCa from those with benign disease. Combined PSA, PSAD, DRE, and trans-rectal ultrasound test failed to discern patients with PCa from those with benign disease (62.86% accuracy). mMRI PIRADS ≥3 showed a sensitivity of 83% and a specificity of 59%. The accuracy in discerning PCa increased up to 75% by BIA test (sensitivity 63.33% and specificity 83.75%). The novel finger probe BIA test is a cheap and reliable test that may help to improve clinical multifeature noninvasive diagnosis for PCa and reduce unnecessary biopsies.

Keywords: prostate cancer diagnosis, bioelectric impedance analysis, prostate specific antigen (PSA), prostate-specific antigen density (PSAD), multiparametric MRI, computational statistical analysis

INTRODUCTION

Prostate cancer (PCa) diagnosis necessarily implies the use of prostate biopsy which is an invasive procedure burdened by potentially relevant complications such as bleeding and systemic infection. On the other hand, currently available noninvasive diagnostic tools seem to be unable to reduce the number of unnecessary biopsies. The decision-making process is mainly based on total prostate-specific antigen (PSA) values and the results of multiparametric MRI (mMRI) (1, 2). PSA levels alone are often unable to differentiate PCa from benign prostate hyperplasia (BPH) while the combination of total PSA and DRE, as well as the combination of PSA, DRE, and trans-rectal ultrasound (TRUS) improve the cancer detection rate to 50% (3, 4). mMRI in naïve patients remains of difficult application due to its high costs and the high number of men who need to be investigated in every day clinical practice, although its accuracy has been improved by the PIRADS V2 score classification (5). Therefore, there is a need for alternative noninvasive tools improving the selection of patient candidate for prostate biopsies. Previous studies on Bioelectric Impedance Analysis (BIA) revealed enthusiastic results mainly in patients with aggressive cancers (6, 7). Phase-sensitive instruments are able to simultaneously measure resistance (R), reactance (Xc), and provide the phase angle degree (PA). Very low PA values indicate cells with altered electrical activity due to different intracellular content, DNA, and water in cancers (8) (**Figure 1**).

Studies previously conducted on PCa characterization were limited by the applicability of proper probes on the prostate surface and the gland anatomic location inside the pelvic bone girdle (9–11). With the aim to reduce previous limitations in the applications of BIA test, we tested the performance of a novel endorectal probe in a series of consecutive patient candidate to prostate biopsy for suspicious PCa.

The objectives of the present study were to test the accuracy of BIA test to detect prostate cancer by analyzing three different parameters such as R, Xc, and PA, to evaluate if the proposed BIA methodology has to be further optimized to obtain clinically meaningful results and to develop a multifeature decision support system including BIA test parameters for the prediction of prostate cancer.

MATERIALS AND METHODS

Patient Selection

All the patients who were candidates for a prostate biopsy for suspicious prostate cancer were consecutively and prospectively selected in the timeframe between March and September 2018. Presumptive diagnosis was based on persistently raised total PSA value (>4 ng/ml) and/or suspicion of cancer at DRE. Patients younger than 45 years and those affected by other neoplasms, electrolyte imbalance, and liver diseases were excluded from the study to avoid the risk of confounding factors. Moreover, patients with declared allergies to antibiotics and/or other compounds such as lidocaine, were excluded from the study. A total of 123 patients with persistently high total PSA levels and negative DRE underwent to mMRI and were classified in accordance with the PIRADS V2 system (5). Moreover, a group of young healthy volunteers were collected from a series of patient candidates to circumcision and selected with the same exclusion criteria. Healthy voluntaries were enrolled if total PSA value was <4 ng/ml, and no earlier history of prostate diseases or prostate-related symptoms was referred. Subjects included in the control group were not age-matched selected due to both the risk of developing familiar prostate cancer also in relatively young men (over45 years old) and the prostate growth that individually starts at the age of 30 but become usually symptomatic after 50 years of age (12). Patients who had undergone to prostate biopsy were age matched and had comparative risks of developing BPH or PCa. All patients selected for a prostate biopsy had a clinical suspicion of prostate cancer due to persistently increased PSA serum levels, and/or gland indurations at DRE, and/or familiar history of prostate cancer. Diagnosis of prostate cancer was performed on the basis of the pathological response after prostate biopsy. The study protocol was developed in accordance with the STROBE Statement and approved by Internal Review Committee (IRC) (1251/2017) and

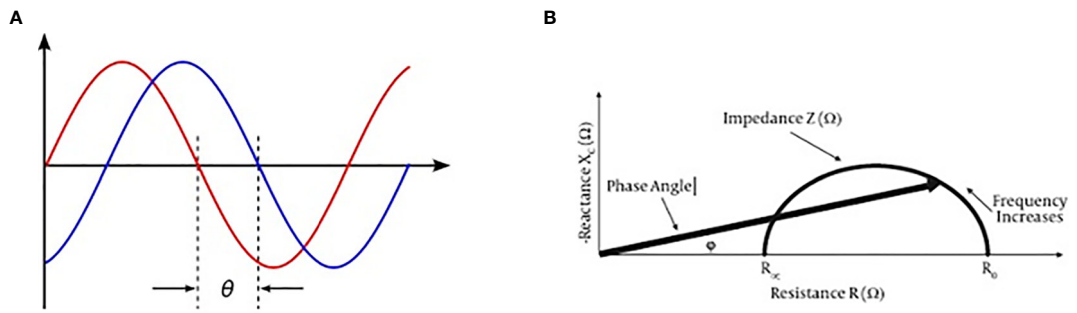


FIGURE 1 | BIA test current voltage and phase angle definition. **(A)** Phase angle consists in a ratio between current out-of-phase and current phase, and it may be expressed by an angle value (ϕ). **(B)** Reactance and resistance measurement by BIA test may express a phase angle value.

then registered (NCT03428087) (13). All patients and healthy volunteers provided their preliminary approval to participate in the study by signing an informed consent form.

For every patient, the following prebioptical parameters were collected: age, Body Mass Index (BMI), baseline total PSA (ng/ml), digital rectal examination (DRE), prostate volume estimated during TRUS examination, PSA density, and PIRADS score when available.

The BIA Test and the New Endorectal Probe

All patients underwent a BIA test using a new endorectal "finger probe" before to perform prostate biopsy. The Akern's BIA tester is tested and validated instrument and was previously used to measure the biometric parameters (BIA 101 ASE®, Akern Srl, Italy) (14). The BIA test was provided with the patient placed in a left flank position as normally adopted for DRE and prostate biopsy procedures. The negative pole electrodes (red) were placed at the base of penile shaft and at the coccyx apex while the positive ones (black) were placed at half inch over the pubic bone and at the novel "finger probe", respectively (**Figure 2**).

The electrodes placement was done to create a restricted electric field in the prostate area and get more reliable results in terms of sensitivity as demonstrated by other authors (9–11, 15). The novel probe was conceived to test the prostate tissue directly and consists of an electrode placement over a single rubber finger glove tip wearable over the rubber gloves normally used to detect prostate abnormalities. Carbon fibers are fixed at the tip of the probe, passed into the rubber finger, and connected to the "receptor"-positive pole BIA electrode. The use of flat and tender fibers other than rigid sensors was planned to allow an easy and sensitive concomitant palpation of the prostate gland. The BIA tester automatically calculated R, Xc, and PA, and different registrations have been made for the two prostate lobes. Because the location of possible cancer tissue is not known a priori, measurements from the BIA test were averaged between the two lobes for further analyses.

Prostate Biopsy

Transrectal prostate biopsy was performed after combined antibiotic prophylaxis with ceftriaxone single shot and oral fosfomycin lasting for a couple of days and local anesthesia with 10% of 5 ml lidocaine. All the patients received a cleansing enema 2 hours prior the biopsy procedure and signed an informed consent to the procedure. In all cases, at least 16 cores were systematically taken with systematic criteria except for the patients with PIRADS V2 score >3 who received adjunctive cores in relation to the number of prostate gland sites described at mMRI (5). Prostatic cores were embedded in formalin solution and then analyzed by two uropathologists. The presence of cancer or other diseases were documented as well as the Gleason score classification in the case of cancer diagnosis confirmation.

Statistical Analysis, ROC Curve, and SVM Classification

All BIA measures were normalized by dividing their value by the prostate volume, which was estimated during the TRUS examination. Therefore, the BIA test was not dependent by the volume of the prostate, which can be significantly affected by the presence of cancer.

Biomarker samples were statistically described using median and median absolute deviation (MAD) given the nonnormality of the majority of samples demonstrated using Kolmogorov-Smirnov tests (16). Accordingly, the Mann-Whitney *U*-test was used to statistically compare continuous variables from two different groups (e.g., patients vs. controls), whereas the Wilcoxon signed-rank test was used to compare differences in paired data (17, 18).

All *p*-values were corrected for multiple comparisons following the Holm-Bonferroni's method (19).

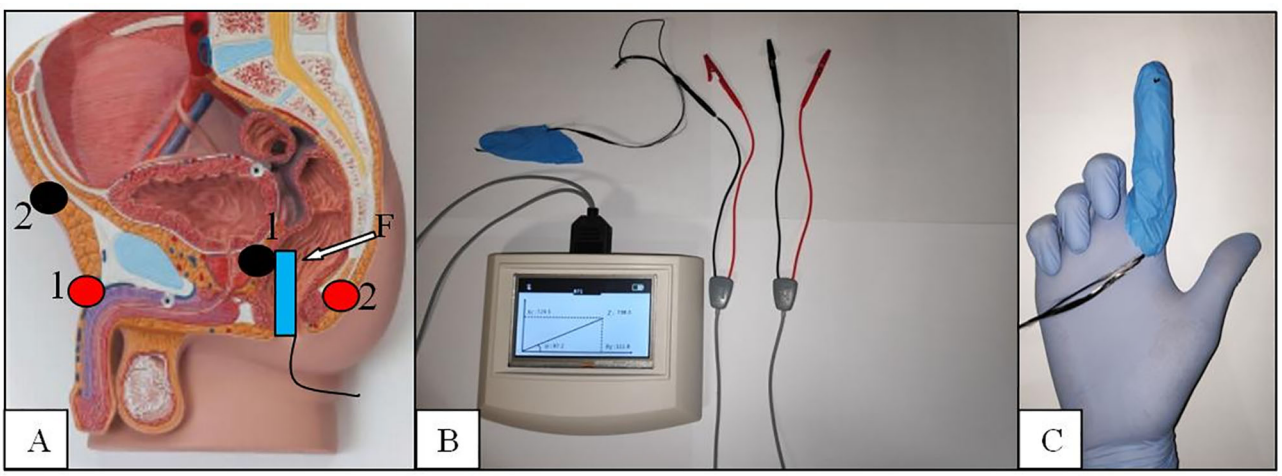


FIGURE 2 | The BIA endorectal probe. **(A)** Electrodes placement red/black 1:1 and 2:2. **(B)** The BIA tester. **(C)** The finger probe with the carbon fibers placed at the tip of the second finger.

A receiver operating characteristic (ROC) curve analysis was performed on PIRADS V2 scores gathered from 123 patients by pairing false-positive rates (1-specificity) and true-positive rates (sensitivity) at different PIRADS V2 score thresholds (5, 20).

To maximize a direct clinical applicability of the proposed study and move to a clinical evaluation at a single-patient level, we implemented a multifeature computational approach that takes into account all features, combine them through a particular mathematical function, and automatically estimate the multidimensional threshold to be used to make a clinical decision for the prediction of cancer presence. The computational methodology is quite common in the bioengineering field and is named support vector machine (SVM). We further extend the implementation of such a decision support system by integrating the so-called recursive feature elimination (RFE) approach. This scores each patient's feature such that it is possible to rank and select the most informative clinical information for the automatic discrimination of patients with prostate cancer and BPH.

The proposed SVM model combining standard PCa biomarkers including BMI, PSAD, and PSA, together with BIA-related parameters was calibrated using data from $N = 123 - 1 = 122$ subjects and then tested using data from the Nth subject for model validation. This validation procedure has been iterated N times following the so-called leave-one-subject out procedure (LOSO), where the calibration and validation sets randomly change at each iteration. The sensitivity and specificity of the proposed multiparametric approach are then calculated after N iterations, based on the observation of true positives/negatives and false positives/negatives.

RESULTS

One hundred-forty men candidate to prostate biopsy for clinical suspicion of PCa were enrolled during the study period. No patients had relevant complication after the biopsy except for persistent bleeding in seminal fluid lasting for at least 1 month. Patients with total PSA levels of <4 ng/ml presented suspicion of cancer at DRE and 6 out 15 PIRADS >3 at mMRI. Cancer was detected in four out of 15 cases. PIRADS score >3 was found in three out of four subjects with PCa. Cancer was found in 31 out of 64 patients with total PSA between 4.1 and 10 ng/ml. PIRADS score >3 was found in 34 out of 58 cases, but only 22 of them presented association with PCa. Similarly, PCa was found in 21 out of 61 patients with total PSA >10 ng/ml. MMRI confirmed the presence of cancer in 18/25 patients although resulted positive in 31 out of 54 subjects. In 60 (42.8%) cases (median BMI, 26.25; IQR, 24.87–28.7), the biopsies resulted positive for prostate cancer while in the remaining 80 (57.2%) cases (median BMI, 25.75; IQR, 24.17–27.87), a nonneoplastic prostatic condition (BPH or inflammation) was diagnosed. ROC curve analysis performed on PIRADS V2 scores obtained from 123 mMRI of patients who underwent prostate core biopsy (no healthy volunteers have been included) showed a major threshold score or equal to 3 *a* sensitivity of 83% and a specificity of 59% and VPP and VPN were 61% and 82%, respectively (**Figure 3**).

By analyzing patients with prostate cancer, in 21 (35%) cases, the disease involved a single lobe (11 the right side and 10 the left side). Conversely, in the other 39 (65%) cases, both lobes were involved by the tumor. Therefore, according to D'Amico risk classification, 31 (51.6%) patients were classified as low risk, nine (15%) as intermediate, and 20 (33.4%) as high risk (21). The 40 young healthy volunteers showed a median age of 37 years (MAD = 4) and a median BMI of 25 (MAD = 1.1). The median prostate volume was 19.23 cm^3 (MAD = 6.5) with a median PSA density of 0.05 (MAD = 0.03). Comprehensive descriptive statistics of patients' characteristics stratified according to the biopsy results are reported in **Table 1**.

Inferential statistics between patients with PCs, BPH, and controls are reported in **Table 2**.

Comparing patients with PCa vs. controls, differences in age, PSA, prostate volume, and PSA density were found. Same statistical differences were found comparing BPH *vs.* controls. Concerning BPH *vs.* patients with PCa, significant differences were found in the prostate volume exclusively ($p < 0.01$); hence, PSAD analysis exclusively was retained for further analyses.

Evaluation of BIA Test Parameters

Table 2 summarizes and compares BIA test parameters gathered from patients with prostate cancer, benign prostatic disease, and healthy volunteers. While no significant differences between groups were found on the BIA parameter PA, significant differences were found in comparing BPH *vs.* controls using R ($p < 0.01$), as well as comparing controls *vs.* patients with PCa and controls *vs.* BPH using Xc ($p < 0.05$).

Concerning the statistical comparison between the three bioimpedance test measurements from the right and left sides of the prostate, we split the dataset in three subsets: (i) a subset of patients with right-sided prostate cancer (**Table 3**); (ii) a subset

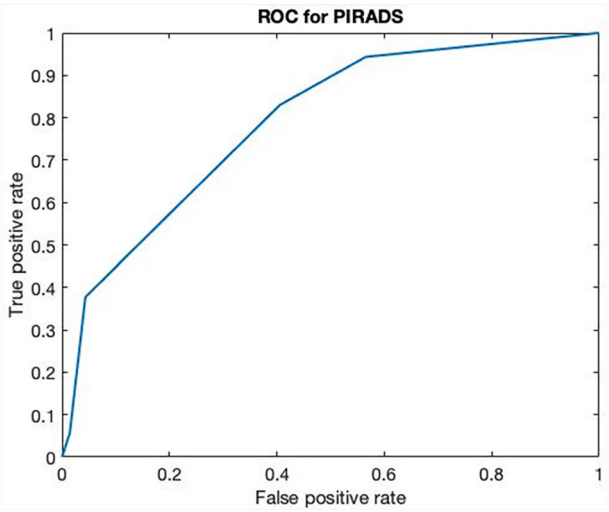


FIGURE 3 | ROC curve of multiparametric MRI. ROC curve of multiparametric MRI on 123 patients with clinical suspicion of PCa (AUC 0.7893).

TABLE 1 | Patients' characteristics.

	PCa	BPH	Healthy controls
Patients (*n*.)	60	80	40
Age (years)	70 ± 5	69 ± 5	37 ± 4
BMI (kg/m^2)	26.01 ± 1.51	26.15 ± 2.15	25 ± 1.10
Prebiopsy total PSA value (ng/ml)			
<4	4	11	35
4–10	31	33	0
>10	25	36	0
Prostate nodule/s at DRE (*n*. pts)			
Right	18	14	0
Left	10	7	0
Bilateral	8	4	0
Negative	24	55	35
Prostate volume at TRUS (ml)	41.48 ± 11.88	53.78 ± 13.30	19.23 ± 6.50
PSA density (PSA/volume)	0.25 ± 0.13	0.17 ± 0.07	0.05 ± 0.03
Prostate biopsy	YES	YES	NO
Low-risk PCa (*n*.)	31	–	–
Intermediate-risk PCa (*n*.)	9	–	–
High-risk PCa (*n*.)	20	–	–
Patients who underwent to prebiopsy mMRI (*n*.)	60	63	–
Patients with mMRI PIRADS ≥3 (*n*./%)	43/71.6	28/44.4	–
Patients with mMRI PIRADS ≥4 (*n*./%)	15/25	2/3.1	–

MAD, median absolute deviation.
Descriptive ranges are expressed as (median ± MAD).

of patients with left-sided prostate cancer (**Table 4**); and (iii) a subset of patients with both-sided prostate cancer (**Table 5**).

It is worthwhile noting that the R of the right side of the prostate was significantly lower than the left side in the left- and both-sided cancer patient group. Moreover, PA of the left side of the prostate was significantly lower than the right side in the both-sided cancer patient group.

Evaluation of a Multifeature Clinical Decision Support System for Diagnosis at a Single-Subject level

Using the parameter set comprising BMI, PSA density, PSA, AGE, R, PA, and Xc, we built an SVM multifeature computational model as described above and derived cancer recognition accuracy, sensitivity, specificity, positive, and negative predictive values (PPV and NPV) (**Table 6**).

For each feature, we performed a ROC curve analysis to evaluate the performance of each single feature in discriminating between patient with PCa from those with benign disease. From each ROC curve, we computed the area under the curve (AUC) and the related 95% confidence interval (95% CI AUC).

TABLE 3 | Median and MAD values of the bioimpedance features calculated on the right and left sides of the prostate in the right-sided prostate cancer group.

Feature	MEDIAN right side	MAD right side	MEDIAN left side	MAD left side	*p*-Value
R	46	9.6	46.9	9.45	n.s.
Xc	8.7	4.1	9.05	3.95	n.s.
PA	9.45	2.9	9.5	2.85	n.s.

The last column shows the results of the Wilcoxon signed-rank test for paired data between the left and right sides of the prostate (n.s., nonsignificant p-value).

TABLE 4 | Median and MAD values of the bioimpedance features calculated on the right and left sides of the prostate in the left-sided prostate cancer group.

Feature	MEDIAN right side	MAD right side	MEDIAN left side	MAD left side	*p*-Value
R	40.5	7.9	40.8	9.55	**<0.01**
Xc	10.35	3.15	10	3.05	n.s.
PA	14.7	3.8	14.35	3.25	n.s.

The last column shows the results of the Wilcoxon signed-rank test between the left and right sides of the prostate (n.s., nonsignificant p-value).
Bold values indicate significant p-values lower than 0.05.

TABLE 5 | Median and MAD values of the bioimpedance features calculated on the right and left sides of the prostate in the both-sided prostate cancer group.

Feature	MEDIAN right side	MAD right side	MEDIAN left side	MAD left side	*p*-Value
R	40.9	10.35	41.7	11.5	**<0.001**
Xc	8.4	2.7	8.5	2.75	n.s.
PA	12.4	3.9	11.55	3.5	**<0.05**

The last column shows the results of the Wilcoxon signed-rank test for paired data between the left and right sides of the prostate (n.s., nonsignificant p-value).
Bold values indicate significant p-values lower than 0.05.

TABLE 2 | Statistical comparison between patients who underwent to prostate biopsy (cancer and BPH) and controls.

Feature	MEDIAN (PCa)	MAD (PCa)	MEDIAN (BPH)	MAD (BPH)	MEDIAN (HC)	MAD (HC)	*p*-Value PCa *vs.* HC	*p*-Value BPH *vs.* HC	*p*-Value BPH *vs.* PCa
AGE	70	5	69	5	37	4	**<0.001**	**<0.001**	n.s.
BMI	26.01	1.51	26.15	2.15	25	1.1	n.s.	n.s.	n.s.
PSA	8.985	2.535	9.11	3.915	0.87	0.45	**<0.001**	**<0.001**	n.s.
Prostate volume at TRUS	41.475	11.88	53.775	13.295	19.23	6.5	**<0.001**	**<0.001**	**<0.001**
PSA density	0.25	0.13	0.17	0.07	0.05	0.03	**<0.001**	**<0.001**	n.s.
R	41.425	10.325	47.125	12.45	34.25	10.95	n.s.	**<0.01**	n.s.
Xc	8.875	3.3	10.425	2.7	6.2	1.55	**<0.05**	**<0.001**	n.s.
PA	12.3	3.875	13.025	3.45	12.15	4.15	n.s.	n.s.	n.s.

n.s., nonsignificant p-value; MAD, median absolute deviation; BMI, Body Mass Index; PSA, prostate-specific antigen; TRUS, trans-rectal ultra sound; RES, resistance; REA, reactance; PHASE, phase angle; PCa, prostate cancer; BPH, benign prostate hyperplasia; HC, healthy controls.
Bold values indicate significant p-values lower than 0.05.

TABLE 6 | Feature ranking according to the RFE criterion.

Feature ranking
BMI
PSA density
R
PSA
PA
AGE
Xc

*Feature ranges for healthy controls and PCa patients are reported in **Tables 1**, **2**.*

According to the 95% CI AUC, we tested also if the AUC was statistically greater than 0.5, i.e., the chance performance value (AUC *p*-value).

Moreover, we identified the threshold cutoff point associated with the best accuracy (*Accuracy* Cut-off point) according to the Youden's index (Sensitivity-Specificity-1) and the related sensitivity and specificity values (**Table 7**).

The average prediction accuracy achieved is shown in **Figure 2**, with a final score as of 75.00% (**Figure 4**).

This was obtained mathematically combining the following four parameters that were identified as clinically relevant: BMI, PSA density, R, and PSA. A comprehensive, ranked clinical feature list for this decision support system is reported in **Table 8**, while the corresponding confusion matrix is in **Table 9**. Sensitivity and specificity of the PCa prediction *vs.* BPH were 63.33% and 83.75%, respectively. The PPV and NPV were 74.51% and 75.28%, respectively. It is worthwhile noting that the resistance, averaged between the right and left prostate lobes, is one of the most informative features and gives a significant contribution to achieve the 75.00% of accuracy.

Importantly, as a counterproof, we obtained a significant decrease in the PCa prediction accuracy of 62.86% while repeating the same SVM-based computational procedure using a feature set that does not include the three biometric measures. As expected, the most informative subfeatures set included BMI, PSA density, and PSA.

TABLE 8 | Comprehensive ranked clinical feature list of the most accurate subset of features: BMI, PSA density, RES, and PSA.

BMI, PSA density, R, PSA	Cancer	BPH
Cancer predicted	**63.33%**	16.25%
BPH predicted	36.67%	**83.75%**

Bold values indicate significant p-values lower than 0.05.

TABLE 9 | Confusion matrix of the most accurate sub-set of features excluding BIA parameters.

BMI, PSA density, PSA	Cancer	BPH
Cancer predicted	**51.67%**	28.75%
BPH predicted	48.33%	**71.25%**

BMI, Body Mass Index; PASD, prostate-specific antigen density; PSA, prostate-specific antigen.
Bold values indicate significant p-values lower than 0.05.

DISCUSSION

BIA of different tissues was originally investigated by Geddes and Baker in the 1960s (22). They carried out an electrical measurement on living tissues demonstrating different values of resistivity. From that period, the BIA test have been used for various purposes such as the lean and fat body mass calculation and other medical applications like skin and breast cancer diagnosis (23–25). Halter et al. measured electrical properties of "*ex vivo*" prostate tissues with the aim of future applications for PCa noninvasive diagnosis. They realized that PCa, BPH, nonhyperplastic glandular tissue, and stromal tissue had different conductivity at all frequencies while mean cancer permittivity was significantly greater than that of benign tissues at high frequencies (15). Other authors demonstrated that best results for cancer diagnosis by BIA test were obtained by measuring the tissue phase angle. Low phase angle suggests cell death or decreased cell integrity, whereas higher phase angle suggests healthy cell (26, 27). A low phase angle has been associated with an impaired outcome in tumor diseases such as pancreatic cancer, colorectal cancer, and lung cancer (6, 7, 15). Tyagi et al. recently demonstrated that low phase angle values measured by BIA test allow for discriminating PCa patients from matched controls and those with advanced stage and high-risk PCa in particular. They investigated a group of subjects using the BIA electrode placement on the right upper and right lower limb. On the other hand, all PCa-diagnosed subjects had a total PSA increased values and other

TABLE 7 | Statistical analysis for each feature.

Feature	AUC	95% CI AUC	*Accuracy* cutoff point	Sensitivity	Specificity	AUC *p*-value
AGE	0.5597	0.4641-0.6553	75	0.825	0.267	0.110
BMI	0.51092	0.4141-0.6078	34.2	0.988	0.0333	0.413
PSA	0.49015	0.3931-0.5871	42.74	0.988	0.067	0.579
PSA density	0.62337	0.5311-0.7156	0.25	0.812	0.533	**0.00437**
R	0.52737	0.4308-0.6239	2.1185	0.925	0.167	0.289
Xc	0.50038	0.4034-0.5973	1.1439	0.975	0.0833	0.497
PA	0.55362	0.4578-0.6494	0.5291	0.925	0.200	0.136

The area under the curve (AUC) was calculated from the ROC curve as well as the related 95% confidence interval (95% CI AUC). According to the 95% CI AUC, the AUC was statistically greater than 0.5, i.e., the chance performance value (AUC p-value).
Bold values indicate significant p-values lower than 0.05.

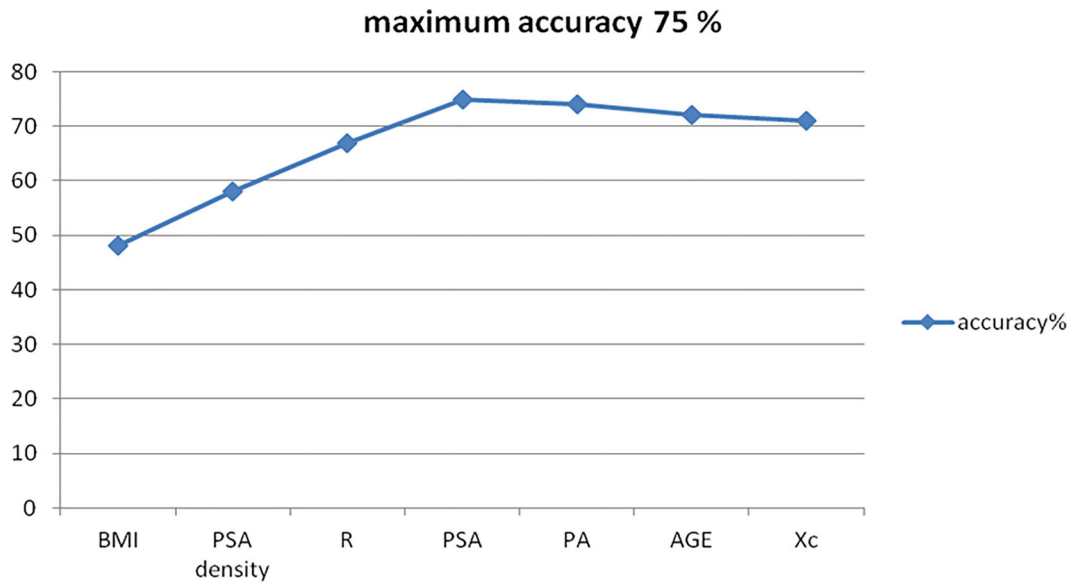


FIGURE 4 | BIA test accuracy for the diagnosis of prostate cancer.

concomitant diseases excluded to avoid the risk of false-positive results (10). Similarly, Khan developed a new composite impedance metrics method with a nine-electrode microendoscopic probe. This novel device was tested on "*in vivo*" and "*ex vivo*" prostate tissue either intraoperatively or after the prostate removal in patients who underwent surgery for PCa or BPH. The results obtained demonstrated a predictive accuracy of 90.79% for PCa (11). For these reasons, we provided an alternative electrode placement and a restricted locoregional electric field in order to improve the BIA test sensitivity and specificity and reduce possible confounding factors. The finger probe allows the obtainment of the tissue resistance, reactance, and phase angle measurements directly from the prostate gland surface through a restricted electric field generated into the pelvic bone girdle. However, prostate tissue presents an extreme variability of electrical absorption due to its water and/or stromal content and the presence of microcalcifications in its tissue context with subsequent false positive results.

Our results demonstrated that the finger probe is a promising, reliable, and easy-to-use tool to improve the accuracy of PCa noninvasive diagnosis together with other standard clinical parameters. In patients where PCa was diagnosed in both prostate lobes (i.e., 65% of cases), BIA PA were found significantly different between the right and left side, while seemed to be comparable when PCa was diagnosed in a single lobe. Our experimental evidences on BIA phase angles do not replicate previous findings reported in (9). This may be justified by the presence of more represented stromal tissue and/or calcifications inside the gland, as well as by the normalization procedure that we have performed prior to the statistical analyses. All BIA measures including R, Xc, and PA, in fact, were normalized by dividing their value by the prostate volume estimated during the TRUS examination to avoid biases. Without normalization, patients with BPH and with PCa vs. healthy controls showed significant differences in terms of BIA PA (p = 0.006 and 0.003, respectively), therefore confirming previous observations by Tyagi et al. (10). BIA resistance values were lower in patients with PCa although, taken alone, it seemed to be unable to differentiate cancer from noncancer patients, while it was significantly different between healthy subjects and the BPH group. BIA reactance values were significantly different between healthy subjects and patients, although taken alone were not significantly different between BPH and PCa patients.

In this sense, likewise for the PSA alone, the BIA test failed to differentiate subjects with clinical suspicion of PCa and prospectively missed the intent of avoiding unnecessary biopsies. Nevertheless, when combined with the other standard clinical parameters including patients' PSA and PSA density, BIA test provided meaningful information for discerning between PCa and BPH patients with an accuracy as high as 75% at a single patient level.

Our results indicate a good PCa prediction using a combination of the following clinical features: BMI, Age, PSA, and PSA density. In this case, sensitivity and specificity are lower than the ones associated with a combination of BMI, PSA density, R, and PSA, thus demonstrating the significant clinical information associated with BIA test.

Study limitations include the limited amount of data, especially gathered from healthy volunteers, the nonage-matched group taken as negative control due to the increased risk of developing prostate diseases in the advanced age and a fixed range of 50 mHz frequency band for the BIA.

The proposed BIA test is a cheap, easy-to-perform method helpful for the multifeature clinical and noninvasive detection of prostate cancer and may be also able to decrease the number of unnecessary biopsies. The use of a novel transrectal "finger-probe" allows to do the BIA test with a minimal discomfort for the patient, contributing to an accuracy as high as 75% for the PCa vs. BPH prediction when properly combined with BMI, total PSA, and PSA density. Interestingly, the test can be easily repeatable. Further studies by varying the BIA tester voltage frequency are necessary to improve the BIA test efficacy. The cheaper cost of the method in comparison with mMRI may be immediately attractive for low-income countries.

AUTHOR'S NOTE

Preliminary version of the manuscript has been previously preprinted at www.BiorXiV.org (https://doi.org/10.1101/2020.02.12.943829), and partial results are previously presented as an abstract at the European Urological Association Congress in Barcelona (European Urology Open Science March 2019, volume 18, issue 1, page e1810).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Internal Review Committee (IRC) (approval 1251/2017). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conceptualization: RB. Data curation: TV. Formal analysis: AG, ES, and GV. Investigation: JD and TV. Methodology: RB. Project administration: RB. Resources: RB, VF, and ES. Software: GV, AG, and ES. Supervision: RB, GV, and VF. Validation: RB and GV. Visualization: TV and AG. Writing original draft: RB and GV. Writing review edition: VF. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Siegel RI, Miller KD, Jemal A. Cancer Statistics. *Cancer J Clin* (2019) 69:7–34. doi: 10.3322/caac.21551
2. Catalona WJ, Richie JP, Ahmann FR, Hudson MA, Scardino PT, Flanigan RC, et al. Comparison of Digital Rectal Examination and Serum Prostate Specific Antigen in the Early Detection of Prostate Cancer: Results of a Multi Center Clinical Trial of 6,630 Men. *J Urol* (1994) 151:1283–90. doi: 10.1016/S0022-5347(17)35233-3
3. Brown LC, Ahmed HU, Faria R, El-Shater Bosaily A, Gabe R, Kaplan RS, et al. Multiparametric MRI to Improve Detection of Prostate Cancer Compared With Transrectal Ultrasound-Guided Prostate Biopsy Alone: The PROMIS Study. *Health Technol Assess* (2018) 22(39):1–176. doi: 10.3310/hta22390
4. Ahmed HU, El Shater Bosaily A, Brown LC, Gabe R, Kaplan R, Parmar MK, et al. Diagnostic Accuracy of Multiparametric MRI and TRUS in Prostate Cancer (PROMIS) : A Paired Validating Confirmatory Study. *Lancet* (2017) 389:815–22. doi: 10.1016/S0140-6736(16)32401-1
5. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura CJ, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *Eur Urol* (2016) 69:16–40. doi: 10.1016/j.eururo.2015.08.052
6. Norman K. Cut-Off Percentiles of Bioelectrical Phase Angle Predict Functionality, Quality of Life and Mortality in Patients With Cancer. *Am J Clin Nutr* (2010) 92:612–9. doi: 10.3945/ajcn.2010.29215
7. Gupta D. BIPA in Clinical Practice. Implications for Prognosis in Advanced Colorectal Cancer. *Am J Clin Nutr* (2004) 80:1634–8. doi: 10.1093/ajcn/80.6.1634
8. Davis MP. Bioelectric Impedance Phase Angle Changes During Hydration and Prognosis in Advanced Cancer. *Am J Hosp Palliat Care* (2009) 26:180–7. doi: 10.1177/1049909108330028
9. Mishra V, Bouyad H, Halter RJ. Electrical Impedance-Based Biopsy for Prostate Cancer Detection. *2011 IEEE 37th Annu Northeast Bioeng Conf (NEBEC)* (2011) 1:1–2. doi: 10.1109/NEBC.2011.5778712
10. Tyagi R, Mishra S, Gaur N, Panwar A, Saini D, Singh K, et al. Bioelectric Impedance Phase Angle in Carcinoma Prostate- a Hospital Based Study. *Int J Med Sci Pub Health* (2016) 5(9):1826–30. doi: 10.5455/ijmsph.2016.30122015335
11. Khan S, Mahara A, Hyams ES, Schned AR, Halter RJ. Prostate Cancer Detection Using Composite Impedance Metric. *IEE Trans Med Imag* (2016) 35(12):2513–23. doi: 10.1109/TMI.2016.2578939
12. Bartoletti R, Gavazzi A, Cai T, Mondaini N, Morelli A, Del Popolo G, et al. Prostate Growth and Prevalence of Prostatic Diseases in Early Onset Spinal Chord Injuries. *Eur Urol* (2009) 56:142–8. doi: 10.1016/j.eururo.2008.01.088
13. *STROBE Statement Strengthening the Reporting of Observational Studies in Epidemiology*. Available at: http://www.strobe-statement.org/.
14. Grimmes S, Martinsen OG. *Bioimpedance and Bioelectricity Basics. 3rd. Edition*. Elsevier Science: Switzerland (2014). Available at: https://consensus.nih.gov/1994/1994bioelectricimpedancebodyta015html.htm.
15. Halter RJ, Schned A, Heaney J, Hartov A, Paulsen KD. Electric Properties of Prostatic Tissues: Single Frequency Admittivity Properties. *J Urol* (2009) 182:1600–7. doi: 10.1016/j.juro.2009.06.007
16. Ong LD, LeClare PC. The Kolmogorov Smirnov Test for the Log-Normality of Sample Cumulative Frequency Distributions. *Health Phys* (1968) 22:570. doi: 10.2466/pr0.1968.22.2.570
17. Mann HB DW. On a Test of Whether One of Two Random Variables Is Stochastically Larger Than the Other. *Ann Math Statist* (1947) 18:50–60. doi: 10.1214/aoms/1177730491
18. Wilcoxon F. Individual Comparisons by Ranking Methods. *Biometrics Bull* (1937) 1(6):80–3. doi: 10.2307/3001968
19. Aickin M, Gensler H. Adjusting for Multiple Testing When Reporting Research Results: The Bonferroni vs Holm Methods. *Am J Public Health* (1996) 15:1107–21. doi: 10.2105/AJPH.86.5.726
20. Metz CE. Basic Principles of ROC Analysis. *Semin Nucl Med* (1978) 8:283–98. doi: 10.1016/S0001-2998(78)80014-2
21. D'Amico AV. Risk-Based Management of Prostate Cancer. *N Engl J Med* (2011) 365(2):169–71. doi: 10.1056/NEJMe1103829
22. Geddes LA, Baker LE. The Specific Resistance of Biological Material. A Compendium of Data for the Biomedical Engineer and Physiologist. *Med Biol Engin* (1967) 5(3):271–93. doi: 10.1007/BF02474537
23. Macfarlane DJ, Chan NT, Tse MA, Joe GM. Agreement Between Bioelectrical Impedance and Dual Energy X-Ray Absorptiometry in Assessing Fat, Lean and Bone Mass Changes in Adults After a Lifestyle Intervention. *J Sports Sci* (2016) 34:1176–81. doi: 10.1080/02640414.2015.1096416
24. Jossinet J, Schmitt M. A Review of Parameters for the Bioelectrical Characterization of Breast Tissue. *Ann NY Acad Sci* (1999) 873:30–41. doi: 10.1111/j.1749-6632.1999.tb09446.x

25. Braun RP, Mangana J, Goldinger S, French L, Dummer R, Marghoob AA. Electrical Impedance Spectroscopy in Skin Cancer Diagnosis. *Dermatol Clin* (2017) 35:489–93. doi: 10.1016/j.det.2017.06.009
26. Baumgartner RN, Chumlea WC, Roche AF. Bioelectric Impedance Phase Angle and Body Composition. *Am J Clin Nutr* (1988) 48:16–23. doi: 10.1093/ajcn/48.1.16
27. Selberg O, Selberg D. Norms and Correlates of Bioimpedance Phase Angle in Healthy Human Subjects, Hospitalized Patients and Patients With Liver Cyrrosis. *Eur J Appl Physiol* (2002) 86:509–16. doi: 10.1007/s00421-001-0570-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership