AI IN BIOLOGICAL AND BIOMEDICAL IMAGING

EDITED BY: Xin Gao, Lihua Li and Min Xu

PUBLISHED IN: Frontiers in Molecular Biosciences and Frontiers in Physiology

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-049-9
DOI 10.3389/978-2-88974-049-9

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

AI IN BIOLOGICAL AND BIOMEDICAL IMAGING

Topic Editors:
Xin Gao, King Abdullah University of Science and Technology, Saudi Arabia
Lihua Li, Hangzhou Dianzi University, China
Min Xu, Carnegie Mellon University, United States

Doctors Gao and Li hold patents related to artificial intelligence.

Citation: Gao, X., Li, L., Xu, M., eds. (2022). AI in Biological and Biomedical Imaging. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-049-9

Table of Contents

frontiers in Molecular Biosciences

EDITORIAL
published: 24 November 2021
doi: 10.3389/fmolb.2021.804476

Editorial: AI in Biological and Biomedical Imaging

Xin Gao[1]*, *Lihua Li*[2] *and Min Xu*[3]

[1]King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, [2]Hangzhou Dianzi University, Hangzhou, China, [3]Carnegie Mellon University, Pittsburgh, PA, United States

Keywords: biomedical imaging, bioinformatics, AI, biological imaging, machine learning

Editorial on the Research Topic

AI in Biological and Biomedical Imaging

Imaging is the visual representation of structures and functions of objects, such as biological molecules, biological ultrastructures, tissues, and the spatial organizations of the objects. It is also an indispensable step towards diagnostics and therapeutics in modern medicine. For example, during the current pandemic caused by COVID-19, CT-scans have been used, in addition to nucleic acid detection, as a main criterion for diagnostics. Unlike computers, the human brain has a remarkably strong ability to understand and interpret the information obtained from imaging data, more so than from interpreting numerical or textual data. On the other hand, AI methods may produce more objective and highly reproducible analysis results with increased automation. Therefore, it is beneficial to develop AI methods to complement manual image analysis.

Imaging is playing an increasingly significant role in both biological and biomedical sciences. With technologies including optical microscopy, fluorescence microscopy, electron tomography, nuclear magnetic resonance, single particle cryo-EM, and X-ray crystallography, biological imaging has provided rich information about biological systems and molecules at various resolutions, all the way from tissue-level, to cellular-level, to organelle-level, to macromolecular-level, to small-molecular-level, and to atomic-level. Imaging also has many diagnostic and therapeutic applications in medicine with different modalities, such as ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and optical coherence tomography (OCT). Such technologies can provide fast, non-invasive, painless and direct information to clinicians and physicians, which is critical to not only diagnosis, but also prognosis and treatment.

With the recent development of AI technologies, especially deep learning, the frontiers on biological and biomedical imaging have been greatly advanced. In this Research Topic, we have collected 16 high quality works on developing or applying state-of-the-art AI techniques for processing, information mining, integrating, diagnosing, comparing, and reviewing biological and biomedical imaging, with their applications in biology, diagnostics and therapeutics.

There are in total 16 papers accepted by this Research Topic. Each paper was handled by one guest editor and reviewed by at least two reviewers. We are very grateful to the reviewers in helping us select these high-quality papers for this Research Topic.

The accepted papers focus on developing AI algorithms and systems to process, analyze, interpret, and mine from both biomedical and biological imaging data. We begin by providing a thought-provoking perspective on the future of radiology diagnostic service (Seong et al.). Radiology has been a leading technology of digital transformation in healthcare, which is again at the crossroad for the next generation of transformation, possibly evolving as a one-stop integrated diagnostic service. AI promises to offer radiology new powerful new digital tools to facilitate the next transformation. This paper proposes three pathways for AI's role in radiology: (1) improving the performance of CAD, (2) improving the productivity of radiology service by AI-assisted workflow, and (3) developing

OPEN ACCESS

Edited by:
William C Cho,
QEH, Hong Kong SAR, China

Reviewed by:
Jason T. Smith,
Rensselaer Polytechnic Institute,
United States

Correspondence:
Xin Gao
xin.gao@kaust.edu.sa

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *29 October 2021*
Accepted: *08 November 2021*
Published: *24 November 2021*

Citation:
Gao X, Li L and Xu M (2021) Editorial: AI in Biological and Biomedical Imaging. Front. Mol. Biosci. 8:804476. doi: 10.3389/fmolb.2021.804476

radiomics that integrate the data from radiology, pathology, and genomics to facilitate the emergence of a new integrated diagnostic service.

We then present papers on 2D biomedical imaging data. We first discuss three papers on pathology imaging. Xu et al. proposed an effective immunohistochemistry pathology microscopic image-generation method that can generate synthetic immunohistochemistry pathology microscopic images from hematoxylin-eosin stained pathology microscopy images without any annotation. CycleGAN is adopted as the basic architecture for the unpaired and unannotated dataset. Moreover, multiple instances learning algorithms and the idea behind conditional GAN are considered to improve performance.

Liu et al. selected Ki-67-expression as the representative of molecular information. They proposed a method that can predict Ki-67 positive cells directly from H&E stained slides by a deep convolutional network model. To train this model, they constructed a dataset containing Ki-67 negative or positive cell images and background images. These images were all extracted from H&E stained WSIs and the Ki-67 expression was acquired from the corresponding IHC stained WSIs. The trained model was evaluated both on classification performance and the ability to quantify Ki-67 expression in H&E stained images.

He et al. proposed a hybrid-attention nested UNet (Han-Net), which consists of two modules: a hybrid nested U-shaped network (H-part) and a hybrid attention block (A-part). H-part combines a nested multi-depth U-shaped network and a dense network with full resolution to capture more effective features. A-part is used to explore attention information and build correlations between different pixels. With these two modules, Han-Net extracts discriminative features, which effectively segment the boundaries of not only complex and diverse nuclei but also small and dense nuclei. The comparison in a publicly available multi-organ dataset shows that the proposed model achieves the state-of-the-art performance compared to other models.

We further accepted two papers on X-ray data analysis. Yang et al. proposed a data-driven diagnostic model for hip dysplasia. Angles including CE, sharp, and Tonnis angle which are commonly measured in clinical diagnosis, are automatically obtained. Samples, each of which consists of these three angle values, are used for clustering according to their densities in a descending order. A three-dimensional normal distribution derived from the cluster is built and regarded as the parametric model for diagnosis of hip dysplasia.

Li et al. proposed an interpretable method called Deetal-Perio to predict the severity degree of periodontitis in dental panoramic radiographs. In their method, alveolar bone loss (ABL), the clinical hallmark for periodontitis diagnosis, could be interpreted as the key feature. To calculate ABL, they also proposed a method for teeth numbering and segmentation. First, Deetal-Perio segments and indexes the individual tooth via Mask R-CNN combined with a novel calibration method. Next, Deetal-Perio segments the contour of the alveolar bone and calculates a ratio for individual tooth to represent ABL. Finally, Deetal-Perio predicts the severity degree of periodontitis given the ratios of all the teeth.

We also accepted one paper on quantifying vascular density in tissue engineered constructs. Strobel et al. developed a semi-automated method that leverages machine learning to identify and quantify vascular metrics in an angiogenesis model imaged with different modalities. Their software, BioSegment, is designed to make high throughput vascular density measurements of fluorescent or phase contrast images.

We then present papers on 3D biomedical imaging data. We start from two papers on CT data analysis. Dou et al. aimed to test whether chest CT manifestation of 2019 novel coronavirus (COVID-19) can be differentiated by a radiologist or a computer-based CT image analysis system. They conducted a retrospective case-control study that included 52 laboratory-confirmed COVID-19 patients and 80 non-COVID-19 viral pneumonia patients. Their results do not support CT findings replacing microbiological diagnosis as a critical criterion for COVID-19 diagnosis.

Cai et al. developed and validated a radiomics-based nomogram to predict the prognosis of colorectal cancer (CRC). A total of 381 patients with colorectal cancer were enrolled and radiomic features were extracted from the vein phase of preoperative computed tomography. Their results show that radiomics score derived from the preoperative CT image was an independent prognostic factor and could be a complement to the current staging strategies of colorectal cancer.

We further accepted four papers on MRI data analysis. Fan et al. proposed a framework of a 3D-Mask region-based convolutional neural network (3D-Mask RCNN) computer-aided diagnosis (CAD) system for mass detection and segmentation with a comparative analysis of performance on patient subgroups with different clinicopathological characteristics. The results suggest that the 3D-Mask RCNN CAD framework has advantages over 2D-based mass detection on both the whole data and subgroups with different characteristics.

Fan et al. predicted responses to NACT in breast cancer by analyzing early changes in tumor heterogeneity modeled by longitudinal dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). Their results suggested that changes in DCE-MRI features that reflect a reduction in tumor heterogeneity following NACT could provide early prediction of breast tumor response.

Wang et al. proposed a novel model structure to capture 3D MRI images' essential information and converted them into lower dimensions. The novel CNN model they proposed could automatically differentiate the rare NMOSD from MS, especially, our model showed better performance than traditional 3D CNN models.

Gu et al. proposed a multi-head self-attention model (MSAM). By integrating the self-attention mechanism and multilayer perceptron method, the MSAM offers a promising tool to enhance the classification of Temporal Lobe Epilepsy (TLE) subtypes. The robustness of MSAM is extensively assessed with various ablation tests, which demonstrates the effectiveness and generalizability of the proposed approach.

We then included a paper on developing diagnostic algorithm for capsule endoscopy, which is a leading diagnostic tool for small bowel lesions. Kong et al. proposed a multi-task framework,

called the multi-task classification and segmentation network (MTCSN), to achieve joint learning of clearness degree and tissue semantic segmentation. Extensive experiments and ablation studies report the significant performance gains of the MTCSN over state-of-the-art methods.

We further accepted a paper on reviewing tactile perception technologies on minimally invasive surgery (MIS), which has been the preferred surgery approach owing to its advantages over conventional open surgery. As a major limitation, the lack of tactile perception impairs the ability of surgeons in tissue distinction and maneuvers. Huang et al. aimed to provide potential tactile perception methods for MIS by reviewing literatures on tactile sensing in MIS and literatures on industrial robotic tactile perception technologies, especially AI methods on tactile images.

We finish our collection of papers by including one paper on biological imaging. Zhou et al. developed a one-shot learning framework, called cryo-ET one-shot network (COS-Net), for simultaneous classification of macromolecular structure and generation of the voxel-level 3D segmentation, using only one training sample per class, from cryo-electron tomography data. Their experimental results demonstrated that COS-Net could efficiently classify macromolecular structures with small amounts of samples and produce accurate 3D segmentation at the same time.

To conclude, we thank the authors and the reviewers for their contribution to this Research Topic. We are confident that the collection of articles in this Research Topic will serve as an inspiring compendium for future AI advancement and deployment in biomedical and biological imaging fields.

AUTHOR CONTRIBUTIONS

The three authors contributed to the writing of the manuscript.

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 04 August 2020
doi: 10.3389/fmolb.2020.00183

Predict Ki-67 Positive Cells in H&E-Stained Images Using Deep Learning Independently From IHC-Stained Images

Yiqing Liu[1†], Xi Li[2†], Aiping Zheng[3†], Xihan Zhu[4], Shuting Liu[1], Mengying Hu[1], Qianjiang Luo[2], Huina Liao[2], Mubiao Liu[5], Yonghong He[1] and Yupeng Chen[6*]*

[1] Department of Life and Health, Tsinghua Shenzhen International Graduate School, Shenzhen, China, [2] Department of Gastroenterology, Peking University Shenzhen Hospital, Shenzhen, China, [3] Department of Pathology, Peking University Shenzhen Hospital, Shenzhen, China, [4] School of Traditional Chinese Medicine, Capital Medical University, Beijing, China, [5] Department of Obstetrics and Gynecology, Guangdong Provincial People's Hospital, Guangzhou, China, [6] Peng Cheng Laboratory, Shenzhen, China

OPEN ACCESS

Edited by:
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Renmin Han,
Shandong University, China
Xingyu Liao,
Central South University, China
Xuefeng Cui,
Shandong University, China

***Correspondence:**
Yonghong He
heyh@sz.tsinghua.edu.cn
Yupeng Chen
chenyp01@pcl.ac.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 10 June 2020
Accepted: 13 July 2020
Published: 04 August 2020

Citation:
Liu Y, Li X, Zheng A, Zhu X, Liu S, Hu M, Luo Q, Liao H, Liu M, He Y and Chen Y (2020) Predict Ki-67 Positive Cells in H&E-Stained Images Using Deep Learning Independently From IHC-Stained Images. Front. Mol. Biosci. 7:183. doi: 10.3389/fmolb.2020.00183

Objective: To obtain molecular information in slides directly from H&E staining slides, which apparently display morphological information, to show that some differences in molecular level have already encoded in morphology.

Methods: In this paper, we selected Ki-67-expression as the representative of molecular information. We proposed a method that can predict Ki-67 positive cells directly from H&E stained slides by a deep convolutional network model. To train this model, we constructed a dataset containing Ki-67 negative or positive cell images and background images. These images were all extracted from H&E stained WSIs and the Ki-67 expression was acquired from the corresponding IHC stained WSIs. The trained model was evaluated both on classification performance and the ability to quantify Ki-67 expression in H&E stained images.

Results: The model achieved an average accuracy of 0.9371 in discrimination of Ki-67 negative cell images, positive cell images and background images. As for evaluation of quantification performance, the correlation coefficient between the quantification results of H&E stained images predicted by our model and that of IHC stained images obtained by color channel filtering is 0.80.

Conclusion and Significance: Our study indicates that the deep learning model has a good performance both on prediction of Ki-67 positive cells and quantification of Ki-67 expression in cancer samples stained by H&E. More generally, this study shows that deep learning is a powerful tool in exploring the relationship between morphological information and molecular information.

Availability and Implementation: The main program is available at https://github.com/liuyiqing2018/predict_Ki-67_from_HE

Keywords: digital pathology, immunohistochemistry, Ki-67, deep learning, fully convolutional network, neuroendocrine tumor

INTRODUCTION

In recent years, deep learning has developed rapidly and has outperformed humans in some medical data analysis tasks (Li et al., 2018; Norgeot et al., 2019; von Chamier et al., 2019). Meanwhile, more and more tissue slides are digitalized by a scanner and saved as whole slide images (WSIs). Thus, it is natural to come up with the idea about applying deep learning algorithms to these WSIs. In fact, many researched tasks have explored the potential of deep learning on histopathological image analysis (Komura and Ishikawa, 2018), such as detection or segmentation of Region of Interest (ROI) (Spanhol et al., 2016), scoring of immunostaining (Mungle et al., 2017), mitosis detection (Roux et al., 2013) and so on.

In terms of pathology, hematoxylin and eosin (H&E), as the gold standard stain in evaluations for many cancer types, is routinely employed worldwide (Xu et al., 2019). In most cases, pathologists rely on H&E for their diagnosis and the majority of algorithms for histopathological image analysis, like cell detection, tissue segmentation and cancer grading, are based on H&E imaging (Ghaznavi et al., 2013). It is easy to acquire and cost effective. However, H&E stained slides only contain basic morphological information (Wittekind, 2003), such as the shapes of cells, tissues and tissue blocks. Molecular information like the expression of antigen (protein) in cells, which is more micro, is not reflected in H&E stained slides, which makes it difficult for pathologists and algorithms to analyze and assess.

To obtain molecular information in slides, immunohistochemistry staining (or IHC staining) is often employed in clinical practice. It allows the visualization of specific proteins on the tissue slide by binding targeted antibodies to corresponding proteins and highlighting the protein-binded antibodies by using chromogens of different colors (Ramos-Vara and Miller, 2014; Xu et al., 2019). Hence, this method can distinguish cells that express particular proteins from other components and therefore augment pathologist interpretation and direct therapy.

If a patient needs further diagnosis (such as confirming tumor subtype) or a targeted treatment plan, then an immunohistochemical test is often needed although he has already had H&E stained slides. It is because tumor subtype classification and making the plan of immunotherapy need some molecular information, which is not directly reflected in H&E staining slides. If this information can be inferred from H&E staining slides by some techniques like deep learning, it will greatly improve diagnostic efficiency and save costs.

If the assumption holds that the differences between positive cells (cells that contain a specific protein) and negative cells (cells that do not contain a specific protein) in IHC-stained slides have correlation with H&E-stained slides from the same regions, then there should be a way to model the relationship between the morphological information of cells in H&E images and IHC stained conditions of the cells. It is then possible to predict whether a cell can express specific proteins directly from a H&E-stained slide, without additional IHC staining process. In fact, some related works have been done to predict molecular information from H&E stained images. Coudray et al. (2018) founded six out of ten most commonly mutated genes in LUAD can be predicted from pathology images. Kather et al. (2019) showed that deep residual learning can predict microsatellite instability directly from H&E histology.

Ki-67 is a cancer antigen that is sometimes considered a good marker of proliferation, helping doctors determine patients' cancer prognosis or their chance of recovery (Scholzen and Gerdes, 2000). However, in clinical practice, not every patient is tested for Ki-67 since it is time and money-consuming.

In this paper, we proposed a method that can predict Ki-67 positive cells directly from H&E stained slides by a deep convolutional network model, which realized a cell-level transformation. After the training process, the model was evaluated both on classification and quantification performance. The classification accuracies for our model on training set and validation set are 0.9780 and 0.9371. As for evaluation on quantification performance, the correlation coefficients of D_{pos}, D_{neg} and R_{pos} between these two different types of images are 0.60, 0.73, and 0.80. The results reflect the consistency of Ki-67-expression between real IHC staining images and the output images of our model using H&E staining images as the inputs.

MATERIALS AND METHODS

The overview of our method is displayed on **Figure 1**. first, Consecutive sections of (formalin-fixed paraffin-embedded) samples obtained from the neuroendocrine tumor of twelve patients were cut and stained with H&E and Ki-67 antibody. Then, the slides were digitalized and a set of Ki-67 positive or negative cells in H&E stained images were annotated based on the Ki-67 expression present in the IHC stained images. After that, these cells along with some background patches were extracted for training the model. In order to quantify Ki-67 expressions in a bigger H&E stained image (sized 7,556 × 3,864 for each), a transformation was applied to our trained convolutional network to convert all the fully connected layers into convolutional layers. In this way, the transformed network can take one ROI as the input and output the classification map of the ROI. In order to compare real IHC staining images and images predicted by our model, we use color channel filtering to convert IHC staining images into three-value colormaps.

Data Preparation

Patient Material

Formalin-fixed paraffin-embedded tumor samples of twelve patients operated for neuroendocrine tumor within the Peking university Shenzhen Hospital, China, were used in the study. The samples were stored in archives of Department of pathology in Peking university Shenzhen hospital and the Head of the Department of Pathology approved the use of the samples. The samples were anonymized and all patient-related data and unique identifiers were removed. The procedures were performed under the supervision and approval of the Ethics Committee in Peking university Shenzhen hospital. Samples represented different histological types: five cases with neuroendocrine tumor

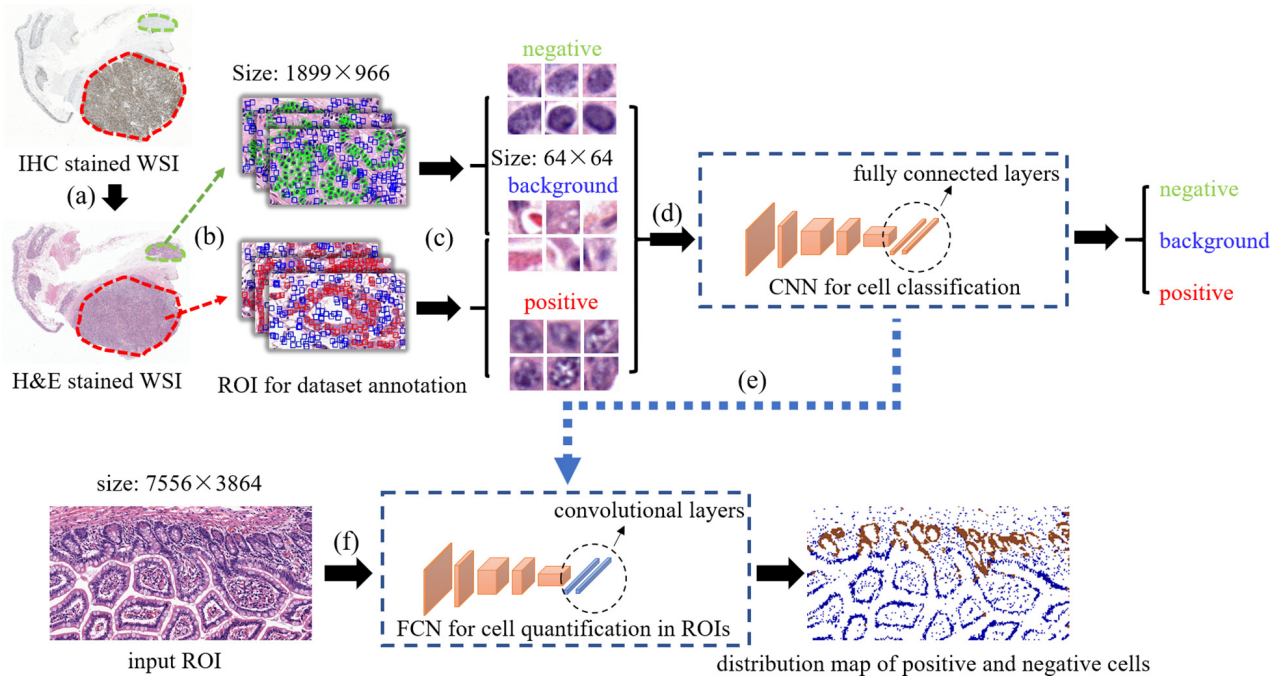


FIGURE 1 | The overview of our method. **(a)** selecting positive/negative regions in H&E stained WSIs with the guidance of IHC stained WSIs, **(b)** ROIs selection for positive/negative/background samples extraction, **(c)** positive/negative/background samples annotation and extraction, **(d)** training CNN for cell classification, **(e)** transforming fully connected layers into convolutional layers, **(f)** taking a test ROI as the input of the transformed model for positive/negative cell distribution prediction.

of rectum(G1), two cases of neuroendocrine tumor of colon(G3), one case of neuroendocrine tumor of small intestine(G1), two cases of neuroendocrine tumor of duodenum(G3), one case of gastric tubular adenocarcinoma with neuroendocrine tumor (G3) and one case of rectal tubular adenocarcinoma with neuroendocrine tumor (G3).

Staining Protocols

From each Formalin-fixed parafinembedded block, we cut two consecutive sections (3.5 μm): One for H&E staining and one for staining with the Anti-Ki67 antibody. For H&E staining, we used undiluted Mayer's hematoxylin and 0.5% eosin. For IHC, we used Anti-Ki67 antibody (Roche, United States), 3,3'-diaminobenzidine as chromogen, and Mayer's hematoxylin as a counterstain with a 1:10 dilution.

Sample Digitization

Matched H&E and IHC stained slides were scanned at 40× with Sqray slide scanner.

Construction of the Dataset

Based on the Ki-67-expression, we selected 300 regions of interest (ROIs) sized 1,889 × 966 from 5 out of 12 H&E stained slides. Then we extracted 5,900 images of positive cells, 6,086 images of negative cells and 6,776 images of background from these ROIs.

The way of selecting positive and negative samples can be described with **Figure 2**. As is shown in **Figure 2**, there are Ki-67 positive regions in Ki-67 stained slides where all cells are Ki-67 positive. We can infer that the corresponding regions in H&E stained slides are also positive. Therefore, positive samples can be obtained by the following steps: First, extract images from the positive regions in H&E stained slides; Then, annotate each cell in these extracted images with a point label by using a open source annotation software Labelme. Finally, extract patches with these annotated points as centers and these patches are what we need. The way of obtaining negative samples is similar with that of obtaining positive samples.

The method of extracting background samples is shown in **Figure 3**. After the samples of positive cells and negative cells are selected, the background samples are selected by random sampling: a series of candidate boxes (shown in blue in **Figure 3**) are randomly generated. If the candidate boxes do not overlap with the boxes of negative cells (shown in green in **Figure 3**) or positive cells (shown in red in **Figure 3**), they will be retained and selected.

The size of these images was all 64 × 64 without any resize operation. The reason why we use 64 × 64 as the patch size is that the distribution of cell size is 40 pixel × 40 pixel ∼70 pixel × 70 pixel so the size of 64 × 64 can cover most situations. In addition, 64 is integer power of 2 which is convenient for computing. After that, the dataset consisting of all the images were split randomly into training set and validation set with the ratio of 8:2. The procedure of constructing the dataset were illustrated in **Figures 1a–c**. **Table 1** summarizes information about the dataset.

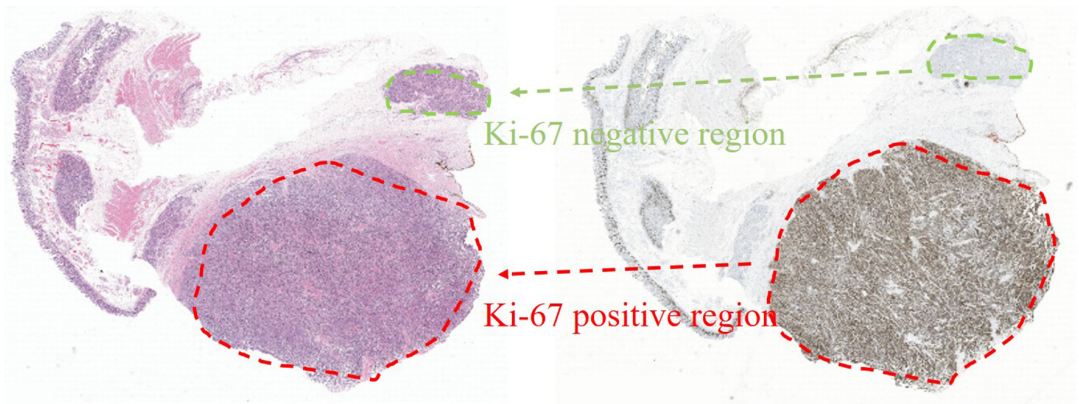


FIGURE 2 | The way of selecting positive and negative samples.

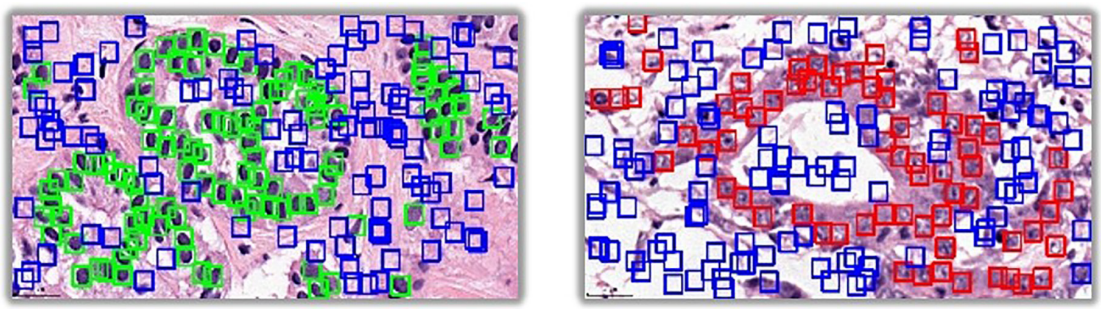


FIGURE 3 | The method of extracting background samples.

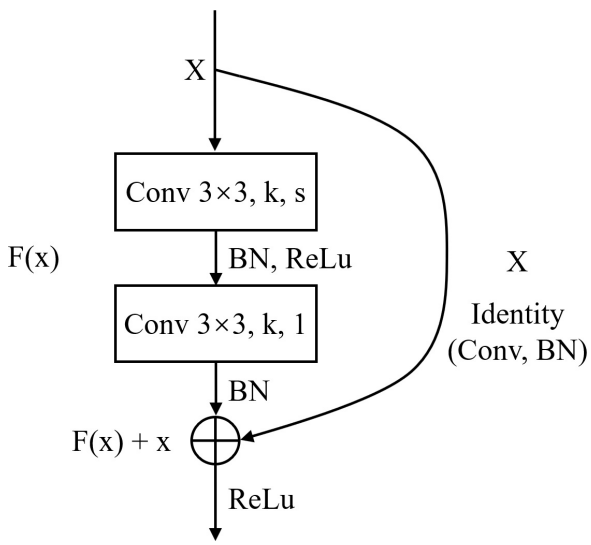


FIGURE 4 | The structure of Block.

TABLE 1 | Information about the dataset.

Parameters	Values
Magnification/number of the H&E-IHC pairs	40×/12
Size/number of the ROIs selected for extracting training and validation set	1,889 × 966/300
Numbers of positive cell/negative cell background images	5,900/6,086/6,776
Ratio of training set to validation set	8:2
Size/number of the ROIs selected for evaluation of cell quantification	7,556 × 3,864/32

Classification Using CNN

Deep Learning is a significant area of Machine Learning research. It uses very deep (in terms of number of layers) neural network to solve problems, especially problems which are related to visual recognition. The key aspect of all the deep learning architectures is the use of Convolutional Neural Network (CNN) (Krizhevsky et al., 2012). CNN is a biologically inspired form of the artificial neural network, that has local connections and shared weights. It is one of the most important tools of machine learning when it comes to the current generation, and it has been very popularly used to solve image recognition tasks, in the field of Computer Vision. The CNN architecture can be obtained by exploiting existing famous networks such as VGG (Simonyan and Zisserman, 2014), Inception (Szegedy et al., 2015) or ResNet (He et al., 2016), or by designing a new network. Both of them have their own strengths and weaknesses. Using existing networks can take advantages of pre-trained weights acquired from training

TABLE 2 | Detailed information about the modified ResNet18 network.

Layer name	Input size/output size	Elements/params
conv1	3 × 64 × 64/64 × 32 × 32	Conv/7 × 7, k = 64, s = 2
		BN/-
		ReLu/-
pool1	64 × 32 × 32/64 × 16 × 16	MaxPool/3 × 3, s = 2
layer1	64 × 16 × 16/64 × 16 × 16	Block/k = 64, s = 1
		Block/k = 64, s = 1
layer2	64 × 16 × 16/128 × 8 × 8	Block/k = 128, s = 2
		Block/k = 128, s = 1
layer3	128 × 8 × 8/256 × 4 × 4	Block/k = 256, s = 2
		Block/k = 256, s = 1
layer4	256 × 4 × 4/512 × 2 × 2	Block/k = 512, s = 2
		Block/k = 512, s = 1
FC	512 × 2 × 2 (flattened)/3	fc/out = 3

*The structure of Block is shown in **Figure 4**.*

the networks on large scale public datasets such as ImageNet (Deng et al., 2009) for transfer learning. This can accelerate the process of training significantly and guarantee the classification accuracy in the stage of inference, with only a small amount of training set. However, these pre-trained weights are usually generated by training the networks on natural images, which have considerable difference from histopathological images. So, the pre-trained models for natural image classification may not entirely appropriate for recognition tasks on histopathological images. Using self-designed network can be more flexible as we can devise a more targeted model according to the characteristics of the dataset. While this approach may not achieve a satisfied result if the training set is limited. ResNet is a well-known deep learning network architecture proposed by He et al. (2016). By using "shortcut connections," this network are easier to optimize, and can gain accuracy from considerably increased depth. In this paper, we adopted a modified ResNet18 as our CNN classifier by removing the last average pooling layer to make the network adaptable for locating positive or negative cells in large-scale H&E stained images (which will be elaborated in the next subsection) and changing the number of the output nodes in the last fully connected layer into 3 since it was a 3-value classification problem. **Table 2** lists detailed information about the modified ResNet18 network including layer name, input and output size, types of elements in each layer and their parameters. k represents the number of the kernels and s represents stride.

Figure 1d shows the stage of training CNN. The input to the first layer is an RGB image containing one positive or negative cell only or not containing any cell. The last layer generated labels, showing the probability of the image whether it represents a positive cell, a negative cell or background. Then a loss function was calculated and back propagation will be conducted to adjust the weighting parameters of the network so as to minimize the loss.

Cell Quantification in ROIs Using Fully Convolutional Network

We had trained a CNN classifier using the samples of positive cell, negative cell and background. However, this classifier had a

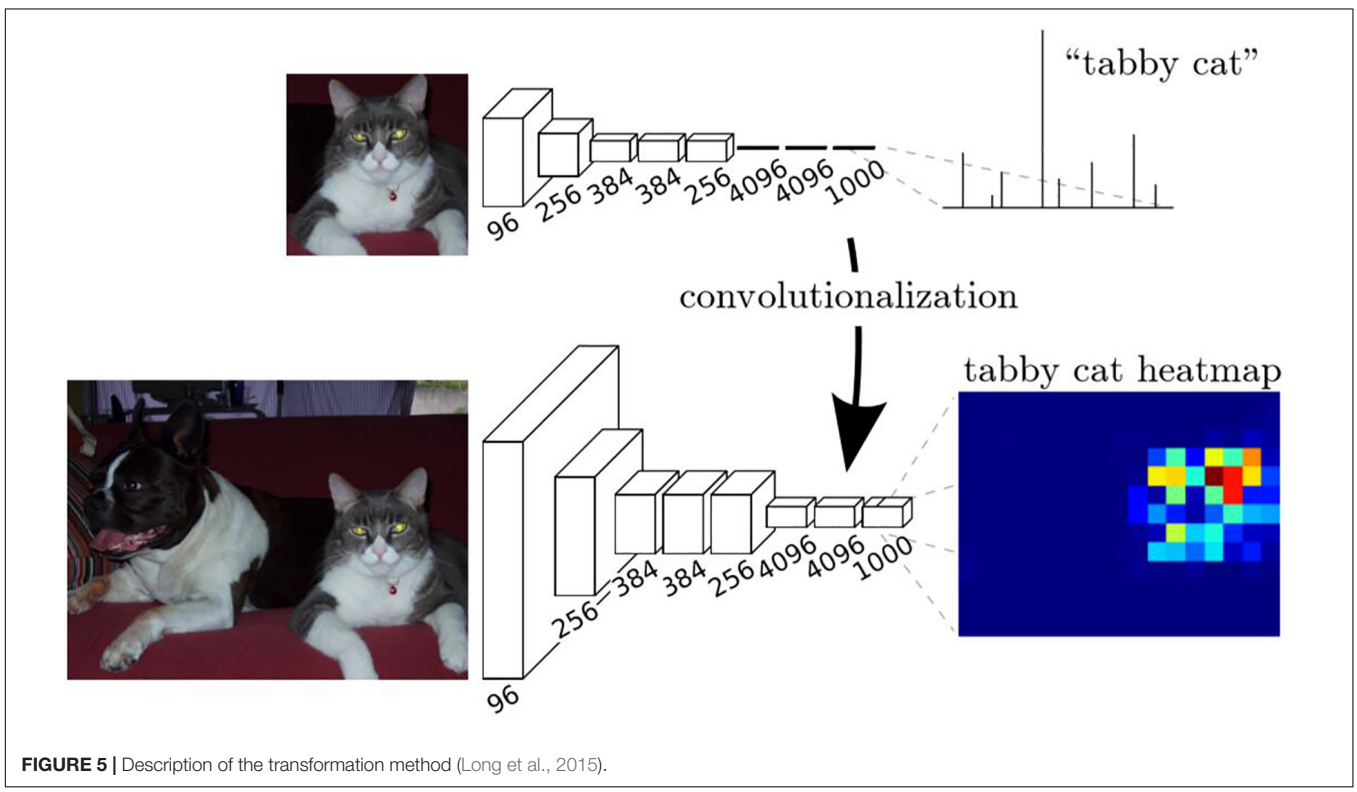


FIGURE 5 | Description of the transformation method (Long et al., 2015).

TABLE 3 | Training details of our modified ResNet18 model.

GPU	TITAN Xp (12GB) × 1
Framework	Pytorch 1.1.0
Data preprocessing	Random Horizontal Flip Random Vertical Flip Normalize (mean vector = [0.485, 0.456, 0.406], standard deviation vector = [0.229, 0.224, 0.225])
Batch size	64
Loss function	Cross Entropy
Learning rate	1e-3 (epoch 1–10), 1e-4 (epoch 11–20)
Optimizer	Adam with weight decay = 1e-5
Training on the pretrained model?	Yes (pretrained model from ImageNet.)
Training epochs	20

TABLE 4 | Classification report on training set.

	Precision	Recall	f1-score	Support
Negative	0.962	0.976	0.969	4834
Positive	0.973	0.963	0.968	4760
Background	0.996	0.993	0.995	5415
Avg/total	0.978	0.978	0.978	15009

fixed size (64 × 64) of input and can only classify images with that size. In order to obtain the classification maps of ROIs (sized 7,556 × 3,864 for each), a transformation method proposed in Long et al. (2015) was applied to our trained CNN to convert all the fully connected layers into convolutional layers, as is shown in **Figure 1e**.

The transformation method can be described in **Figure 5**. In training stage, the network learns a classification task. The input is a fixed size image, and the output is the corresponding category (cat for example) of the image. In the inference stage, the fully connected layer of the trained network is rearranged into a convolutional layer. In this way, the network can take any size of the image as input and output a probability map, representing the predicted probability of the target at each pixel in the input image.

Thus, the transformed CNN can take one ROI as the input and output the classification map of the ROI. The procedure was displayed in **Figure 1f**.

Statistical Methods

To evaluate the classification result, we adopted accuracy, precision, recall, F1-score and confusion matrix. For illustration purposes, we will use T and F to indicate whether the network prediction is correct or not. P and N are used to indicate whether the sample is negative or positive. Therefore, TP (True Positive) means positive and correctly predicted, while FN (False Negative) means negative and wrongly predicted. The same is true for TN and FP. In this way, accuracy, precision and recall can be expressed as following:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{Precesion} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score is defined as a harmonic mean of precision and recall:

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{preicision} + \text{recall}}$$

Confusion matrix, which is represented by an n × n matrix, is a specific table layout that allows visualization of the performance of an algorithm. Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class. A value in i column j row represent how many samples in class j is predicted to be class i.

To evaluate the quantification result, first we calculated dense of positive cells (D_{pos}), proportional area of negative cells (D_{neg})

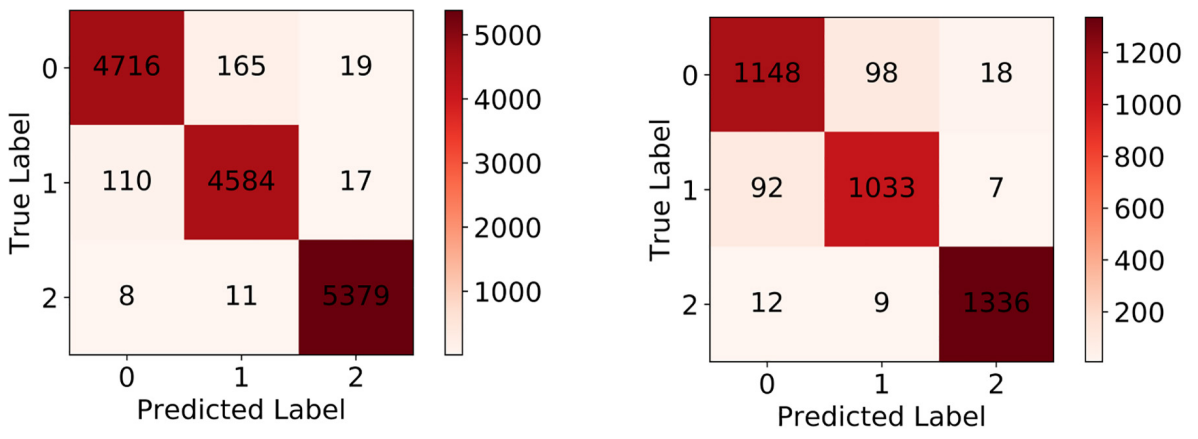


FIGURE 6 | Left: Confusion matrix of training set. **Right:** Confusion matrix of validation set.

TABLE 5 | Classification report on validation set.

	Precision	Recall	f1-score	Support
Negative	0.908	0.917	0.913	1252
Positive	0.913	0.906	0.909	1140
Background	0.985	0.982	0.983	1361
Avg/total	0.937	0.937	0.937	3753

and positive rate (R_{pos}) in H&E ROIs and IHC ROIs respectively, which are defined as:

$$\begin{cases} D_{\text{pos}} = \frac{S_{\text{pos}}}{S_{\text{ROI}}} \\ D_{\text{neg}} = \frac{S_{\text{neg}}}{S_{\text{ROI}}} \\ R_{\text{pos}} = \frac{S_{\text{pos}}}{S_{\text{pos}}+S_{\text{neg}}} \end{cases}$$

where S_{pos} is the area covered by positive cells, S_{neg} is the area covered by negative cells, S_{ROI} is the area of a whole ROI. The areas covered by positive cells or negative cells were obtained simply by color channel filtering. Then we computed pairwise correlation coefficient (r), as is defined in the equation below, for measuring correlation between the quantification results of H&E ROIs and that of IHC ROIs.

$$r(X,\ Y) = \frac{Cov(X,\ Y)}{\sqrt{Var\,[X]\ Var[Y]}}$$

where $cov(X, Y)$ is defined as

$$Cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

EXPERIMENTS AND RESULTS

Experimental Setup

In this section, we tested the performance of cell classification and cell quantification of our model. **Table 3** shows the training details of our modified ResNet18 model.

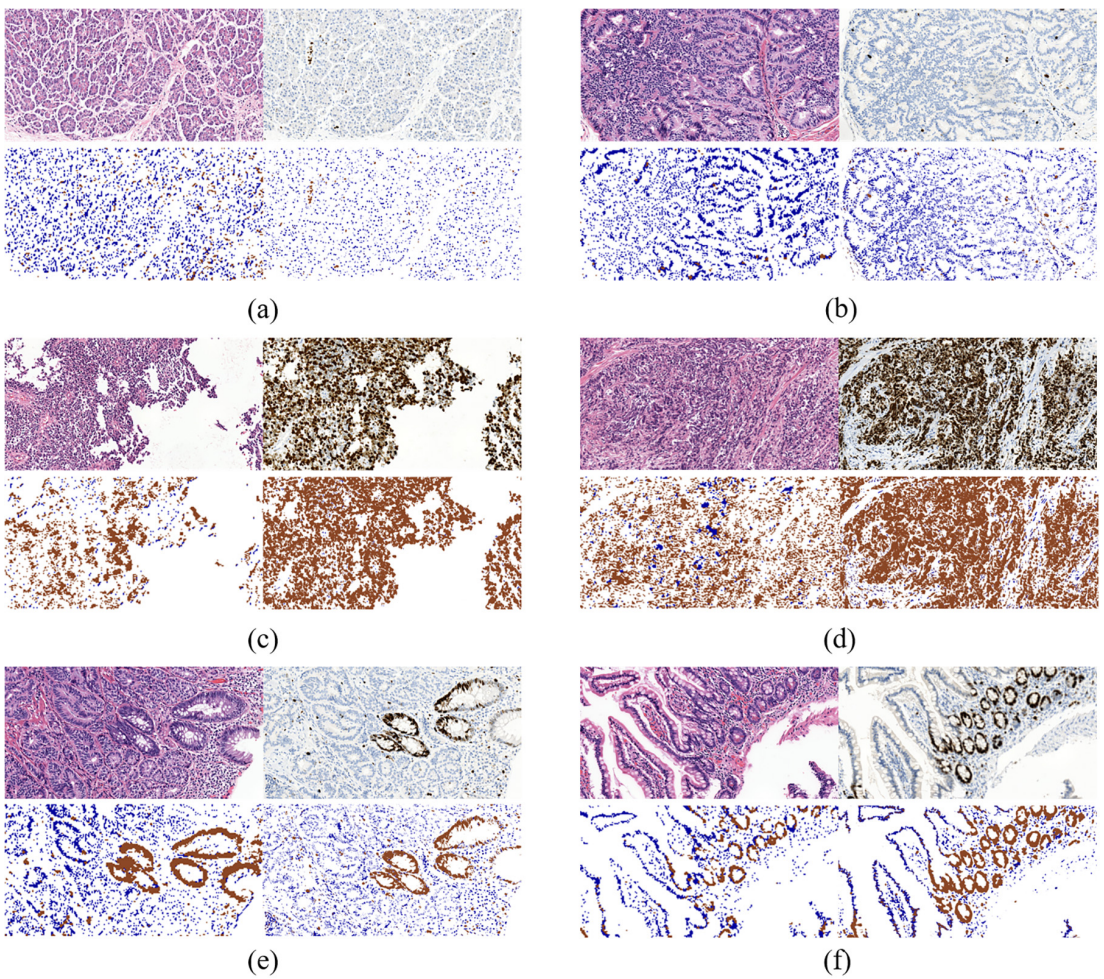


FIGURE 7 | Three typical cases of Ki-67-expression: middle rate of Ki-67 positive cells **(a,b)**, low rate of Ki-67 positive cells **(c,d)** and high rate of Ki-67 positive cells **(e,f)**. Each case contains four sub-figures, representing H&E stained ROIs (top left), IHC stained ROIs (top right), quantification results in H&E stained ROIs (bottom left) and quantification results in IHC stained ROIs (bottom right).

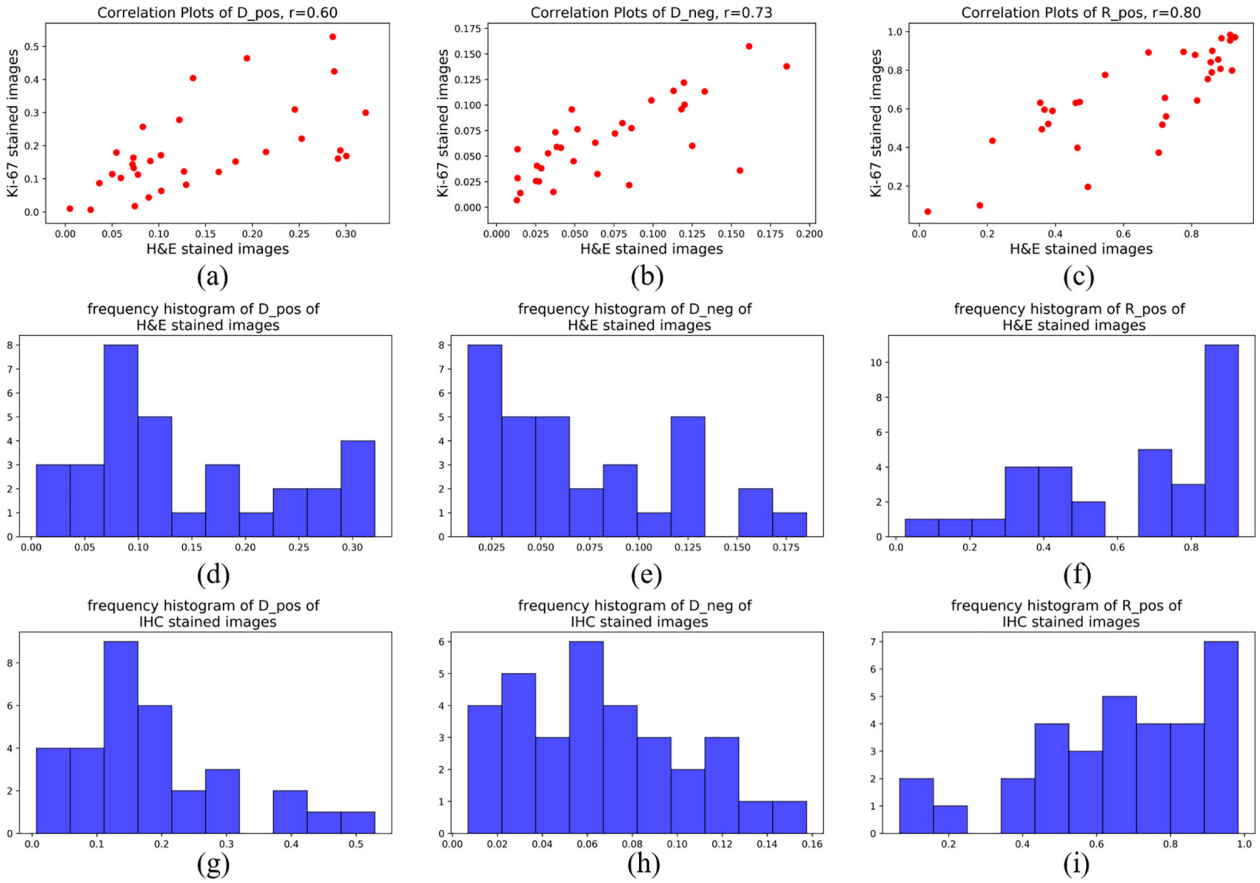


FIGURE 8 | Frequency histogram of **(a)** Dpos, **(b)** Dneg and **(c)** Rpos of H&E stained images; Frequency histogram of **(d)** Dpos, **(e)** Dneg and **(f)** Rpos of IHC stained images; Correlation-plots of H&E stained images and IHC stained images of **(g)** Dpos, **(h)** Dneg and **(i)** Rpos.

Evaluation on Single Cell Classification

After the training process, we fed the validation set into our trained model to evaluate its classification performance. The classification accuracies for our model on training set and validation set are 0.9780 and 0.9371. **Table 4** and **Figure 6** left show the classification report and confusion matrix of the results of the train set respectively. **Table 5** and **Figure 6** right show those of the validation set. We also performed a 10-fold cross-validation analysis. We randomly split the training set (15,009 images) in to 10 subsets (nine sets of 1,501 images and one set of 1,500 images). In 10 training rounds, the average accuracy was 0.9310 (range: 0.9167–0.9427, std = 0.0085). These results are consistent with the previous results obtained from the validation set of 3,753 images, which suggest that the performance of our model is robust to how we split our dataset for training and test.

Evaluation on Cell Quantification in ROIs

To further evaluate our model, we compared the quantification results in H&E stained ROIs performed by the model and the results in IHC stained ROIs performed by color channel filtering. **Figure 7** displays three typical cases of Ki-67 expression. It's obvious that the quantification results in H&E stained ROIs are basically consistent with the quantification results in IHC stained ROIs.

Specifically, the results in middle rate cases and low rate cases are better than those in high rate cases. It is because in ROIs with high rate of Ki-67 positive cells, the distribution of positive and negative cells is more irregular than the other two types of cases, which makes it difficult to distinguish positive cells from negative cells in H&E stained ROIs in the process of annotation. It is because when we label the cells, due to the lack of correspondence between scattered cells in the H&E stained regions and Ki-67 stained regions, it was not completely determined whether a certain cell was a positive cell or a negative cell in H&E stained ROIs unless all the cells in these regions are all positive or negative. Though there are a mass of positive cells in Ki-67 positive regions, a small number of negative cells are inevitably mixed in with positive cells in this type of ROIs, which makes labeling more difficult. While the similar situations appear less in Ki-67 negative regions. In other words, negative cells in Ki-67 negative regions can be extracted with more confidence than positive cells in Ki-67 positive regions.

In addition, there are many glandular-like structures in ROIs with low or medium density of negative cells. If all the cells on a gland in KI-67 stained ROIs are negative or positive, then all the

cells in the corresponding gland in H&E stained ROIs are also marked as negative or positive, according to the correspondence between glands in H&E stained ROIs and Ki-67 stained ROIs.

For statistical evaluation, we calculated D_{pos}, D_{neg} and R_{pos} in 32 pairs of H&E-staining ROIs and IHC-staining ROIs. **Figure 8** shows the frequency histograms and correlation plots of these three indexes in H&E stained images and IHC stained images. The correlation coefficients of D_{pos}, D_{neg} and R_{pos} between these two different types of ROIs are 0.60, 0.73 and 0.80. The results reflect the quantitative consistency of Ki-67 expression between the two types of staining images. Moreover, The correlation coefficients of R_{pos} has the highest value indicates that the evaluation indexes considering both positive and negative cells can more stablely reflect the relationship between H&E stained ROIs and Ki-67 stained ROIs.

DISCUSSION

In this paper, we made an attempt to build a relationship between H&E stained slides and Ki-67 antibody stained slides. We introduced a modified ResNet18 model to predict Ki-67 expression directly from H&E stained images without any IHC staining process. Our results show that morphological information has close relation with molecular information, which are consistent with the opinion proposed in Fuchs and Buhmann (2011) that tissue and cell morphologies displayed in histopathological images are a function of underlying molecular drivers. Once their relationship is established, it should be possible to faithfully predict the distribution of specific protein abundance directly in samples only using a basic morphology staining. However, it's just the beginning of our research on this topic. Challenges still exist, including:

1. Performance of the model is highly dependent on the quality of input images. Low quality images may result in less accurate results. The quality of images is influenced by many factors, such as standardization of making slides, quality of stains and accuracy of scanners.
2. The relationship between morphological information and molecular information may be very complex, considering the diversity of different lesions, tissues, cells and antibodies. For the moment, our research has only focused on one specific relationship so much work should be done if we want our model to be more generalized.
3. At present we can only distinguish between positive cells and negative cells in some certain regions of a H&E stained images guided by the corresponding IHC stained image. It's hard to verify the positive degree of a cell in a H&E stained image even with the help of IHC staining, which hampers a more precise inference of the model.

Our future work will mainly focus on the following aspects. first, Enlarge our dataset to contain more samples. So, the model trained on the new dataset will have stronger ability of robust and generalization; Second, Conduct more experiments on samples with different tissues and stains to promote our conclusion to a more general situation; Last but not least, Optimize our model. For example, semi-supervised learning can be adopted to alleviate the workload of annotation.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YH presented the initial idea for the article and provided financial support. YL wrote all the code, conducted experiments and wrote most of the article. XL and AZ communicated with the hospital and obtained access to the data. They also provided guidance on pathological diagnosis, verified the medical significance of our work, and wrote part of the article. YC provided support on computing resources and guidance on deep learning techniques. XZ and ML reviewed the data and the results of the experiment and gave valuable suggestions for the revision of the article. QL and HL made and collected all the slides. SL scanned all the slides and proposed the method of data labeling. MH finished the work of data labeling. All authors contributed to the article and approved the submitted version.

FUNDING

This research was made possible with the financial support from National Science Foundation of China (NSFC) (61875102, 81871395, and 61675113), Science and Technology Research Program of Shenzhen City (JCYJ20170816161836562, JCYJ20170817111912585, JCYJ20160427183803458, JCYJ20170412171856582, and JCYJ20180508152528735), Oversea cooperation foundation, Graduate School at Shenzhen, Tsinghua University (HW2018007).

REFERENCES

Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., et al. (2018). Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat. Med.* 24, 1559–1567. doi: 10.1038/s41591-018-0177-5

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., et al. (2009). "Imagenet: a large-scale hierarchical image database," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, (Miami, FL: *IEEE*), 248–255.

Fuchs, T. J., and Buhmann, J. M. (2011). Computational pathology: challenges and promises for tissue analysis. *Comp. Med. Imag. Graph.* 35, 515–530. doi: 10.1016/j.compmedimag.2011.02.006

Ghaznavi, F., Evans, A., Madabhushi, A., and Feldman, M. (2013). Digital imaging in pathology: whole-slide imaging and beyond. *Ann. Rev. Pathol. Mechan. Dis.* 8, 331–359. doi: 10.1146/annurev-pathol-011811-120902

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Piscataway, NJ: IEEE), 770–778.

Kather, J. N., Pearson, A. T., Halama, N., Jäger, D., Krause, J., Loosen, S. H., et al. (2019). Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* 25, 1054–1056. doi: 10.1038/s41591-019-0462-y

Komura, D., and Ishikawa, S. (2018). Machine learning methods for histopathological image analysis. *Comput. Struct. Biotechnol. J.* 16, 34–42. doi: 10.1016/j.csbj.2018.01.001

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Proceedinds of the Advances in Neural Information Processing Systems*, (Tahoe, NV: NIPS), 1097–1105.

Li, Y., Xu, F., Zhang, F., Xu, P., Zhang, M., Fan, M., et al. (2018). DLBI: deep learning guided Bayesian inference for structure reconstruction of super-resolution fluorescence microscopy. *Bioinformatics* 34, i284–i294. doi: 10.1093/bioinformatics/bty241

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Boston, MA: IEEE), 3431–3440.

Mungle, T., Tewary, S., Das, D. K., Arun, I., Basak, B., Agarwal, S., et al. (2017). MRF-ANN: a machine learning approach for automated ER scoring of breast cancer immunohistochemical images. *J. Microsc.* 267, 117–129. doi: 10.1111/jmi.12552

Norgeot, B., Glicksberg, B. S., and Butte, A. J. (2019). A call for deep-learning healthcare. *Nature medicine* 25, 14–15. doi: 10.1038/s41591-018-0320-3

Ramos-Vara, J. A., and Miller, M. A. (2014). When tissue antigens and antibodies get along: revisiting the technical aspects of immunohistochemistry—the red, brown, and blue technique. *Vet. Pathol.* 51, 42–87. doi: 10.1177/0300985813505879

Roux, L., Racoceanu, D., Loménie, N., Kulikova, M., Irshad, H., Klossa, J., et al. (2013). Mitosis detection in breast cancer histological images An ICPR 2012 contest. *J. Pathol. Inform.* 4:8. doi: 10.4103/2153-3539.112693

Scholzen, T., and Gerdes, J. (2000). The Ki-67 protein: from the known and the unknown. *J. Cell. Physiol.* 182, 311–322. doi: 10.1002/(sici)1097-4652(200003)182:3<311::aid-jcp1>3.0.co;2-9

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv* [Preprint]. arxiv:1409.1556 (accessed June 10, 2020).

Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. (2016). "Breast cancer histopathological image classification using convolutional neural networks," in *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, (Vancouver: IEEE), 2560–2567.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition*, (Boston, MA: IEEE), 1–9.

von Chamier, L., Laine, R. F., and Henriques, R. (2019). Artificial intelligence for microscopy: what you should know. *Biochem. Soc. Trans.* 47, 1029–1040. doi: 10.1042/bst20180391

Wittekind, D. (2003). Traditional staining for routine diagnostic pathology including the role of tannic acid. 1. value and limitations of the hematoxylin-eosin stain. *Biotechn. Histochem.* 78, 261–270. doi: 10.1080/10520290310001633725

Xu, Z., Moro, C. F., Bozóky, B., and Zhang, Q. (2019). GAN-based virtual re-staining: a promising solution for whole slide image analysis. *arXiv* [Preprint]. arxiv:1901.04059 (accessed June 10, 2020).

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 22 October 2020
doi: 10.3389/fmolb.2020.571180

Effective Immunohistochemistry Pathology Microscopy Image Generation Using CycleGAN

Zidui Xu[1], Xi Li[2], Xihan Zhu[3], Luyang Chen[4], Yonghong He[1] and Yupeng Chen[5]**

[1] Department of Life and Health, Tsinghua Shenzhen International School, Shenzhen, China, [2] Department of Gastroenterology, Peking University Shenzhen Hospital, Shenzhen, China, [3] School of Traditional Chinese Medicine, Capital Medical University, Beijing, China, [4] Department of Computer Science, Pennsylvania State University, University Park, PA, United States, [5] Peng Cheng Laboratory, Shenzhen, China

Immunohistochemistry detection technology is able to detect more difficult tumors than regular pathology detection technology only with hematoxylin-eosin stained pathology microscopy images, – for example, neuroendocrine tumor detection. However, making immunohistochemistry pathology microscopy images costs much time and money. In this paper, we propose an effective immunohistochemistry pathology microscopic image-generation method that can generate synthetic immunohistochemistry pathology microscopic images from hematoxylin-eosin stained pathology microscopy images without any annotation. CycleGAN is adopted as the basic architecture for the unpaired and unannotated dataset. Moreover, multiple instances learning algorithms and the idea behind conditional GAN are considered to improve performance. To our knowledge, this is the first attempt to generate immunohistochemistry pathology microscopic images, and our method can achieve good performance, which will be very useful for pathologists and patients when applied in clinical practice.

Keywords: immunohistochemistry pathology microscopy image, medical image generation, CycleGAN, conditional GAN, multiple instances learning

OPEN ACCESS

Edited by:
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Peng Yang,
Baidu, United States
Juntao Liu,
Shandong University, China
Fa Zhang,
Institute of Computing Technology, China

****Correspondence:***
Yonghong He
heyh@sz.tsinghua.edu.cn
Yupeng Chen
chenyp01@pcl.ac.cn

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *10 June 2020*
Accepted: *19 August 2020*
Published: *22 October 2020*

Citation:
Xu Z, Li X, Zhu X, Chen L, He Y and Chen Y (2020) Effective Immunohistochemistry Pathology Microscopy Image Generation Using CycleGAN. Front. Mol. Biosci. 7:571180. doi: 10.3389/fmolb.2020.571180

INTRODUCTION

Immunohistochemistry (IHC) detection technology, such as staining with Ki-67 reagent, plays an important role in tumor detection. About 5–10% of patients with tumors cannot be detected accurately only with hematoxylin-eosin (HE) stained pathology microscopic images. Luckily, with the rapid development of IHC detection technology, many difficult tumors can be detected, especially undifferentiated or poorly differentiated tumors. Although IHC detection technology is a more accurate method, making Ki-67 pathology microscopic images costs a large amount of money and time. Considering the surprising performance of deep learning technology in medical image analysis region (Wang et al., 2016; Liu et al., 2017; Li and Ping, 2018), especially generating synthetic medical images using generative adversarial networks (GAN) and its variants (Goodfellow et al., 2014; Hoffman et al., 2017; Nie et al., 2017; Frid-Adar et al., 2018a,b; Han et al., 2018; Kazeminia et al., 2018; Mahmood et al., 2018; Shin et al., 2018), generating synthetic Ki-67 pathology microscopy images from HE pathology microscopy images using GAN would be a good choice.

However, generating synthetic Ki-67 microscopy images from HE pathology microscopy images is a challenging task. The first reason is the unpaired dataset. Because a pathology slice is not usually allowed to be stained twice, in this paper, the HE pathology microscopy image and its corresponding Ki-67 pathology microscopy image are not pixel-aligned. The second reason is the

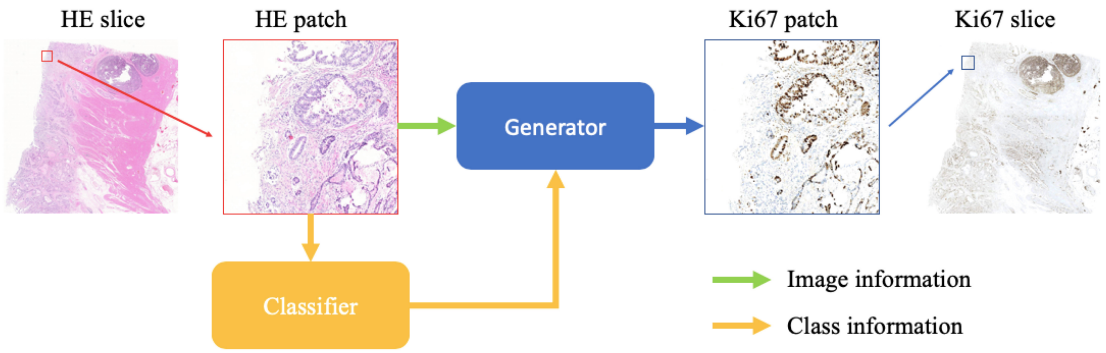


FIGURE 1 | The evaluation process of our method. After patch extraction, the HE pathology patches will be fed into a trained classifier at first. Then the class-related information will be combined with the HE pathology patches, and they are the input of the trained generator, which will generate synthetic Ki-67 pathology patches.

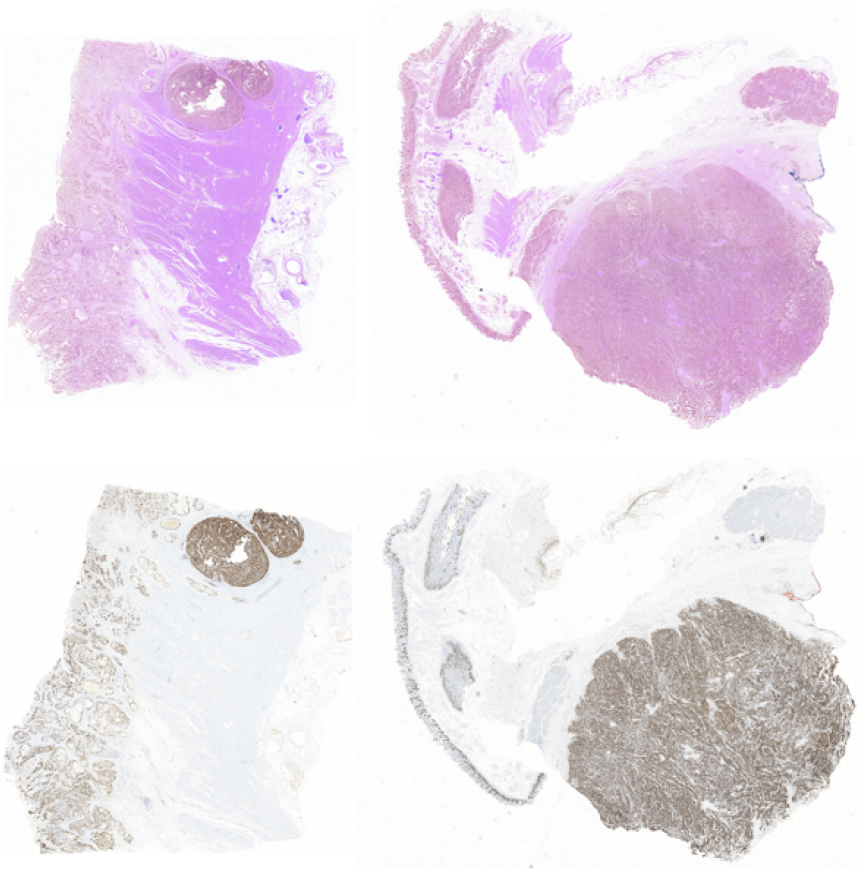


FIGURE 2 | A representation of two HE pathology microscopy images (top) and their corresponding Ki-67 pathology microscopy images (bottom) in our dataset.

difficulty of achieving professional pathological annotations. Unlike any other common computer vision tasks, the annotations of pathology microscopy images can only be completed and checked by professional pathologists, so it is extremely hard to get a large fully and accurately annotated microscopy image dataset. The third reason is the difficulty of considering class-related information from HE pathology microscopy images to Ki-67 pathology microscopy images. The adversarial training process will align some feature vectors between different domains (Song et al., 2019), but the class-related feature vectors are what we need, and we need to handle them.

In recent years, with the development of deep learning technology, many researchers have tried their best to address these drawbacks. CycleGAN was proposed for unpaired image datasets when generating synthetic images (Zhu et al., 2017). By introducing cycle loss functions during the adversarial training

process, the generator finds an accurate mapping between two different domains with unpaired datasets. In this view, CycleGAN is the proper way for unpaired pathology microscopy image datasets. With incomplete or lacking annotations of pathology microscopy images, semi-supervised learning-based methods, unsupervised learning-based methods and self-supervised learning-based methods have been introduced to work on datasets with partial annotations or without any annotation and these methods have proven to be useful (Campanella et al., 2019; Xu G. et al., 2019). Among these methods, multiple instances learning (MIL) algorithms have been applied successfully with unannotated pathology microscopy images, so they have been adopted in this paper (Xu G. et al., 2019). Actually, the adversarial training process aims to align some feature vectors extracting from real images or fake images (Nguyen et al., 2017). When the feature vector that makes the largest contribution to the discriminator are aligned well, then the discriminator would neglect other feature vectors. In our work, we found that when training with small pathology microscopy patches, morphological feature vectors are selected for alignment during adversarial training process. However, what we need is to align class-related feature vectors. The idea behind conditional GAN is extended to address this problem by treating the class label as the additional channel of the input patch, and the class label can be obtained from MIL algorithms (Mirza and Osindero, 2014). By teaching the generator to focus on class-related information, the model will achieve better performance.

We show the schematic representation of our method during the evaluation step in this paper (**Figure 1**). When we generate a synthetic Ki-67 pathology microscopy image from a specific HE pathology microscopy image, the HE pathology microscopy patch will be fed into a classifier to get its class label. Then this class-related information will be combined with the HE pathology microscopy patch, and they are the input of the generator which will produce a synthetic Ki-67 pathology microscopy patch.

In this paper, the major contributions are in three phases: first, we apply CycleGAN to generate synthetic Ki-67 pathology microscopy images from unpaired HE pathology microscopy images. By introducing the cycle loss function and adjusting the architecture of the networks, CycleGAN is able to find an accurate mapping from the HE domain to the Ki-67 domain. Second, we apply a MIL algorithm to train two classifiers from unannotated HE and Ki-67 pathology microscopy images separately. Both classifiers are used to distinguish tumor patches from normal patches. Last, we apply the idea behind conditional GAN. By

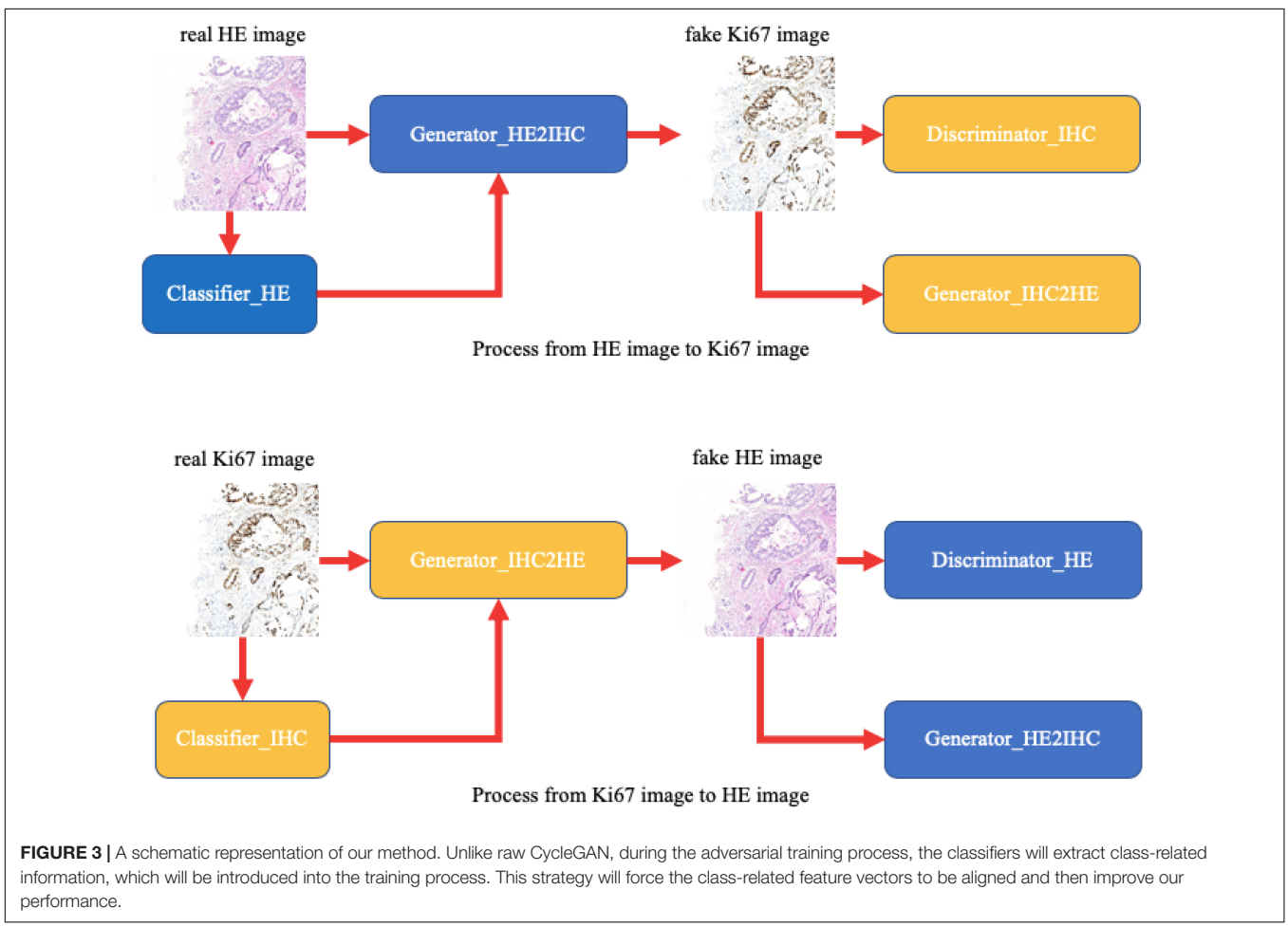


FIGURE 3 | A schematic representation of our method. Unlike raw CycleGAN, during the adversarial training process, the classifiers will extract class-related information, which will be introduced into the training process. This strategy will force the class-related feature vectors to be aligned and then improve our performance.

TABLE 1 | MIL and CycleGAN dataset.

	Training dataset		Evaluation dataset	
	HE	**Ki-67**	**HE**	**Ki-67**
MIL	307/87	359/78	35/35	46/40
CycleGAN	34121	33293	4396	4031

The table above shows the number of samples used during the MIL training process. During the MIL training process, they are the number of positive samples/the number of negative samples. The training data are at the large patch level (2,240 × 2,240 pixels) during the MIL training process and at small patch level (2,24 × 2,24 pixels) during the CycleGAN training process.

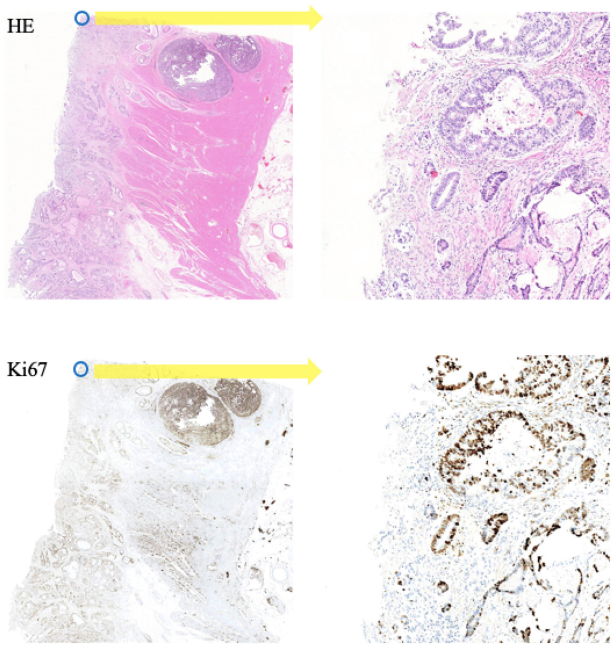


FIGURE 4 | A representation of our dataset. Because the Ki-67 pathology slice is the consecutive slice near HE pathology slice, these two pathology microscopy images are almost same at image level, but they are different at pixel level or patch level.

TABLE 2 | CycleGAN configuration.

Generator	**Discriminator**
Conv (in_c = 3, out_c = 64, k = 7, s = 1, *p* = 3)	**Conv(in_c = 3, out_c = 64, k = 7, s = 1, *p* = 3)**
InstanceNorm(64)	LeakyReLU(0.2)
ReLU()	Conv(in_c = 64, out_c = 128, *k* = 3, *s* = 2, *p* = 1)
Conv(in_c = 64, out_c = 128, *k* = 3, *s* = 2, *p* = 1)	InstanceNorm(128)
InstanceNorm(128)	LeakyReLU(0.2)
ReLU()	Conv(in_c = 128, out_c = 256, k = 3, s = 2, p = 1)
Conv(in_c = 128,out_c = 256, *k* = 3, *s* = 2, *p* = 1)	InstanceNorm(256)
InstanceNorm(256)	LeakyReLU(0.2)
ReLU()	Conv(in_c = 256,out_c = 512, *k* = 3, *s* = 2, *p* = 1)
Conv(in_c = 256, out_c = 256, *k* = 3, s = 1, *p* = 1)	InstanceNorm(512)
InstanceNorm(256)	LeakyReLU(0.2)
ReLU()	Conv(in_c = 512,out_c = 1, *k* = 4, *s* = 1, *p* = 1)
Conv(in_c = 256,out_c = 256, *k* = 3, *s* = 1, *p* = 1)	AveragePool()
InstanceNorm(256)	
Conv_transpose(in_c = 256, out_c = 128, *k* = 3, *s* = 2, p = 1, p_o = 1)	
InstanceNorm(128)	
ReLU()	
Conv_transpose(in_c = 128, out_c = 64, *k* = 3, *s* = 2, p = 1, p_o = 1)	
InstanceNorm(64)	
ReLU()	
Conv(in_c = 64, out_c = 3, *k* = 7, *s* = 1, *p* = 3)	
Tanh()	

The table above shows the detailed configuration of generator and discriminator in CycleGAN framework during CycleGAN training process. Specifically, Conv is convolution, InstanceNorm is instance normalization, ReLU is ReLU activation function, Conv_transpose is transposed convolution, Tanh is Tanh activation function, AveragePool is the average pooling function, in_c is input channel, out_c is output channel, k is kernel size, s is stride size, p is padding size, and p_o is output padding size.

treating class-related information as the additional channel of input patches, class-related feature vectors are forced to be aligned well and this strategy is essential to our performance.

The remaining sections are organized as following: section "Materials and Methods" introduces the dataset and our method in detail. In section "Results," we conduct experiments and show our experimental results. Finally, we conclude and discuss our work in section "Discussion."

MATERIALS AND METHODS

Dataset

In this paper, we conduct experiments on a neuroendocrine tumor dataset. Formalin-fixed paraffin-embedded tumor samples from 10 patients with neuroendocrine tumors in the Peking University Shenzhen Hospital, China, are used in this work. The samples are stored in the archives of Department of Pathology in Peking University Shenzhen Hospital, and the head of the Department of Pathology approved the usage of the samples in this work. The samples are anonymized. All patient-related data and unique identifiers are removed. These procedures were performed under the supervision and approval of the Ethics Committee in Peking University Shenzhen Hospital.

From each formalin-fixed paraffin-embedded block, we cut two consecutive sections: one for staining with HE and the other for staining with the anti-Ki-67 antibody. During the HE staining process, we used undiluted Mayer's hematoxylin and 0.5% eosin. During the IHC staining process, we used anti-Ki-67 antibody (Roche United States).

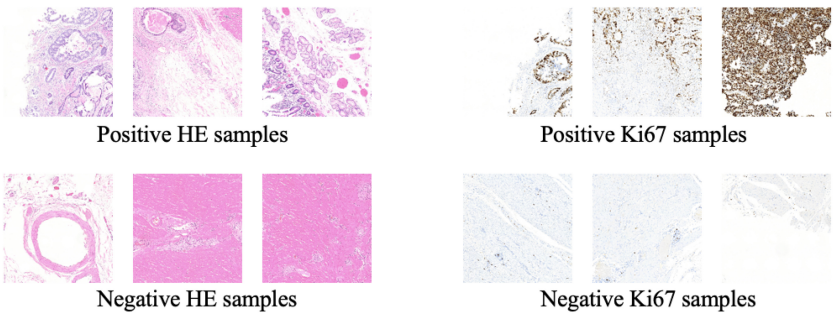


FIGURE 5 | Some training samples during MIL training process.

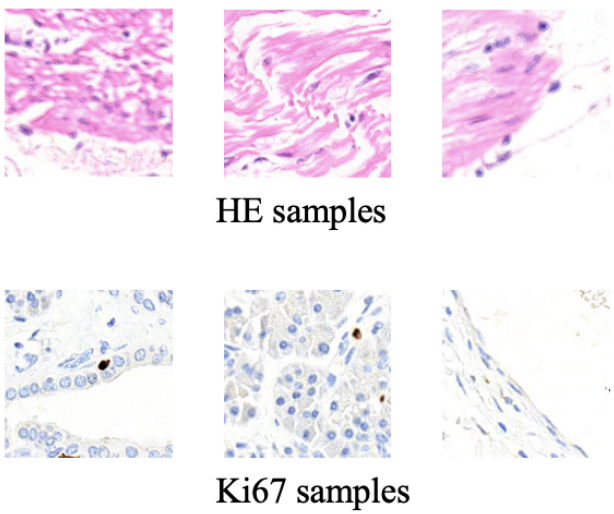


FIGURE 6 | Some training samples during CycleGAN training process.

In this way, we got 10 HE pathology microscopy images and 10 corresponding Ki-67 pathology microscopy images of neuroendocrine tumors. In all, 7 HE pathology microscopy images and 7 corresponding Ki-67 pathology microscopy images are used as training data, and the rest are used as evaluating data. In this paper, we list the representation of 2 HE pathology microscopy images and their corresponding Ki-67 pathology microscopy images (**Figure 2**).

Preprocess

During the MIL training period, for each domain (HE or Ki-67), we want to train a binary classifier to classify tumor patches or normal patches. However, our data are unannotated. So two preprocess steps are necessary. The first step is foreground extraction. There exist three kinds of patches from HE pathology microscopy images or Ki-67 pathology microscopy images: tumor patches including at least one tumor cell, normal patches including only normal cells, and background patches including only background. But the MIL algorithm is used to train a binary classifier to classify two kinds of patches, so the background patches should be removed. In this paper, an OTSU algorithm (Xu G. et al., 2019) is used to extract foreground from HE pathology microscopy images and Ki-67 pathology microscopy images. The second step is extracting large patches with weak annotations. In this paper, HE pathology slices and Ki-67 pathology slices are all positive slices containing tumor cells. But the MIL algorithm works with positive samples and negative samples with weak annotations. For this reason, the foreground of HE pathology microscopy images and Ki-67 pathology microscopy images is cropped into large patches (2,240 $\times$ 2,240 pixels), and then we can label these large patches manually. With this preprocess, we can get two weakly annotated datasets (one HE large patch dataset and one Ki-67 large patch dataset), and they can be used to train two classifiers (one is used in the HE domain, and the other one is used in the Ki-67 domain) using the MIL algorithm.

During the CycleGAN training period, the background of pathology microscopy images should not be removed because the generator is required to generate a synthetic background at the same time. The extraction of large patches is also removed. The pathology microscopy images should be cropped into small patches (224 $\times$ 224 pixels) directly.

Method

In this paper, we provide a schematic representation of our method (**Figure 3**). At first, we train two binary classifiers classifying tumor patches or normal patches with unannotated HE and Ki-67 pathology microscopy images using the MIL algorithm. During the CycleGAN training period, the input patches will be fed into the above classifiers in order first to get its class-related information. Then, following what conditional GAN does, class-related information will be considered as the additional channel of the input patch. This strategy will force class-related feature vectors to be aligned accurately during the adversarial training process. Note that our method is working with an unannotated and unpaired dataset, and thus it can be applied to many other tasks.

MIL for Getting Class-Related Information

There have been many methods based on deep learning technology for tumor cell detection in pathology microscopy images. A dataset with annotations from professional

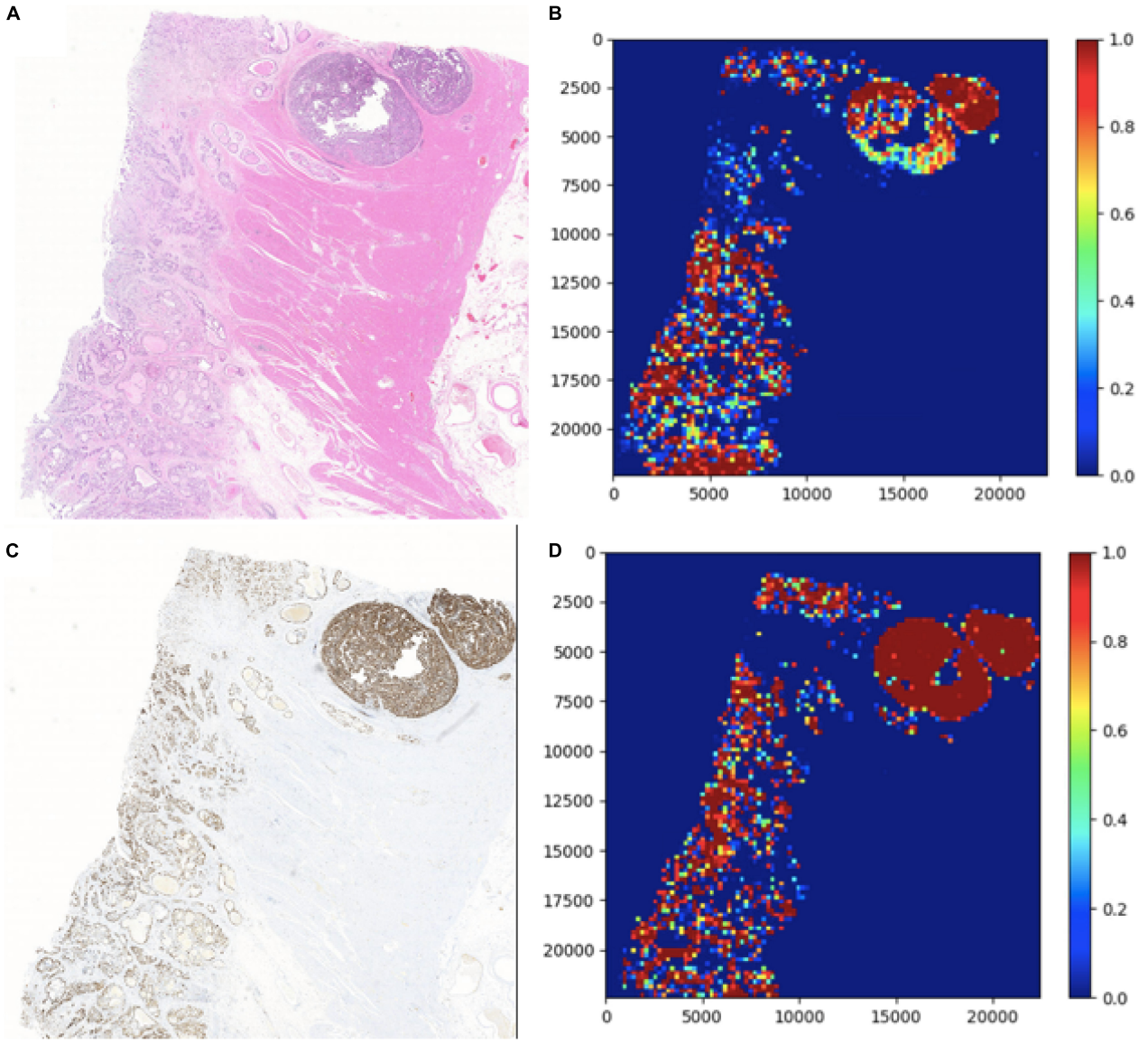


FIGURE 7 | The image-level visualization of classification results in HE and Ki-67 domain: **(A)** is an HE pathology microscopy image, **(B)** is its heatmap, **(C)** is a Ki-67 pathology microscopy image, and **(D)** is its heatmap. Red is for tumor patches, and blue is for normal patches.

pathologists is essential for the models' feasibility and performance, but it is too hard to get such a dataset. To address this drawback, in this paper, we apply the MIL algorithm to collect class-related information from an unannotated dataset.

During the MIL training period, the HE and Ki-67 pathology microscopy images are cropped into large patches (2,240 × 2,240 pixels), and these large patches are manually labeled as tumor patches or normal patches. A large patch is assigned a tumor label if it contains at least one tumor cell, while it is assigned a normal label if it contains only normal cells. After we have labeled all large patches, small patch extraction is necessary for low GPU memory, and each small patch is set at 224 × 224 pixels. Each iteration consists of an evaluation step and a training step. During the evaluation step, for each large patch, a classifier evaluates all small patches from this large patch and then one selected small patch is labeled. The small patch with the largest predictive probability is the one selected, and it should be considered the representative small patch for the large patch. For a large tumor patch, the selected small patch is assigned the tumor label. For large normal patch, the selected small patch is assigned the normal label. After the evaluation step, a training dataset of selected small patches with labels can be used to train the classifier. The trained classifier would evaluate all small patches in the next iteration. Finally, we will achieve two classifiers for classifying tumor patches or normal patches from HE and Ki-67 pathology microscopy images. And they can be used to get class-related information from unannotated data.

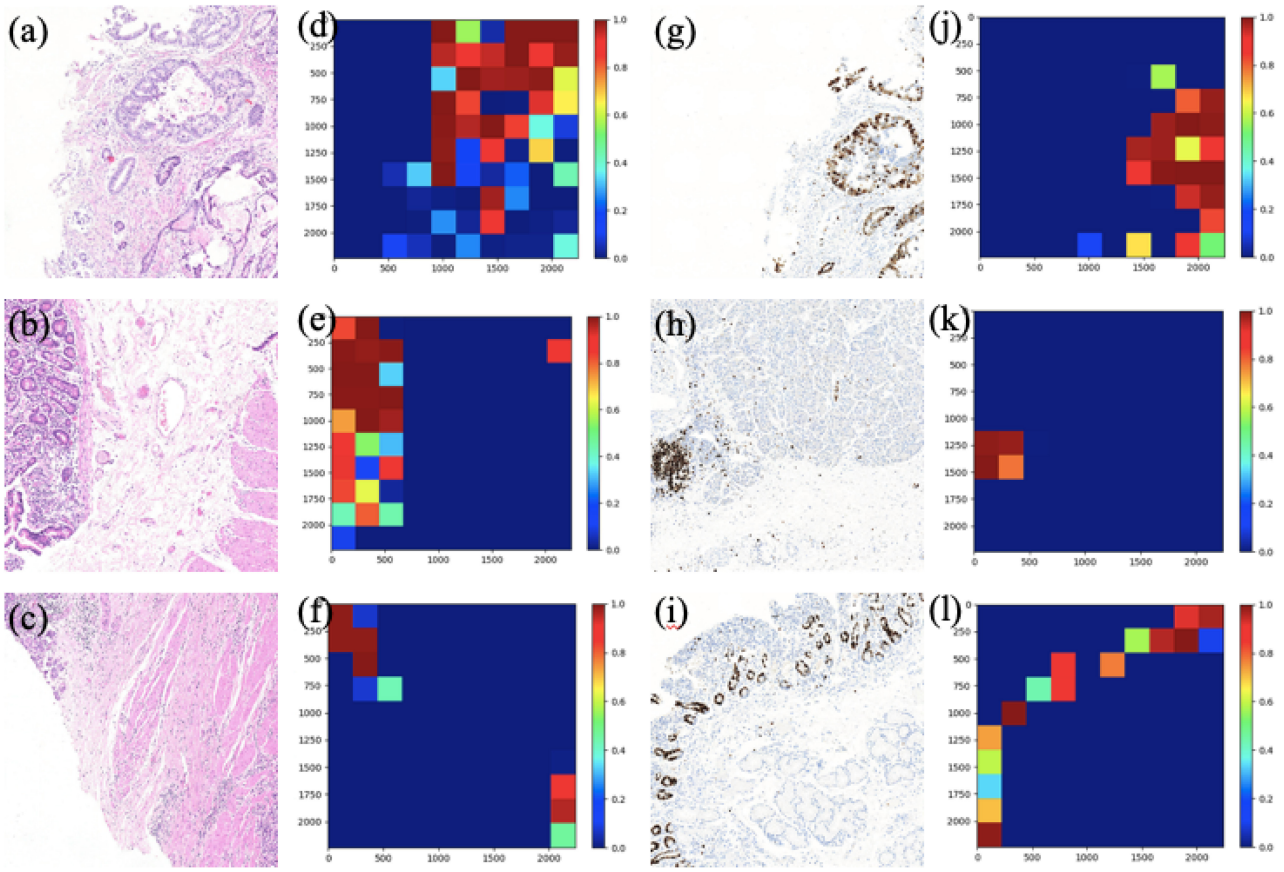


FIGURE 8 | The patch-level visualization of classification results in HE and Ki-67 domain. Here we list six cases in three rows, three HE cases (left), and three Ki-67 cases (right): **(a–c)** are the HE pathology patches; **(d–f)** are their heatmaps; **(g–i)** are the Ki-67 pathology patches; and **(j–l)** are their heatmaps. Red is for tumor patches, and blue is for normal patches.

During the MIL training process, the number of training samples is listed in **Table 1**. What's more, the classifiers used are ResNet34 (He et al., 2016).

CycleGAN for an Unpaired Dataset

In this paper, the corresponding Ki-67 pathology slice is the consecutive slice near the HE pathology slice. This means they are same mostly at the image level, but at the pixel level, they are different and unpaired, just as shown in **Figure 4**. As a result, CycleGAN is an appropriate solution for our task.

During the CycleGAN training period, HE and Ki-67 pathology microscopy images are directly cropped into small patches (224 × 224 pixels), including background. In this paper, the generator in raw CycleGAN has been replaced by a more complicated generator because we found that in raw CycleGAN, the discriminator would learn much faster than the generator when training with our data, so the generator could not learn anything and fails to generate synthetic Ki-67 pathology microscopy images of high quality.

During CycleGAN training process, the number of training samples is listed in **Table 1** and the detailed network configuration is shown in **Table 2**.

Conditional GAN for Class-Related Alignment

In the generation of synthetic images, the adversarial training process of GAN and its variants can be treated as learning an accurate mapping of some feature vectors between different domains. From the viewpoint of domain adaptation, the discriminator in GAN focuses on some feature vectors while neglecting other feature vectors that are less important. The situation is the same with CycleGAN. In our work, we found that raw CycleGAN would neglect class-related feature vectors during the adversarial training process. However, we need to consider the class-related information because Ki-67 pathology microscopy images are used to detect tumor cells. In order to align class-related feature vectors, the idea behinds conditional GAN is introduced to take class-related information into consideration (Xu Z. et al., 2019). In detail, the input patch of our CycleGAN is not only the HE microscopy patch or the Ki-67 microscopy patch, but also their class-related information generated from the classifier training with the MIL algorithm.

Some training samples during MIL training process and CycleGAN training process will be listed to better understand the differences between them. **Figure 5** shows some training samples

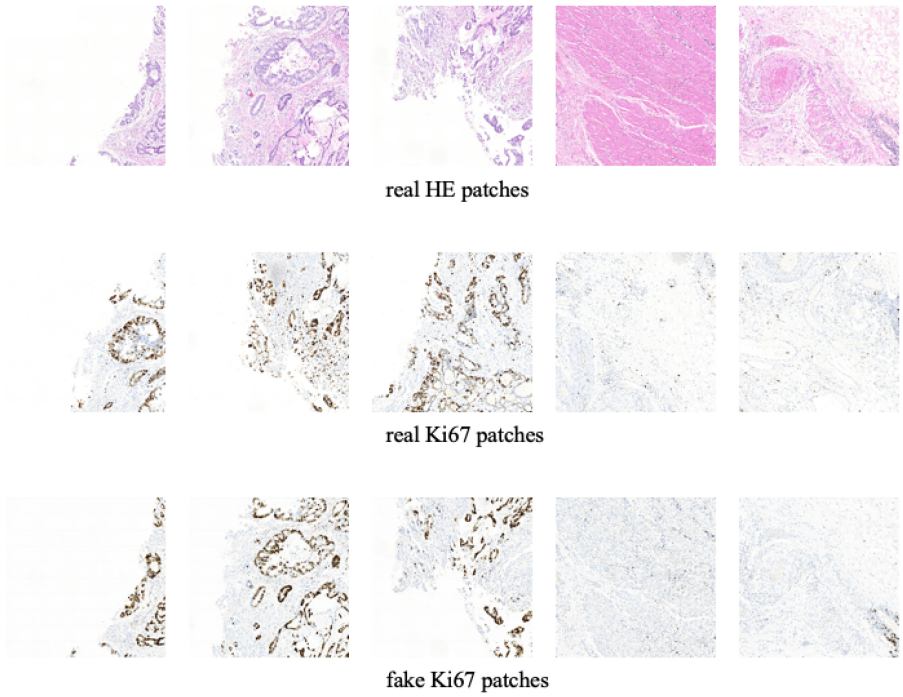


FIGURE 9 | The patch-level visualization of image generation results. Here we list five cases in five columns. At each column, the top image is the HE pathology patch, the middle image is the corresponding Ki-67 pathology patch, and the bottom one is the synthetic Ki-67 pathology patch generated with our proposed method.

during the MIL training process while **Figure 6** shows some training samples during the CycleGAN training process.

Training Configuration

Our method is implemented in the PyTorch framework on an Ubuntu platform. All experiments are conducted on a computer equipped with an NVIDIA GTX 2080 Ti graphic card with 11 GB of memory. During the training stage of our method, the deep neural networks are trained with Adam stochastic gradient descent algorithm. We use the learning rate of 0.0001 for 100 training epochs during the MIL training process and CycleGAN training process, and we will save the models each epoch. Among these 100 saved models, the one achieving the best result on the validation set is selected as the final model.

As for the loss function, cross entropy loss is used to train two classifiers during the MIL training process. During CycleGAN training process, just as with normal CycleGAN, the loss function for the discriminator is binary cross-entropy loss, and the loss function for the generator is mean square error loss.

RESULTS

Metrics

Unlike usual image generation tasks, in this paper, the evaluation metric is lacking because a pathology slice is not stained twice usually. It means that there is no way to get the pixel-aligned Ki-67 pathology microscopy image from a specific HE pathology microscopy image. To address this question, we have got the consecutive Ki-67 pathology slice near HE pathology slice. In this paradigm, the HE pathology microscopy image and its corresponding Ki-67 pathology microscopy image appear similar at image level. But they are different at the patch level. Considering this phenomenon, we evaluate our proposed method by image-level visualization and patch-level visualization. The image-level visualization can be used to evaluate the alignment of global feature vectors, and the patch-level visualization can be used to evaluate the alignment of class-related feature vectors. Moreover, we calculate the ratio of positive cells to all cells in a real Ki-67 pathology microscopy image, and its corresponding fake Ki-67 pathology microscopy image. These two ratios should be as close as possible.

Visualization of Patch Classification

In this section, two figures (**Figures 7**, **8**) show the experimental results of the binary classifiers in the HE domain and the Ki-67 domain using the MIL algorithm. **Figure 7** shows the classification results in the HE domain and the Ki-67 domain, and they are at the image level. The image-level results show the average classification performance. **Figure 8** shows the classification results in the HE domain and the Ki-67 domain, and they are at the patch level. The patch-level results show the classification performance in several regions with different densities of positive cells. From the visualization results, we can infer that we have got two classifiers for the tumor patch or the normal patch

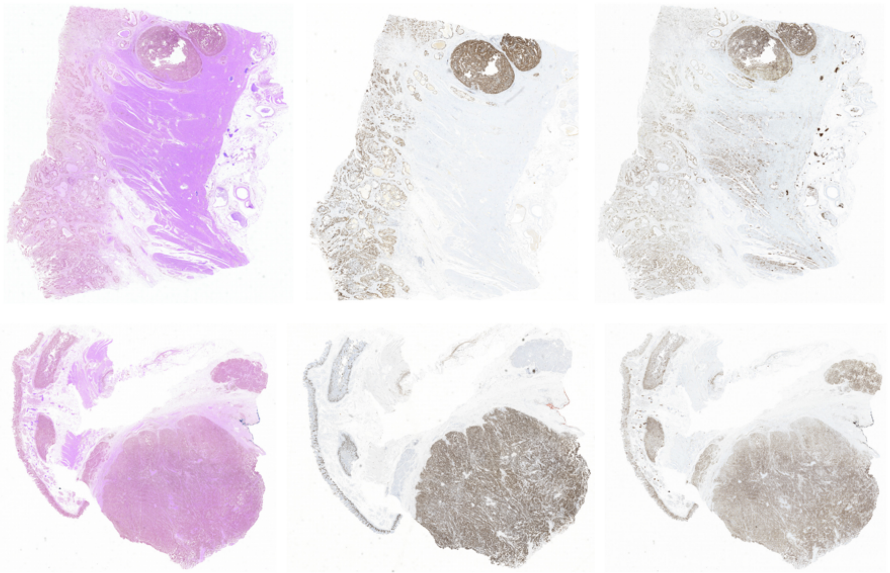


FIGURE 10 | The image-level visualization of image generation results. Here we show two different cases in two rows. In each row, the left image is the HE pathology microscopy image, the middle image is the corresponding Ki-67 pathology microscopy image, and the right one is the synthetic Ki-67 pathology microscopy image generated with our proposed method.

classification in the HE domain and the Ki-67 domain based on unannotated data.

Visualization of Image Generation

In this section, we show the experimental results of our proposed Ki-67 pathology microscopy images generation method from HE pathology microscopy images, including patch-level visualization and image-level visualization. **Figures 9**, **10** show the patch-level visualization and the image-level visualization of our experimental results. **Figure 9** shows five cases of our experimental results, in five columns. In each column, the top image is the HE patch, the middle image is the corresponding Ki-67 patch and the bottom image is the synthetic Ki-67 patch generated by our proposed method. Our proposed method is able to generate synthetic Ki-67 patches of high quality. By comparison with real Ki-67 patches, we can find that normal cells and tumor cells in HE patches can be transformed into normal cells and tumor cells in synthetic Ki-67 patches correctly. **Figure 8** shows our experimental results at image level. In **Figure 10**, we show two different cases, in two rows. In each row, the left image is the original HE pathology microscopy image, the middle image is the corresponding Ki-67 pathology microscopy image, and the right image is the synthetic Ki-67 pathology microscopy image generated with our proposed method. We can easily find that our proposed method can work well in regions including background or a high rate of positive cells or a medium rate of positive cells. However, this result also shows that it cannot work well in regions with a low rate of positive cells. Clinically, when we examine Ki-67 pathology microscopy images, the color, dark or light, of positive cells is not important because the doctors are asked to count positive cells to make a diagnosis. Quantification results taking this clinical usage into consideration will be listed by illustrating the ratio of positive cells to all cells.

TABLE 3 | Quantification results 1 of our proposed method.

	Training dataset		Evaluation dataset	
	Real Ki-67	**Synthetic Ki-67**	**Real Ki-67**	**Synthetic Ki-67**
Ratio	0.3744	0.3850	0.3046	0.2771

Ratio, ratio of the number of positive cells to the number of all cells. The results above show the potential value of our proposed method in clinical practice.

TABLE 4 | Quantification results 2 of our proposed method.

	Example 1		Example 2	
	Real Ki-67	**Synthetic Ki-67**	**Real Ki-67**	**Synthetic Ki-67**
Ratio	0.2570	0.2518	0.3522	0.3025

Ratio, ratio of the number of positive cells to the number of all cells. The results above show the potential value of our proposed method in clinical practice.

Quantification Results

In this section, the ratio of positive cells to all cells will be set as the evaluating metric for our proposed method because of the clinical usage of Ki-67 pathology microscopy images. Each test HE pathology microscopy image will be fed into the well-trained generator to generate a synthetic Ki-67 pathology microscopy image. For quantification results, we separately count the ratio above from real Ki-67 pathology microscopy image and its corresponding synthetic Ki-67 pathology microscopy image, and the results are listed in **Tables 3**, **4**.

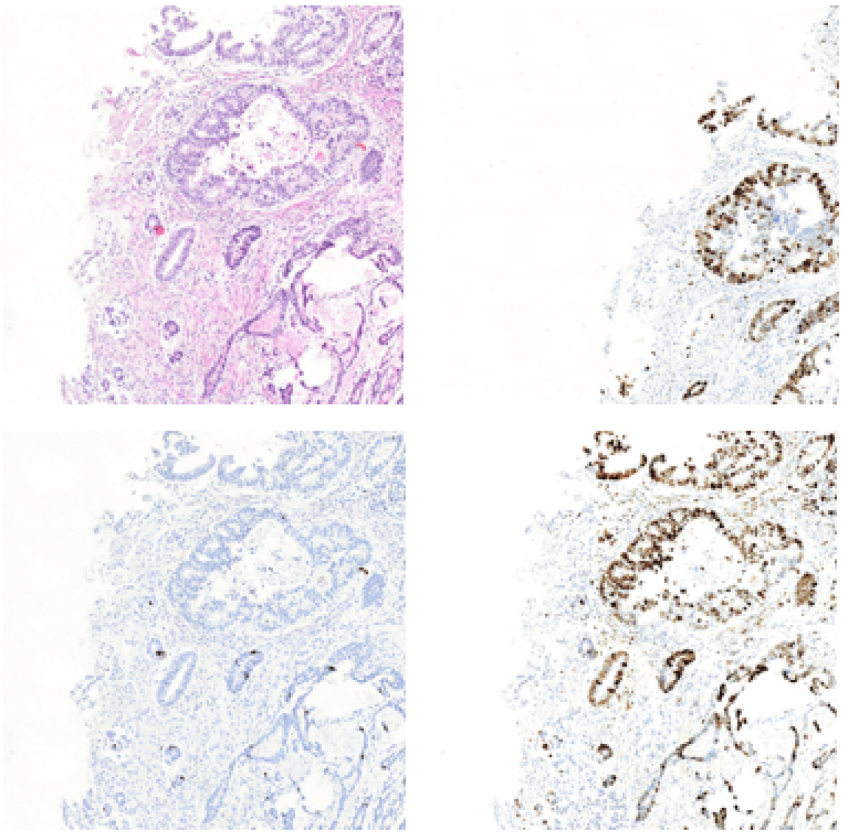


FIGURE 11 | The representation of image generation results using raw CycleGAN and our proposed method. The top left image is the HE pathology patch, and the top right one is its corresponding Ki-67 pathology patch. The bottom left image is the result generated with only CycleGAN and the bottom right one is the result generated with our proposed method.

Ablation Study

In this section, we will list the experimental results of Ki-67 pathology microscopy images generation from HE pathology microscopy images using only CycleGAN and our method. **Figure 11** shows the visualization of the experimental results from only CycleGAN and our method at patch level. The top left one is the real HE patch, the top right one is the corresponding Ki-67 patch, the bottom left one is the synthetic Ki-67 patch generated only with CycleGAN, and the bottom right one is the synthetic Ki-67 patch generated with our method. The comparison means that our proposed method is effective in Ki-67 pathology microscopy image generation from HE pathology microscopy images.

DISCUSSION

This is the first attempt to apply CycleGAN for synthetic Ki-67 pathology microscopy images generation with an unpaired dataset. Moreover, the MIL algorithm has been adopted to extract class-related information from an unannotated dataset. Importantly, the idea behind conditional GAN is used to force the class-related feature vectors to be aligned during the adversarial training process. With all these methods, our proposed method is able to generate synthetic Ki-67 pathology microscopy images of high quality. Because our proposed method is working with an unannotated and unpaired dataset, our method can be applied to many other regions.

Although we propose a synthetic Ki-67 pathology microscopy images generation method and the performance is good enough, there exist many future works for us to address. The first one is the evaluation metric, which can evaluate its clinical usage. The second one is a larger dataset, which is essential to performance improvement. When addressing these problems, we believe we can obtain a model that can be used in clinical practice.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

ZX designed and conducted the experiments, analyzed the data, and drafted the manuscript. YH, XL, XZ, and YC guided the

experiments and manuscript modification. All authors read and approved the final manuscript.

FUNDING

This research was supported by the National Science Foundation of China (NSFC) (81571837, 81871395, and 61675113), Science and Technology Research Program of Shenzhen City (JCYJ20170816161836562, JCYJ20160427183803458, JCYJ20170412171856582, JCYJ20170817111912585, and JCYJ20180508152528735), Oversea Cooperation Foundation, Graduate School at Shenzhen, Tsinghua University (HW2018007).

ACKNOWLEDGMENTS

Data was supported by the Peking University Shenzhen Hospital and Tsinghua University.

REFERENCES

Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A. P., Silva, V. W., Busam, K. J., et al. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* 25, 1301–1309. doi: 10.1038/s41591-019-0508-1

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018a). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 321, 321–331. doi: 10.1016/j.neucom.2018.09.013

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018b). "Synthetic data augmentation using GAN for improved liver lesion classification," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 289–293. doi: 10.1109/ISBI.2018.8363576

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial networks. *Adv Neural Inf. Process. Syst.* 3, 2672–2680.

Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., et al. (2018). "GAN-based synthetic brain MR image generation," in *International Symposium on Biomedical Imaging*, Washington, DC. doi: 10.1109/ISBI.2018.8363678

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

Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., et al. (2017). "Cycada: cycle-consistent adversarial domain adaptation," in *Proceedings of the 35th International Conference on Machine Learning, PMLR* 80, 1989–1998.

Kazeminia, S., Baur, C., Kuijper, A., Ginneken, B. V., and Mukhopadhyay, A. (2020). GANs for medical image analysis. *Artif. Intell. Medicine* 109:101938. doi: 10.1016/j.artmed.2020.101938

Li, Y., and Ping, W. (2018). Cancer metastasis detection with neural conditional random field. Medical Imaging with Deep Learning. [Preprint]. arXiv:1806.07064.

Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G. E., and Stumpe, M. C. (2017). "Detecting cancer metastases on gigapixel pathology images," in *International Conference on Medical Image Computing and Computer Assisted Intervention*. [Preprint]. arXiv:1703.02442.

Mahmood, F., Chen, R. J., and Durr, N. J. (2018). Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. *IEEE Trans. Med. Imaging* 37, 2572–2581. doi: 10.1109/tmi.2018.2842767

Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. [Preprint]. arXiv:1411.1784.

Nguyen, T. D., Le, T., Vu, H., and Phung, D. (2017). "Dual discriminator generative adversarial nets. NIPS", in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, United States, 2667–2677.

Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., et al. (2017). "Medical image synthesis with context-aware generative adversarial networks," in *Medical Image Computing and Computer Assisted Intervention*, Quebec City. doi: 10.1007/978-3-319-66179-7_48

Shin, H. C., Tenenholtz, N. A., Rogers, J. K., Schwarz, C. G., Senjem, M. L., Gunter, J. L., et al. (2018). *Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks. Simulation and Synthesis in Medical Imaging. SASHIMI 2018. Lecture Notes in Computer Science*, New York: Springer International Publishing, 11037, 1–11. doi: 10.1007/978-3-030-00536-8_1

Song, C., He, K., Wang, L., and Hopcroft, J. E. (2019). "Improving the generalization of adversarial training with domain adaptation," in *International Conference on Learning Representations*, Arizona: ICLR.

Wang, D., Khosla, A., Gargeya, R., Irshad, H., and Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. [Preprint]. arXiv:1606.05718.

Xu, G., Song, Z., Sun, Z., Ku, C., Yang, Z., Liu, C., et al. (2019). "CAMEL: a weakly supervised learning framework for histopathology image segmentation," in *International Conference on Computer Vision*, Piscataway, NJ: IEEE, 10682–10691. doi: 10.1109/ICCV.2019.01078

Xu, Z., Moro, C. F., Bozóky, B., and Zhang, Q. (2019). Gan-based virtual re-staining: a promising solution for whole slide image analysis. *arXiv*:1901.04059.

Zhu, J., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *International Conference on Computer Vision*, Venice. doi: 10.1109/ICCV.2017.244

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 11 November 2020
doi: 10.3389/fmolb.2020.599333

Mass Detection and Segmentation in Digital Breast Tomosynthesis Using 3D-Mask Region-Based Convolutional Neural Network: A Comparative Analysis

Ming Fan[1], Huizhong Zheng[1], Shuo Zheng[1], Chao You[2], Yajia Gu[2], Xin Gao[3], Weijun Peng[2]* and Lihua Li[1]**

[1] Institute of Biomedical Engineering and Instrumentation, Hangzhou Dianzi University, Hangzhou, China, [2] Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai, China, [3] Computational Bioscience Research Center (CBRC), Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

OPEN ACCESS

Edited by:
Heye Zhang,
Sun Yat-sen University, China

Reviewed by:
Chenxi Huang,
Xiamen University, China
Jianwen Luo,
Tsinghua University, China

***Correspondence:**
Ming Fan
ming.fan@hdu.edu.cn
Weijun Peng
cjr.pengweijun@vip.163.com
Lihua Li
lilh@hdu.edu.cn

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 27 August 2020
Accepted: 21 October 2020
Published: 11 November 2020

Citation:
Fan M, Zheng H, Zheng S, You C, Gu Y, Gao X, Peng W and Li L (2020) Mass Detection and Segmentation in Digital Breast Tomosynthesis Using 3D-Mask Region-Based Convolutional Neural Network: A Comparative Analysis. Front. Mol. Biosci. 7:599333. doi: 10.3389/fmolb.2020.599333

Digital breast tomosynthesis (DBT) is an emerging breast cancer screening and diagnostic modality that uses quasi-three-dimensional breast images to provide detailed assessments of the dense tissue within the breast. In this study, a framework of a 3D-Mask region-based convolutional neural network (3D-Mask RCNN) computer-aided diagnosis (CAD) system was developed for mass detection and segmentation with a comparative analysis of performance on patient subgroups with different clinicopathological characteristics. To this end, 364 samples of DBT data were used and separated into a training dataset (n = 201) and a testing dataset (n = 163). The detection and segmentation results were evaluated on the testing set and on subgroups of patients with different characteristics, including different age ranges, lesion sizes, histological types, lesion shapes and breast densities. The results of our 3D-Mask RCNN framework were compared with those of the 2D-Mask RCNN and Faster RCNN methods. For lesion-based mass detection, the sensitivity of 3D-Mask RCNN-based CAD was 90% with 0.8 false positives (FPs) per lesion, whereas the sensitivity of the 2D-Mask RCNN- and Faster RCNN-based CAD was 90% at 1.3 and 2.37 FPs/lesion, respectively. For breast-based mass detection, the 3D-Mask RCNN generated a sensitivity of 90% at 0.83 FPs/breast, and this framework is better than the 2D-Mask RCNN and Faster RCNN, which generated a sensitivity of 90% with 1.24 and 2.38 FPs/breast, respectively. Additionally, the 3D-Mask RCNN achieved significantly ($p < 0.05$) better performance than the 2D methods on subgroups of samples with characteristics of ages ranged from 40 to 49 years, malignant tumors, spiculate and irregular masses and dense breast, respectively. Lesion segmentation using the 3D-Mask RCNN achieved an average precision (AP) of 0.934 and a false negative rate (FNR) of 0.053, which are better than those achieved by the 2D methods. The results suggest that the 3D-Mask RCNN CAD framework has advantages over 2D-based mass detection on both the whole data and subgroups with different characteristics.

Keywords: digital breast tomosynthesis, breast cancer, mass detection, deep learning, mass segmentation

INTRODUCTION

Breast cancer is the most common malignancy in women. Full-field digital mammography (FFDM) is commonly used to screen for breast cancer (Nystrom et al., 2002). However, mammography has an inherent limitation when tissue overlaps, especially in dense breasts, which causes mammography to miss some suspicious cancerous lesions (Carney et al., 2003). Digital breast tomosynthesis (DBT) is an emerging breast cancer screening and diagnostic modality that takes quasi-three-dimensional imaging that can be used to provide a detailed assessment of the dense tissue within the breast. DBT has decreased the effect of overlapping tissue on screening, thereby improving lesion detection, characterization and diagnosis and making this approach superior to digital mammography (DM) (Michell et al., 2012; Haas et al., 2013). The integration of DBT into the diagnostic setting is associated with improved diagnostic performance of breast cancer due to the increased specificity (Bahl et al., 2019; Conant et al., 2019). The combination of DBT and mammography resulted in significant gains in the sensitivity and specificity of cancer detection compared with DM alone (Fontaine et al., 2019; Li et al., 2019; Skaane et al., 2019). Due to its improvements in patient diagnosis efficiency, DBT is becoming the standard of care in both screening and diagnostic breast images (Chong et al., 2019).

Early detection of masses on DBT can facilitate improved treatment and management in breast cancer. Additionally, segmentation of breast masses from the background tissue is important for accurate mass characterization and interpretation. However, the increment of the 3D information of breast tissue for DBT also increases the image reading workload by 2-fold (Tagliafico et al., 2017). Manual detection/segmentation of the breast region is therefore becoming impractical under a large number of samples/slices. Consequently, there is a need for computational methods to assist in the evaluation of DBT, both to address the workload issues and to maximize the performance of cancer detection and segmentation.

To this end, studies developed computer-aided diagnosis (CAD) system in DBT to facilitate mass detection and/or segmentation in a clinical setting. The conventional CAD studies have focused on 2D analysis of the slices of DBT using a variety of hand-crafted features (Reiser et al., 2006; Varela et al., 2006). Previous study used classical seed region growing algorithm to enhance the contour of a mass from a given region of interest (ROI) with the ability to adaptively adjust the threshold value (Berber et al., 2013). A Gaussian mixture models based on handcrafted intensity and texture measures were developed to segment breast masses in DBT (Pohlmann et al., 2017).

Compared to the conventional CAD using handcrafted features, deep learning-based CAD methods, which are based on end-to-end learning using a large amount of data, have an important role in DBT (Geras et al., 2019) due to their accuracy and efficiency. The deep CAD framework is reported to achieve much better performance than that achieved by using handcrafted features to detect masses in DBT (Yousefi et al., 2018). Moreover, a layered pathway evolution method was proposed to compress a deep convolutional neural network (DCNN) to classify masses in DBT (Samala et al., 2018). Previous studies developed a CAD system for mass detection and diagnosis using a DCNN with transfer learning from mammograms (Samala et al., 2016, 2019). A U-net based deep architecture was utilized to automatically segment breast masses on DBT data (Lai et al., 2020). To improve efficacy and accuracy in deep learning-based mass detection/segmentation, recent studies used CAD system based on one of the most successful object detection method, Faster RCNN [24] on mammograms (Ribli et al., 2018) and DBT (Fan et al., 2019). The existing studies were mainly performed using a DCNN based on 2D slices of DBT images for mass detection/segmentation. Nevertheless, volumetric, higher-dimensional information are more complicated so as to capture more sufficient, high-level features from 3D images. However, whether the 3D deep learning methods are superior to the traditional mass detection methods remains unknown.

There is also controversy regarding the efficiency of CAD methods for detecting masses in DBT from patients with different characteristics. For example, the DBT increases the cancer detection rate but is less effective for women with extremely dense breasts (Vourtsis and Berg, 2019). A recent study reported that DBT enabled the detection of more cancers in all density and age groups compared with DM, especially cancers classified as spiculated masses and architectural distortions (Osteras et al., 2019). DBT and DM screening increased the detection rate of histologically favorable tumors compared with that attained by DM screening (Hofvind et al., 2018). Therefore, it is of great interest to evaluate and compare the performances of deep learning-based mass detection and segmentation methods using DBT in patients with various characteristics, including different age ranges, breast densities, mass shapes and mass sizes.

Here, we proposed a framework for a 3D-MaskRCCN-based CAD system extended from our previous work of Faster RCNN on 2D slices of DBT (Fan et al., 2019), for the detection and segmentation of breast masses. To evaluate the effectiveness of 3D mask detection, we compared the results of the 3D-Mask RCNN, 2D-Mask RCNN, and Faster RCNN on images from patients with different characteristics. Our study was performed to enhance the efficiency and effectiveness mass detection/segmentation with DBT data and to facilitate an improved understanding of the 3D deep learning-based methods on different types of breast cancers.

MATERIALS AND METHODS

Histological Analysis

Malignant and benign tumors were determined by biopsies using histological analysis. The breast density was determined according to the Breast Imaging Reporting and Data System (BI-RADS) ACR categories and/or quantification, which ranged from 1 to 4. Breasts with up to 25% mammary gland parenchyma were classified as ACR 1 (almost entirely fat), and those with 26–50% gland parenchyma (average density) were defined as ACR 2. The breasts with 51–75% gland parenchyma were classified as ACR 3 and those with more than 75% gland parenchyma (high density) were classified as ACR 4. The ACR type 3 and 4 breasts were

categorized as dense breasts while ACR type 1 and 2 breasts were categorized as non-dense breasts.

Dataset

The imaging and clinical data were collected from Fudan University Affiliated Cancer Center with Institutional Review Board (IRB) approval. **Table 1** shows the characteristics of the samples used in this study. A total of 364 samples were collected (the mean age was 52.31 years, the age ranged from 18 to 88 years, and the median age was 51 years). Among samples, 75 were benign and 289 were determined to be malignant tumors by biopsy. The dense and non-dense breasts represented 75.8 and 24.2% of the total samples, respectively. There were 123 round/oval, 113 spiculate and 128 irregular masses. The data were randomly separated into the training dataset (n = 201) to train a deep learning-based CAD system and the testing dataset (n = 163) to test the effectiveness of the CAD. There were no significant differences in the ages, histologic types, mass types, and breast densities of the training and the testing datasets ($p > 0.05$).

Craniocaudal (CC) and mediolateral oblique (MLO) view DBT images were acquired by a Selenia Dimensions TM unit (Hologic, American) using a total tomographic angular range of 20° with a 1° increment of rotation and 20 projection views (PVs). The number of slices ranged from 20 to 124 (mean = 62.70 and median = 62), and the number of slices containing a lesion ranged from 6 to 111 (mean = 34.85 and median = 32). For each slice, the in-plane resolution was 106 × 106 μm. The entire DBT data set included 716 views from 364 breasts with 364 masses. The breast mass sizes ranged from 10.15 to 140.90 mm (mean = 36.45 mm and median = 33.40 mm).

Image Preprocessing

Digital breast tomosynthesis images were reconstructed into a unified spacing slice (1.0 mm) using the simultaneous algebraic reconstruction technique (SART) (Zhang et al., 2006). To save computational memory and avoid the calculation of large-scale convolutions for the background pixels in the deep learning-based CAD system, the skin and the background were excluded from the breast region using a dynamic multiple thresholding-based breast boundary method (Wu et al., 2010; Kus and Karagoz, 2012).

Mask RCNN-Based CAD System

3D-Mask RCNN Architecture

As an extension of the Faster RCNN (Ren et al., 2017) which performs the object detection of rectangular boxes as both a regression and classification problem task, Mask RCNN adds an additional branch that outputs the object mask (He et al., 2020). We developed a 3D-Mask RCNN-based breast mass detection and segmentation model, which is shown in **Figure 1**. Due to the substantial amount of memory required by the 3D convolution kernel in the network, we used small regions of the images (referred to as patches) with sizes of 256 × 256 × 64. The patches were used for training the 3D-Mask RCNN to obtain the mass detection model. Then, the model was applied to the patches in the testing set, which were subsequently recombined to reconstruct the entire DBT. The prediction probabilities of each patch are used to obtain the mass probability for the DBT, and the probable mass region was obtained by a bounding box.

The original Mask RCNN model was modified into the 3D version (**Figure 1**). In the network, the Residual Networks

TABLE 1 | Patient clinicopathological characteristics.

Characteristics	All (n = 364)	Training dataset (n = 201)	Testing dataset (n = 163)	p
Age (Years)				0.807
<40	47 (12.91%)	22 (6.04%)	25 (6.87%)	
40–49	116 (31.87%)	65 (17.86%)	51 (14.01%)	
50–59	96 (26.37%)	54 (14.84%)	42 (11.54%)	
60–69	76 (20.88%)	43 (11.81%)	33 (9.07%)	
>70	29 (7.97%)	17 (4.67%)	12 (3.30%)	
Histological type				0.98
Benign	75 (20.60%)	42 (11.54%)	33 (9.07%)	
Malignant	289 (79.40%)	159 (43.68%)	130 (35.71%)	
Mass type				0.73
Round/oval	123 (33.79%)	69 (18.96%)	54 (14.84%)	
Spiculate mass	113 (31.04%)	59 (16.21%)	54 (14.84%)	
Irregular	128 (35.16%)	73 (20.05%)	55 (15.11%)	
Breast density				0.63
Non-dense[1]	276 (75.82%)	51 (14.01%)	37 (10.16%)	
Dense[2]	88 (24.18%)	150 (41.21%)	126 (34.62%)	
Tumor maximum diameter (mm)				0.38
$10 \leq d < 30$	267	154	113	
$30 \leq d < 50$	337	184	153	
$50 \leq d$	112	56	56	

[1]*ACR 1 and 2.*

[2]*ACR 3 and 4.*

The differences in the patient characteristics of the groups in the training and testing datasets were compared using a χ^2 test.

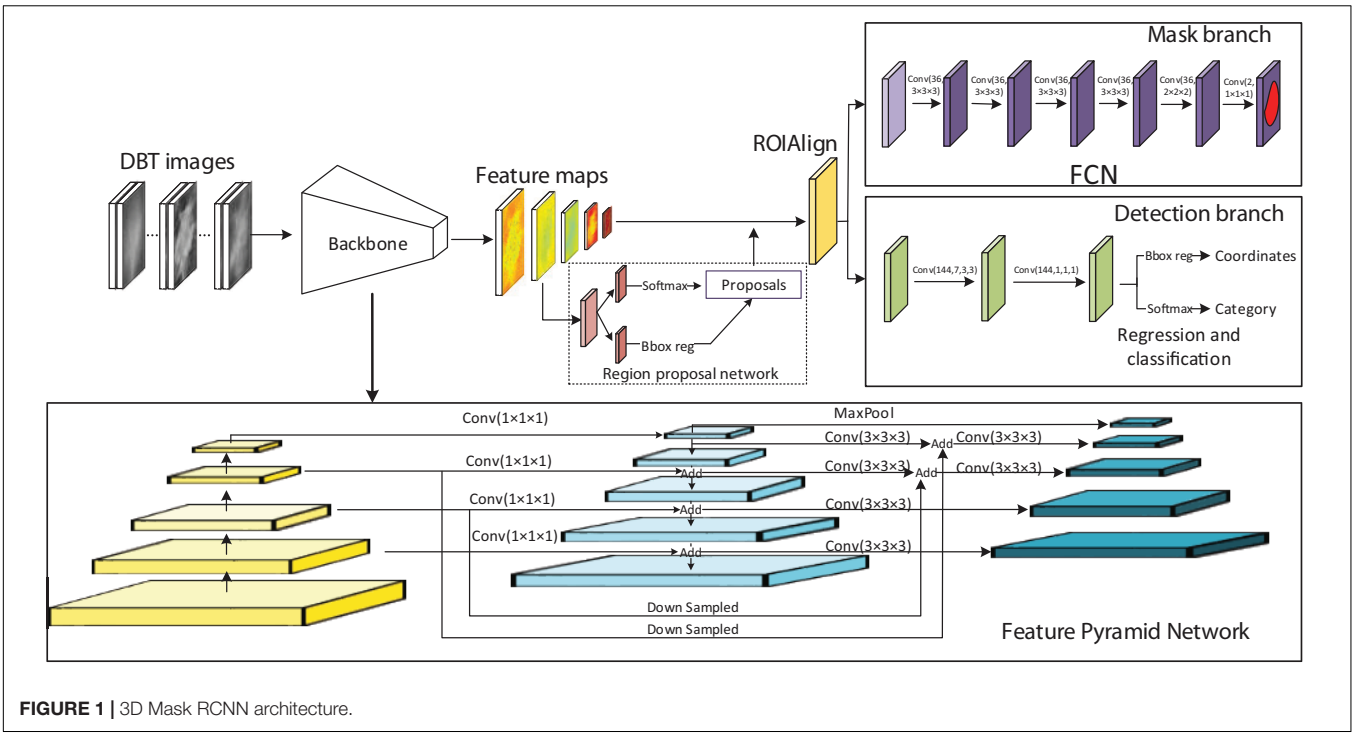


FIGURE 1 | 3D Mask RCNN architecture.

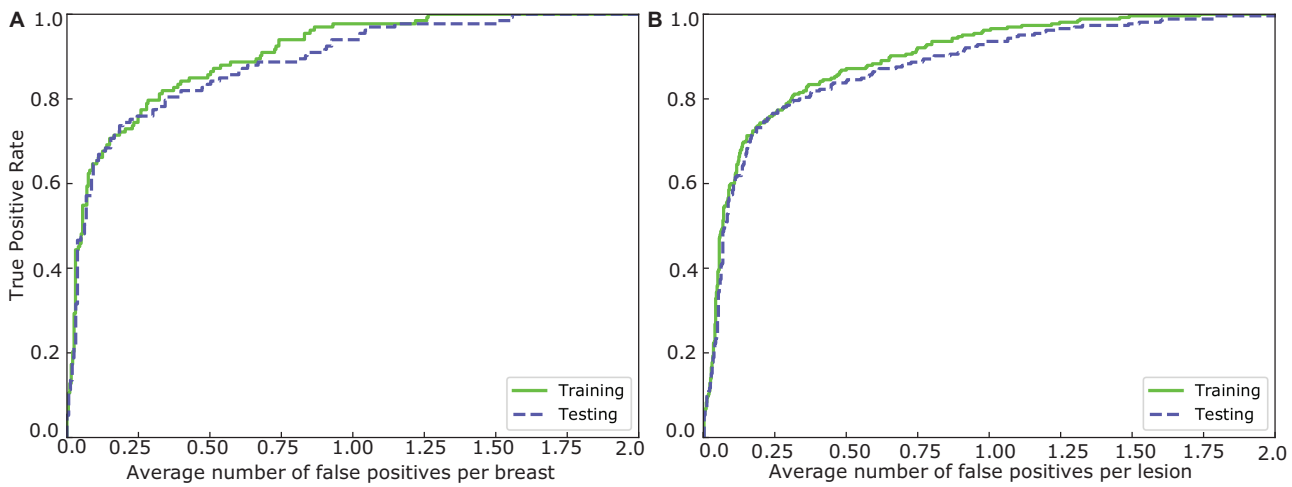


FIGURE 2 | FROCs for the training and testing sets in **(A)** breast-based and **(B)** lesion-based mass detection.

(ResNet)-Feature Pyramid Network (ResNet-FPN) backbone was used to extract different scales of the feature pyramid (He et al., 2016; Lin et al., 2017). FPN combines bottom-up features with top-down features in different scale. The ResNet has a structure of a depth of 50 layers in the 4th stage (ResNet-50-C4). This ResNet along with the FPN improve both the accuracy and speed of feature extraction. A region proposal network (RPN) was used to generate candidate bounding boxes from the input image. A quantization-free layer, i.e., RoIAlign was adopted to align the extracted feature maps with the inputs. This layer reduces misalignments between ROIs and the extracted features of RoIPool layer. The detection branch conducted mass detection for each proposed ROIs using a classifier network and bounding-box regression to obtain the probabilities and position information for the boxes. The mask branch obtained probabilities and position information from the feature maps and predicted a segmentation mask from each ROI using a fully connected network (FCN) in a pixel-to-pixel manner. The rectified linear unit (ReLU) were used as activation function in all the layers. The 3D-Mask RCNN were compared with 2D-Mask RCNN, Faster RCNN and Spatial Fuzzy C-Means (SFCM) (Zhang and Li, 2014) methods.

TABLE 2 | Mean numbers of FPs per lesion and breast at different sensitivities from the FROC curves.

Sen (%)	3D DBT						2D DBT			
	3D-Mask RCNN		2D-Mask RCNN		Faster RCNN		2D-Mask RCNN		Faster RCNN	
	L	B	L	B	L	B	L	B	L	B
60	**0.10**	**0.08**	0.22	0.19	0.29	0.17	**0.35**	0.28	0.83	**0.17**
70	**0.16**	**0.22**	0.38	0.34	0.54	0.33	**0.72**	0.68	1.63	**0.33**
80	**0.33**	**0.34**	0.69	0.77	1.18	0.97	**1.25**	1.26	2.99	**0.97**
85	**0.57**	**0.57**	0.91	1.06	1.66	1.49	**1.95**	2.33	4.26	**1.49**
90	**0.80**	**0.83**	1.30	1.24	2.37	2.38	**2.77**	3.08	5.36	**2.38**
95	**1.10**	**1.02**	2.38	2.80	4.05	3.30	**4.25**	4.28	7.04	**3.30**

Sen, sensitivity; L, lesion-based mass detection; B, breast-based mass detection. The lowest FP values for the mass detection methods are in bold.

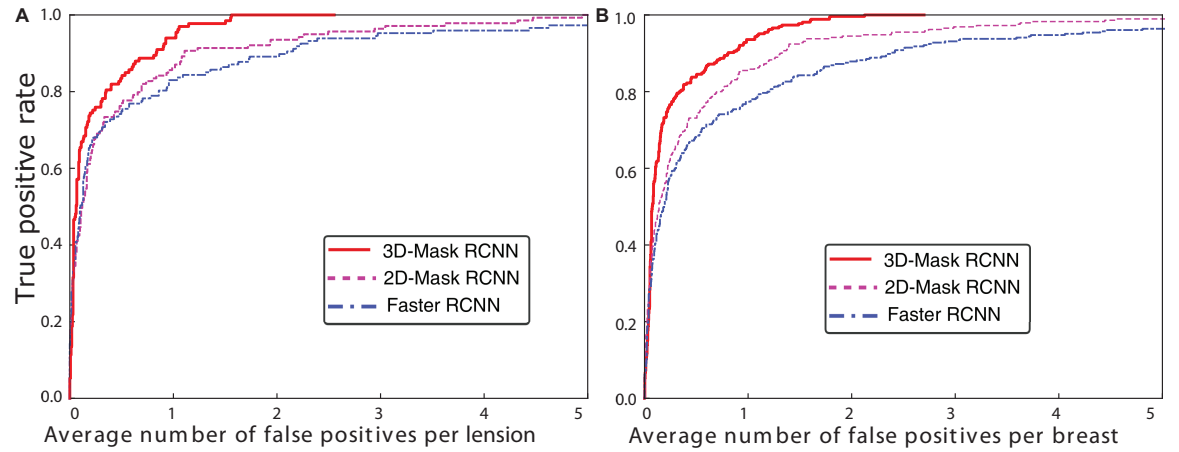


FIGURE 3 | FROC curves for the 3D-Mask RCNN-, Faster RCNN- and 2D-Mask RCNN-based CAD systems in **(A)** lesion-based and **(B)** breast-based detection.

3D-Mask RCNN Training

The hyperparameters of the networks including batch size, RPN train anchors, number of epochs, learning rate, and backbone are shown in **Supplementary Table S1**. To train the Mask RCNN, the network was initialized using the strategy proposed by He et al. (2015) which has good performance and was trained using the Adam optimizer (Kingma and Ba, 2014). For mass segmentation, the mask loss was defined only with the positive ROI. The initial learning rate was 0.001 and it was reduced by a factor of 0.5 after every 50 epochs. This learning rate was changed during training to achieve increased performance and faster training. Each mini-batch had 32 proposed ROIs. A mask branch for predicting an object mask was added to the RCNN. The same end-to-end training was performed to jointly train the RPN and the whole network.

In the training of the Faster RCNN, which was used in comparison with our framework, the weights were randomly initialized using a zero-mean Gaussian distribution. The initial learning rate determined by experiments was set as 0.001 for all the layers and was reduced by a factor of 0.5 after every 50 epochs. Each mini-batch had 2 images per GPU with 256 sampled ROIs. The loss function was divided into two parts: the first part was the classification loss, and the second part was the bounding box loss, with the Smooth L1 regularization defined in Ren et al. (2017). End-to-end training that jointly trains the RPN and Faster RCNN was performed to train the whole network.

Performance Analysis

The detected target was compared with the true masses marked by an experienced radiologist. More specifically, an experienced radiologist manually annotated the 3D bounding boxes, and the true positive (TP) objects were represented by the ROIs extracted from the radiologist-marked locations. The background or non-mass regions were labeled as negative cases. For mass segmentation, the detection was determined to be a TP if its intersection over union (IOU) for the true masses was greater than 50%. The ratio of the positive to negative ROIs was 3 to 2.

We calculated a free-response receiver operating characteristic (FROC) curve defined as the plot of the sensitivity versus the average number of false positives per breast/lesion. The FROC curve was computed by varying the thresholds of the object prediction confidence (Bandos et al., 2010). The lesion-based FROC (the same lesion imaged in the CC and MLO views was regarded as different targets for detection) and the breast-based curve (the same lesion imaged in two views of a breast was considered to be one target and the detection of one or both was regarded as a TP) were both assessed. The average precision (AP) and false negative

TABLE 3 | Comparison of the FPs of the 2D-Mask RCNN, 3D-Mask RCNN and the Faster RCNN-based CAD at a sensitivity of 90%.

Characteristic	2D-Mask RCNN[1]	Faster RCNN[1]	3D-Mask RCNN[1]	3D-Mask RCNN vs. 2D-Mask RCNN[2]	3D-Mask RCNN vs. Faster RCNN[2]
Age					
<40	1.6	2.8	**1.00**	0.291	0.1
40–49	2.62	2.76	**1.21**	**0.024**	**0.026**
50–59	2.83	2.09	**0.69**	0.100	**0.047**
60–69	0.27	1.3	0.48	**0.015**	0.062
>70	0	0.08	**0**	0.097	0.192
Histological type					
Benign	1.6	4.09	**1.36**	0.058	**0.029**
Malignant	1.28	1.76	**0.67**	**0.015**	**0.019**
Mass type					
Round/oval	0.78	1.2	**0.41**	0.061	**0.017**
Spiculate mass	0.9	1.7	**0.75**	**0.015**	**0.022**
Irregular	3	4.54	**1.20**	**0.023**	**0.016**
Breast density					
Non-dense	**0.18**	1	0.35	0.101	0.064
Dense	2.47	2.82	**1.03**	**0.005**	**0.010**
Maximum diameter (mm)					
$10 \leq d < 30$	1.3	3.15	**0.77**	**0.007**	**0.030**
$30 \leq d < 50$	1.26	2.24	**0.60**	**0.007**	**0.010**
$d \geq 50$	1.98	2.71	**1.19**	**0.037**	**0.039**

[1]False positive (FP) values at a sensitivity of 90%. [2]P values are provided for comparisons of differences between the area under the free-response receiver operating characteristic (FROC). The lowest FP values and significant p values are in bold.

rate (FNR) were used to evaluate the effectiveness of the segmentation methods.

The comparisons of the performances of the two CAD systems were conducted by calculating the differences between the area under the FROC (Bornefalk and Hermansson, 2005; Bandos et al., 2009) using the Bootstrap test to resample the prediction score of the detection system under non-parametric assumptions. The statistical significance of the performance difference between our 3D-Mask RCNN and the other two 2D deep CAD systems was estimated based on the breast-based FROC curves.

Ten-fold cross-validation (CV) was used for the training dataset to tune the hyperparameters of the deep learning-based CAD system. In each CV cycle, the deep learning-based CAD systems were trained using nine subsets as the training set and one subset as the testing set. The hyperparameter with the best performance in the 10-fold CV was used to train the CAD system using all the samples in the training set. Then, the trained model was applied to the testing dataset to evaluate its effectiveness on mass detection/segmentation.

RESULTS

Assessment and Comparison of Mass Detection Methods

Comparison of the Deep Mass Detection Methods on All Samples in Testing Set

Mass detection was performed on all the samples in the testing set using the 3D-Mask RCNN, and the results were compared with those of the 2D-Mask RCNN and Faster RCNN (**Figure 2**). The results of the 3D-Mask RCNN CAD system achieved a sensitivity of 90% with 0.80 FPs/lesion and 0.83 FPs/breast, respectively. The mean numbers of FPs per breast at different sensitivities, which were determined based on the FROC curves, are shown in **Table 2** and **Figure 3**. Our 3D-Mask RCNN-based CAD system clearly has better detection performance (in terms of fewer FPs) than that of the 2D-Mask RCNN or the Faster RCNN at all the sensitivities.

For these methods, the detection performances were compared, and statistical tests showed a significant difference in the areas under the breast-based FROC curves between the 3D-Mask RCNN and 2D-Mask RCNN methods (p = 0.005). Moreover, the 3D-Mask RCNN method achieved significantly better detection performance than that of the Faster RCNN-based system with a p value of 0.007. These results demonstrated that 3D-Mask RCNN-based CAD outperformed the 2D methods of the 2D-Mask RCNN and Faster RCNN (**Supplementary Table S2**).

In addition to the deep learning-based mass detection using 3D information of DBT data, the effectiveness of these methods is also examined on 2D slices of DBT. To this end, we evaluated the detection performances of the 2D-Mask RCNN and Faster RCNN using the imaging slice that shows the lesion with maximum diameter among that of all the DBT slices. The results showed that the Mask RCNN achieved better performance than the Faster RCNN in terms of fewer FPs/lesion, whereas an inverse result was observed for the breast-based evaluation, which showed fewer FPs/breast for the Faster RCNN compared with the Mask RCNN (**Table 2**). The two methods for both the lesion- and breast-based mass detection on 2D slices of DBT showed lower performance detection compared with that based on 3D volume of DBT (**Table 2**).

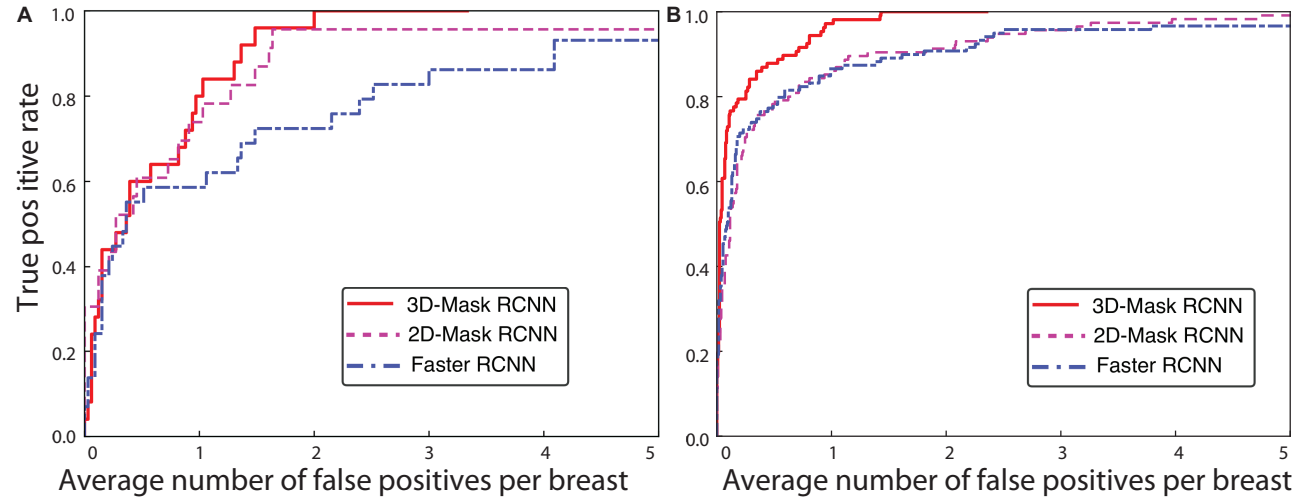


FIGURE 4 | FROC curves for the 3D-Mask RCNN-, Faster RCNN- and 2D-Mask RCNN-based CAD systems for **(A)** benign and **(B)** malignant tumors.

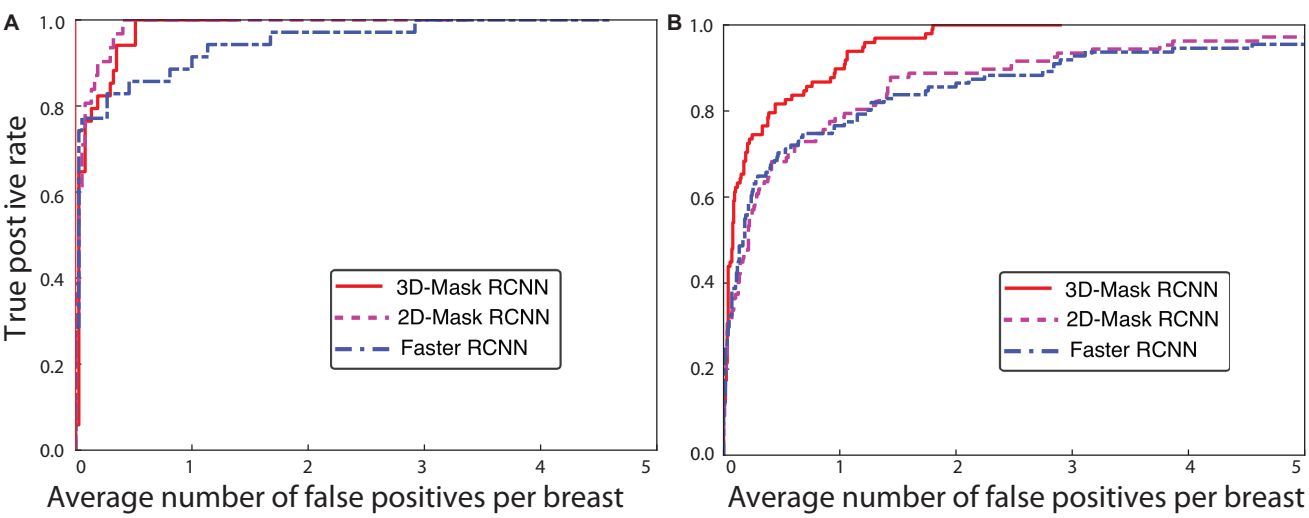


FIGURE 5 | FROC curves for the 3D-Mask RCNN-, Faster RCNN- and 2D-Mask RCNN-based CAD systems for patients with **(A)** non-dense and **(B)** dense breasts.

Comparison of the Deep Mass Detection Methods on Patients With Various Characteristics

A comprehensive comparison of the mass detection results of the 3D and 2D deep learning methods was performed on samples with different patient characteristics. The FPs per breast at a sensitivity of 90% for the CAD systems are shown in **Table 3**. The performances of the 3D-Mask RCNN, 2D-Mask RCNN and Faster RCNN on patients with different characteristics including benign/malignant tumors, breast densities and ages are shown in **Figures 4–6**, respectively.

The 3D-Mask RCNN achieved fewer FPs/breast for almost all the age ranges. For the patients with ages from 40 to 49 years, the 3D-Mask RCNN had significantly better performance than that of the 2D-Mask RCNN (p = 0.024) and the Faster RCNN (p = 0.026). Additionally, the mass detection performance in terms of the fewest FPs/breast was higher for malignant tumors than benign tumors for all the detection methods (**Figure 4**). The 3D-Mask RCNN also achieved significantly better performance in the detection of malignant tumors than the 2D-Mask RCNN and the Faster RCNN with p values of 0.015 and 0.019, respectively. All the methods achieved lower mass detection performance for irregular tumors than other mass types with the highest FPs/breast. Among the mass types, the 3D-Mask RCNN model achieved significantly better detection performance on spiculate or irregular masses than either the 2D-Mask RCNN or the Faster RCNN-based CAD system ($p < 0.05$).

Furthermore, it was observed that the detection performance is lower in samples with dense breasts than those with non-dense breasts in terms of low FPs for all the detection

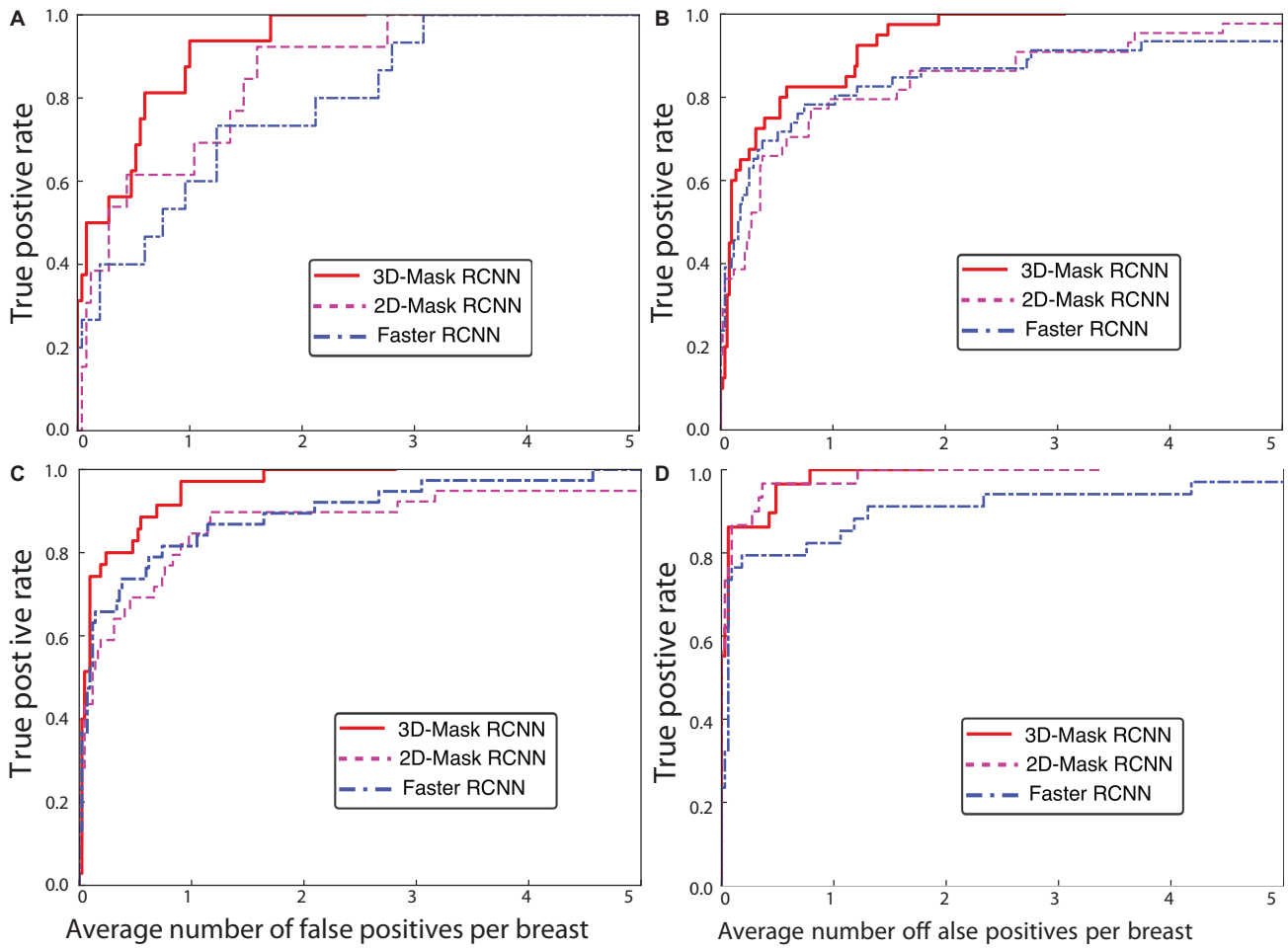


FIGURE 6 | FROC curves for the 3D-Mask RCNN-, Faster RCNN- and 2D-Mask RCNN-based CAD systems for patients with ages **(A)** less than 40 years, **(B)** 40–49 years, **(C)** 50–59 years, and **(D)** 60–69 years.

methods (**Figure 5**). The 3D-Mask RCNN achieved significantly better detection performance than the 2D-Mask RCNN- and the Faster RCNN-based (p = 0.005 and 0.010, respectively) CAD in patients with dense breasts. Furthermore, deep learning-based mass detection performs better for larger masses than smaller masses (**Figure 6**). Again, our 3D-Mask RCNN method has better mass detection performance than the other methods for all the diameter sizes ($p < 0.05$).

Case Study of the Mass Detection of the Deep Learning-Based Mass Detection Methods

Figure 7 shows the examples of the mass detection for the 3D-Mask RCNN, 2D-Mask RCNN and Faster RCNN methods. From this figure, tumors with low densities are easier for the CAD system to detect (**Figures 7A–C**). The three methods also showed high prediction scores on dense breast with characteristic of malignant, oval and small tumor (**Figure 7D**). However, the Faster RCNN method showed a false positive detection result (**Figure 7C**) while the 2D Mask RCNN showed lower detection score than the other methods (**Figure 7A**). **Figure 8** illustrates the mass detection results in patients with dense breast, but with different age ranges, mass shapes and histological types. The results showed that 3D deep mass detection achieved better performance than those of the other two 2D methods, which failed to detect masses in patients with dense breasts and spiculate tumors. However, all three methods failed to detect the masses of patients with large lesion sizes and dense breasts (**Figure 8C**). Additionally, the results showed low detection scores for all the three methods on patients with large lesion sizes, dense breasts and irregular shapes (**Figure 8D**). **Figure 9** illustrates the detection results for patients with dense breasts. Our 3D-Mask RCNN CAD system outperformed the other 2D methods on the four cases. The 2D-Mask RCNN and Faster RCNN achieved lower detection performance than the 3D-Mask RCNN in the detection of large tumors (**Figures 9B,C**). Compared with the 3D method, it is more difficult for the 2D or Faster RCNN to discriminate lesions and background regions in patients with dense breasts with smaller mass size (**Figures 9 A,D**).

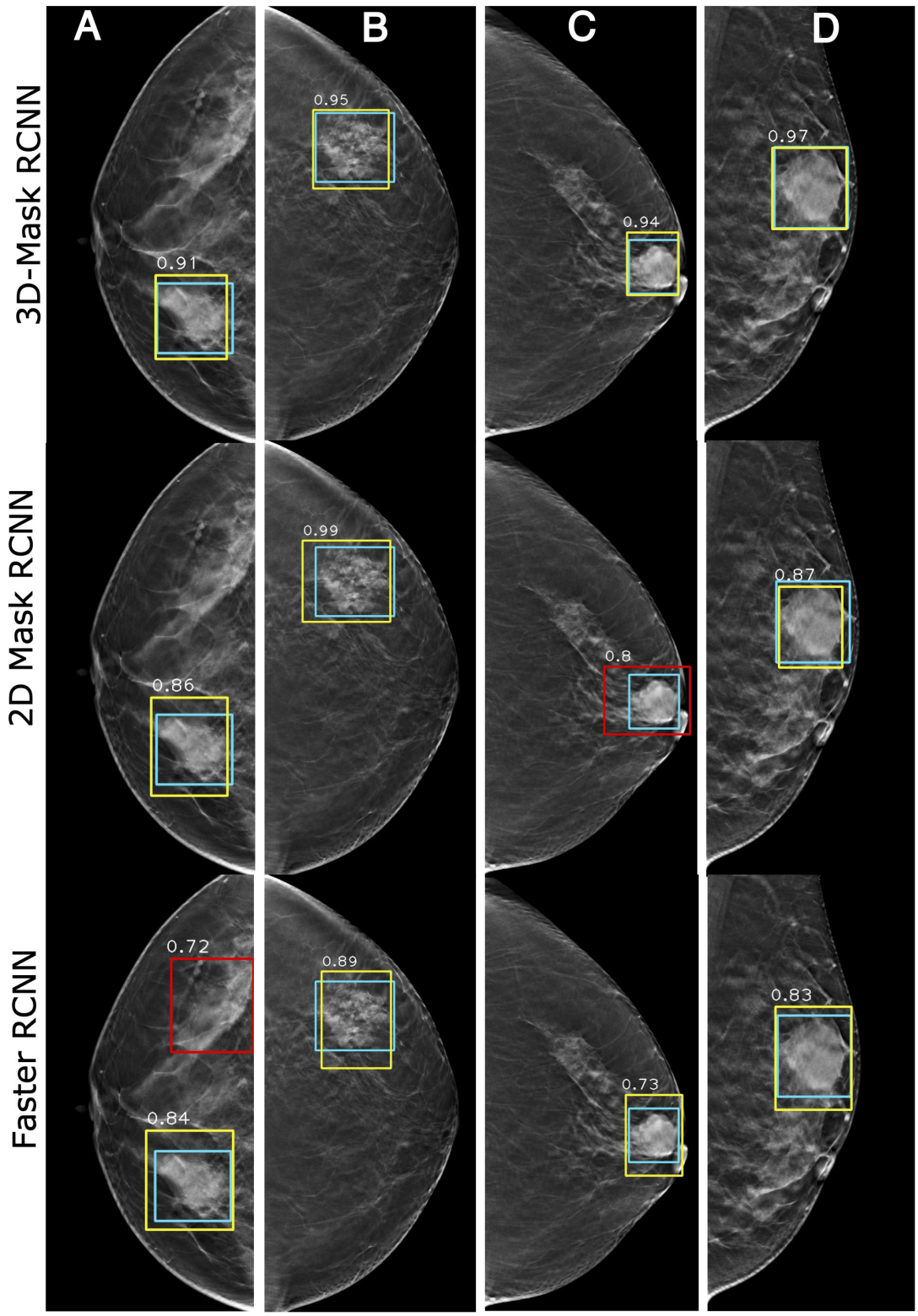


FIGURE 7 | Examples of mass detection results using the 3D-Mask RCNN, 2D-Mask RCNN and Faster RCNN on patients with different densities. Four patients are included (from left to right) with the following characteristics: **(A)** aged 55 years, low density breast, malignant tumor, spiculate mass, and maximum tumor diameter of 43.24 mm; **(B)** aged 54 years, low density breast, malignant tumor, irregular mass, and maximum tumor diameter of 28.36 mm; **(C)** aged 69 years, low density breast, malignant tumor, oval mass, and maximum tumor diameter of 28.36 mm; and **(D)** aged 45 years, dense breast, malignant tumor, oval mass, and maximum diameter of 41.42 mm (green: ground truth box, yellow and red: detection box).

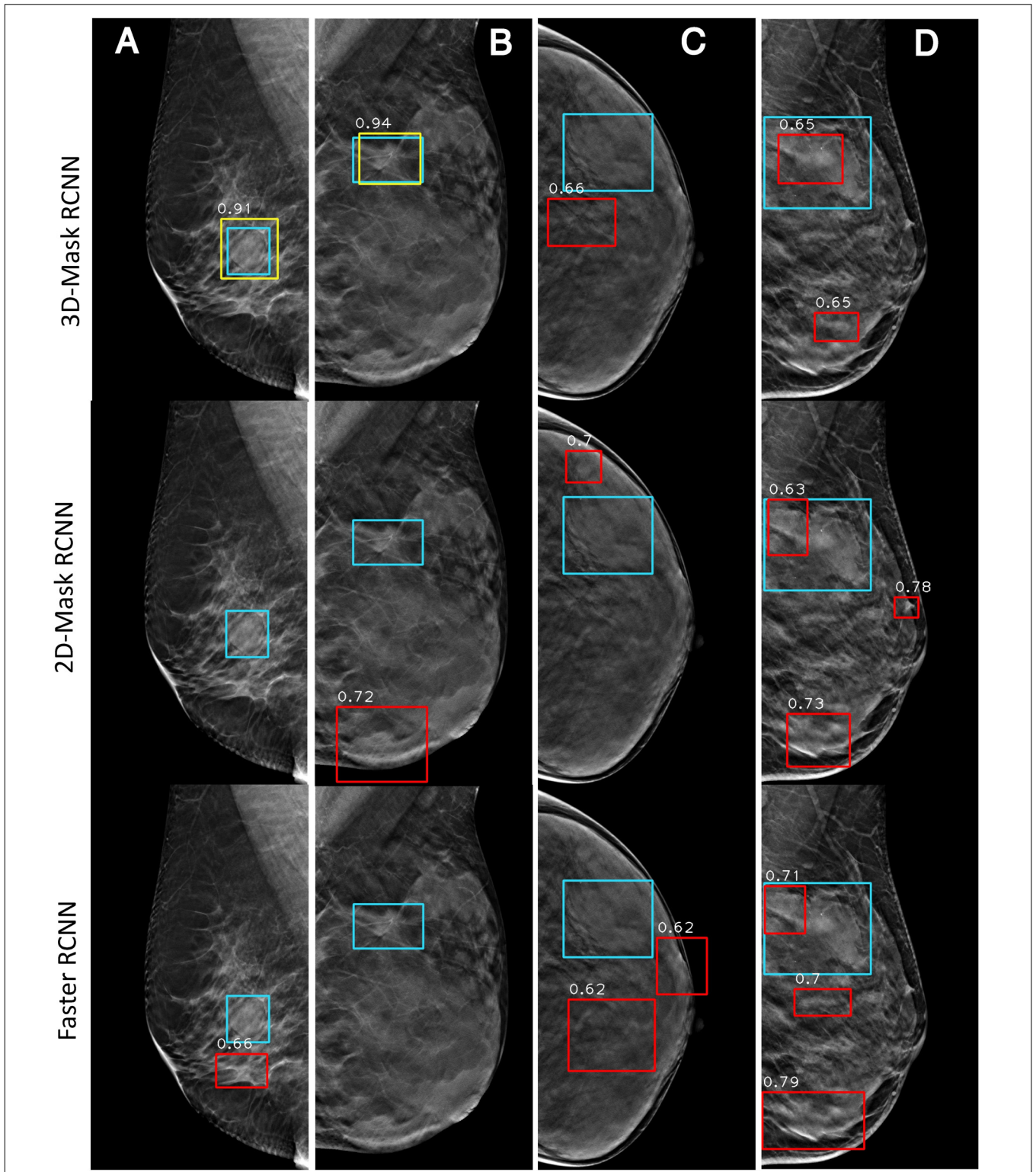


FIGURE 8 | Examples of mass detection results for the 3D-Mask RCNN, 2D-Mask RCNN and Faster RCNN in patients with dense breast. Four patients are included (from left to right) with the following characteristics: **(A)** aged 60 years, dense breast, benign tumor, oval mass, and maximum tumor diameter of 24.62 mm; **(B)** aged 30 years, dense breast, malignant tumor, spiculate mass, and maximum tumor diameter of 38.35 mm; **(C)** aged 43 years, dense breast, malignant tumor, irregular mass, and maximum tumor diameter of 46 mm; and **(D)** aged 51 years, dense breast, malignant tumor, irregular mass, and maximum tumor diameter of 59.82 mm (green: ground truth box, yellow and red: detection box).

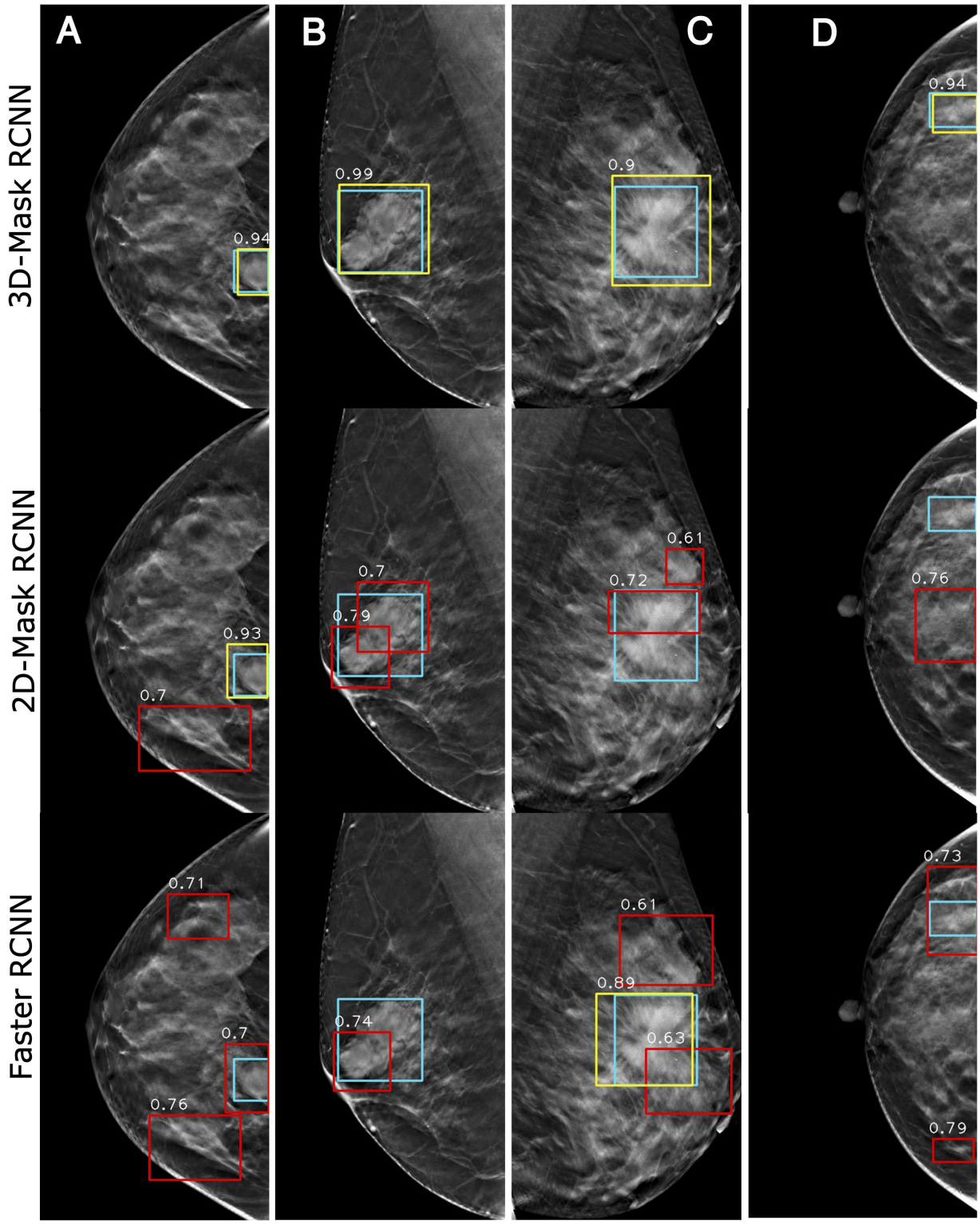


FIGURE 9 | Examples of mass detection results for the 3D-Mask RCNN, 2D-Mask RCNN and Faster RCNN in patients with dense breast. Four patients are included (from left to right) with the following characteristics: **(A)** aged 33 years, dense breast, benign tumor, oval mass, and maximum tumor diameter of 24.24 mm. **(B)** aged 52 years, dense breast, benign tumor, spiculate lesion, and maximum tumor diameter of 50.35 mm; **(C)** aged 48 years, dense breast, malignant tumor, spiculate lesion, and maximum tumor diameter of 62.72 mm; and **(D)** aged 40 years, dense breast, benign tumor, irregular mass, and maximum tumor diameter of 26.58 mm (green: ground truth box, yellow and red: detection box).

TABLE 4 | Tumor segmentation results of the 3D-Mask RCNN, 2D-Mask RCNN and SFCM-based CAD method.

Characteristics	3D-Mask RCNN		2D-Mask RCNN		SFCM	
	AP	FNR	AP	FNR	AP	FNR
All	**0.934**	**0.053**	0.730	0.260	0.674	0.317
Age						
<40	**0.934**	**0.055**	0.734	0.255	0.684	0.307
40–49	**0.934**	**0.052**	0.731	0.259	0.671	0.320
50–59	**0.929**	**0.057**	0.756	0.233	0.676	0.315
60–69	**0.921**	**0.065**	0.743	0.246	0.689	0.303
>70	**0.950**	**0.036**	0.717	0.272	0.623	0.367
Histological type						
Benign	**0.937**	**0.049**	0.723	0.267	0.654	0.338
Malignant	**0.932**	**0.055**	0.735	0.254	0.658	0.332
Mass type						
Round/oval	**0.933**	**0.054**	0.722	0.267	0.670	0.321
Spiculate mass	**0.939**	**0.049**	0.717	0.271	0.687	0.303
Irregular	**0.931**	**0.055**	0.727	0.263	0.686	0.306
Breast density						
Low[1]	**0.931**	**0.055**	0.743	0.246	0.668	0.323
High[2]	**0.933**	**0.054**	0.727	0.262	0.669	0.322
Maximum diameter (mm)						
$10 \leq d < 30$	**0.931**	**0.056**	0.734	0.255	0.668	0.323
$30 \leq d < 50$	**0.930**	**0.056**	0.730	0.259	0.674	0.317
$d \geq 50$	**0.935**	**0.052**	0.726	0.263	0.674	0.317

AP, Average Precision; FNR, False Negative Rate. The highest AP or lowest FNR values for the mass detection methods are in bold.

Assessment and Comparison of Mass Segmentation Methods on DBT

Lesion segmentation was performed using the 3D-Mask RCNN-, 2D-Mask RCNN- and SFCM-based clustering methods. **Table 4** illustrates the tumor segmentation results for all the samples in the testing set and, in the subgroups according to the age range, histological type, mass type, breast density and lesion size. From the table, the 3D segmentation method clearly achieved superior performance compared with the 2D-Mask RCNN- and SFCM-based methods with higher APs and lower FNRs.

Running Time Evaluation and Comparison

The training and testing of the deep CAD model were performed on a Linux workstation with 16 CPU cores (2.1 GHz) and 6 NVIDIA 1080Ti GPUs with 11 GB of memory. The execution time for the 3D-Mask RCNN, the 2D-Mask RCNN and the Faster RCNN were approximately 350, 260, and 245 h, respectively. The detailed description of the running times was illustrated in **Table 5**.

TABLE 5 | Running time comparison of the deep mass detection networks.

Name	Size (MB)	Parameters	Time per image (ms)
2D-Mask RCNN	244	4.93e + 07	195
3D-Mask RCNN	320	5.27e + 07	100
Faster RCNN	533	1.41e + 08	210

DISCUSSION

The Mask RCNN framework was developed to detect masses in breasts. It has been shown that the 3D-Mask RCNN is superior to the other two deep learning-based methods, namely, the 2D-Mask RCNN and the Faster RCNN. Moreover, we have assessed the detection performance in various subgroups of patients with different age ranges, breast densities, histological types and tumor shapes. The results suggested that the 3D-Mask RCNN achieved significantly better performances than the 2D deep CNN models on specific groups according to clinicopathological features.

A previous study used the Faster RCNN model with VGG16 on the INBreast dataset to detect malignant masses and calcifications (Ribli et al., 2018). A deep CNN with multiple instance learning (Yousefi et al., 2018) achieved better performances than the handcrafted features-based CAD systems on 2D slices of DBT images. Samala et al. (2016) presented a DCNN-based approach for mass detection using DBT images. A recent study used the deep learning (Samala et al., 2019) method with transfer learning to discriminate between the malignant and benign masses in DBT images. However, these studies were performed based on the 2D analysis of a deep neural network. In this study, we showed that the 3D deep learning method is superior to the 2D methods in both mass detection and segmentation. It is interesting to note that the Mask RCNN has better lesion-based detection performance while the Faster RCNN achieved better breast-based mass detection in DBT images.

The systematic analyses of the CAD systems showed that the mass detection performances are correlated with patient characteristics, such as age, histological type, mass type, breast density, and mass size. The breast masses of patients who are 40–59 years old are more difficult to detect. Moreover, CAD detection was less accurate with more FPs for the samples with benign tumors, irregular shapes, and dense breasts. Smaller ($10 \leq d < 30$) and larger ($d \geq 50$) tumors were difficult for the deep learning-based detection methods to detect (**Table 3**). We observed that the 3D-Mask RCNN has significantly better ($p < 0.05$) mass detection performance than the other 2D methods, especially for specific groups that were more difficult to detect (i.e., those aged 40–59, benign tumors, irregular tumors and dense breasts). This may be explained by the fact that DBT reduces the tissue overlap and increases the lesion conspicuity, particularly in dense breasts, which makes it rational that 3D methods have better detection performance than those of the 2D methods. Moreover, 3D-Mask RCNN method take advantage of volumetric information of DBT, which is better than 2D methods in discriminating the masses of irregular shapes from normal tissues (e.g., fibroglandular) with fine textures/structures, especially in dense breasts. Since age is inversely associated with breast density (Checka et al., 2012), it is reasonable that detection performance is higher in older patients (aged 60 years and older) than that in the others. The 3D-Mask RCNN also achieved better segmentation performance than those

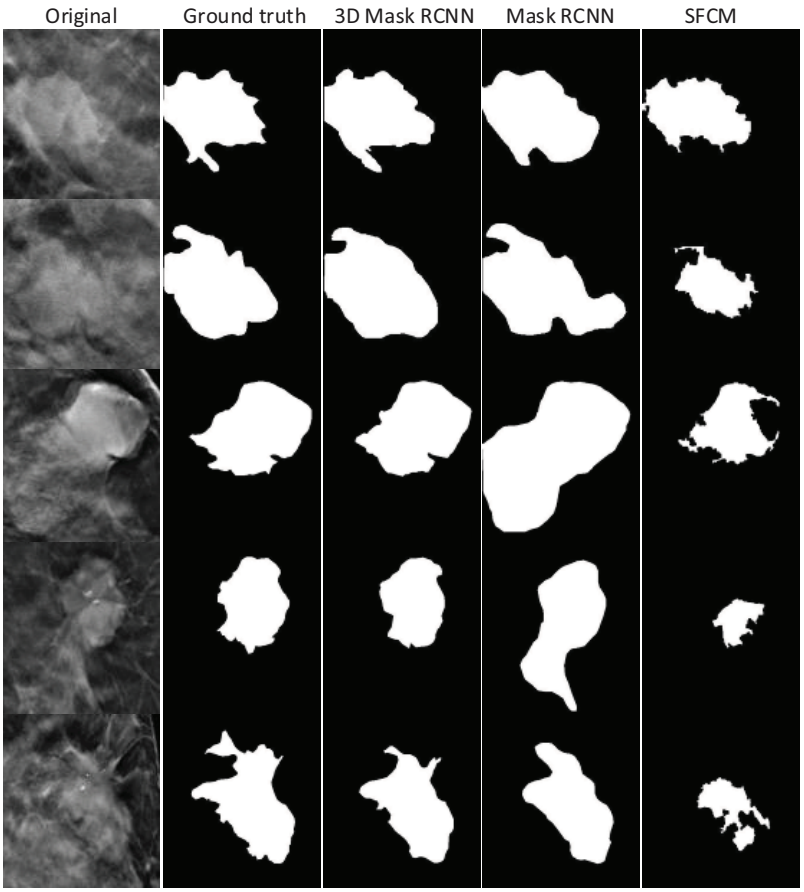


FIGURE 10 | Tumor segmentation results for the 3D-Mask RCNN, 2D-Mask RCNN, and SFCM. Five patients are included (from top to bottom).

of the 2D methods in the entire testing set and the subgroups (**Table 4**). The case study also illustrates that the 3D-Mask RCNN had fewer false positives than the other methods (**Figure 10**).

The limitations of this study should be addressed. First, the sample size in this study is relatively small, especially when the subgroup analysis was conducted on patients with different clinicopathological characteristics. A data cohort with a larger size should be used in the future to refine the results of our study. Second, we do not perform transfer learning in this study. A future study with transfer learning from mammograms might enhance the accuracy of the mass detection/segmentation in patients. Third, we used image patches for detection to save computer memory, and thus, future studies that focus on the entire image should be conducted.

CONCLUSION

In summary, we proposed a 3D-Mask RCNN-based mass detection and segmentation framework for detecting and segmenting tumor masses. A comparison of the 3D- and 2D-based methods under different subgroups based on age ranges, lesion sizes, lesion shapes, and breast densities was conducted. We illustrated that the 3D-Mask RCNN has better performance than the 2D methods, especially for subgroups with specific clinicopathologic characteristics that show higher FPs, and the improvement is significant.

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 studies involving human participants were reviewed and approved by Shanghai cancer hospital, Fudan University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MF, WP, and LL designed the study. HZ, SZ, and CY performed the image processing, deep learning methods, and statistical

analyses. MF and LL wrote the manuscript. YG and XG analyzed the data. All authors reviewed the manuscript.

FUNDING

This work was supported in part by grants from the National Key R&D program of China (2017YFC0109402), the National Natural Science Foundation of China (61731008, 61871428, and U1809205), the Natural Science Foundation of Zhejiang Province of China (LJ19H180001), and King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No URF/1/1976-04 and URF/1/3007-01.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.599333/full#supplementary-material

REFERENCES

Bahl, M., Mercaldo, S., Vijapura, C. A., McCarthy, A. M., and Lehman, C. D. (2019). Comparison of performance metrics with digital 2D versus tomosynthesis mammography in the diagnostic setting. *Eur. Radiol.* 29, 477–484. doi: 10.1007/s00330-018-5596-7

Bandos, A., Rockette, H. T., and Gur, D. (2010). Area under the free-response ROC curve (FROC) and a related summary index. *Biometrics* 65, 247–256. doi: 10.2307/25502264

Bandos, A. I., Rockette, H. E., Song, T., and Gur, D. (2009). Area under the free-response ROC curve (FROC) and a related summary index. *Biometrics* 65, 247–256. doi: 10.1111/j.1541-0420.2008.01049.x

Berber, T., Alpkocak, A., Balci, P., and Dicle, O. (2013). Breast mass contour segmentation algorithm in digital mammograms. *Comput. Methods Progr. Biomed.* 110, 150–159. doi: 10.1016/j.cmpb.2012.11.003

Bornefalk, H., and Hermansson, A. B. (2005). On the comparison of FROC curves in mammography CAD systems. *Med. Phys.* 32, 412–417. doi: 10.1118/1.1844433

Carney, P. A., Miglioretti, D. L., Yankaskas, B. C., Kerlikowske, K., Rosenberg, R., Rutter, C. M., et al. (2003). Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann. Intern. Med.* 138, 168–175. doi: 10.7326/0003-4819-138-3-200302040-00008

Checka, C. M., Chun, J. E., Schnabel, F. R., Lee, J., and Toth, H. (2012). The relationship of mammographic density and age: implications for breast cancer screening. *Am. J. Roentgenol.* 198, W292–W295. doi: 10.2214/AJR.10.6049

Chong, A., Weinstein, S. P., McDonald, E. S., and Conant, E. F. (2019). Digital breast tomosynthesis: concepts and clinical practice. *Radiology* 292, 1–14. doi: 10.1148/radiol.2019180760

Conant, E. F., Barlow, W. E., Herschorn, S. D., Weaver, D. L., Beaber, E. F., Tosteson, A. N. A., et al. (2019). Association of digital breast tomosynthesis vs digital mammography with cancer detection and recall rates by age and breast density. *JAMA Oncol.* 5, 635–642. doi: 10.1001/jamaoncol.2018.7078

Fan, M., Li, Y., Zheng, S., Peng, W., Tang, W., and Li, L. (2019). Computer-aided detection of mass in digital breast tomosynthesis using a faster region-based convolutional neural network. *Methods* 166, 103–111. doi: 10.1016/j.ymeth.2019.02.010

Fontaine, M., Tourasse, C., Pages, E., Laurent, N., Laffargue, G., Millet, I., et al. (2019). Local tumor staging of breast cancer: digital mammography versus digital mammography plus tomosynthesis. *Radiology* 291, 594–603. doi: 10.1148/radiol.2019182457

Geras, K. J., Mann, R. M., and Moy, L. (2019). Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives. *Radiology* 293, 246–259. doi: 10.1148/radiol.2019182627

Haas, B. M., Kalra, V., Geisel, J., Raghu, M., Durand, M., and Philpotts, L. E. (2013). Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology* 269, 694–700. doi: 10.1148/radiol.13130307

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2020). Mask R-CNN. *IEEE Trans. Patt. Anal. Mach. Intellig.* 42, 386–397. doi: 10.1109/TPAMI.2018.2844175

He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 1026–1034.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, 770–778.

Hofvind, S., Hovda, T., Holen, A. S., Lee, C. I., Albertsen, J., Bjorndal, H., et al. (2018). Digital breast tomosynthesis and synthetic 2D mammography versus digital mammography: evaluation in a population-based screening program. *Radiology* 287, 787–794. doi: 10.1148/radiol.2018171361

Kingma, D., and Ba, J. (2014). "Adam: a method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations*, Vancouver, BC.

Kus, P., and Karagoz, I. (2012). Fully automated gradient based breast boundary detection for digitized X-ray mammograms. *Comput. Biol. Med.* 42, 75–82. doi: 10.1016/j.compbiomed.2011.10.011

Lai, X., Yang, W., and Li, R. (2020). DBT masses automatic segmentation using U-net neural networks. *Comput. Math. Methods Med.* 2020:7156165. doi: 10.1155/2020/7156165

Li, X., Qin, G. G., He, Q., Sun, L., Zeng, H., He, Z. L., et al. (2019). Digital breast tomosynthesis versus digital mammography: integration of image modalities enhances deep learning-based breast mass classification. *Eur. Radiol.* 30, 778–788. doi: 10.1007/s00330-019-06457-5

Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 936–944.

Michell, M. J., Iqbal, A., Wasan, R. K., Evans, D. R., Peacock, C., Lawinski, C. P., et al. (2012). A comparison of the accuracy of film-screen mammography, full-field digital mammography, and digital breast tomosynthesis. *Clin. Radiol.* 67, 976–981. doi: 10.1016/j.crad.2012.03.009

Nystrom, L., Andersson, I., Bjurstam, N., Frisell, J., Nordenskjold, B., and Rutqvist, L. E. (2002). Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* 359, 909–919. doi: 10.1016/S0140-6736(02)08020-0

Osteras, B. H., Martinsen, A. C. T., Gullien, R., and Skaane, P. (2019). Digital mammography versus breast tomosynthesis: impact of breast density on diagnostic performance in population-based screening. *Radiology* 293, 60–68. doi: 10.1148/radiol.2019190425

Pohlmann, S. T. L., Lim, Y. Y., Harkness, E., Pritchard, S., Taylor, C. J., and Astley, S. M. (2017). Three-dimensional segmentation of breast masses from digital breast tomosynthesis images. *J. Med. Imaging* 4:034007. doi: 10.1117/1.JMI.4.3.034007

Reiser, I., Nishikawa, R. M., Giger, M. L., Wu, T., Rafferty, E. A., Moore, R., et al. (2006). Computerized mass detection for digital breast tomosynthesis directly from the projection images. *Med. Phys.* 33, 482–491. doi: 10.1118/1.2163390

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Ribli, D., Horvath, A., Unger, Z., Pollner, P., and Csabai, I. (2018). Detecting and classifying lesions in mammograms with deep learning. *Sci. Rep.* 8:4165. doi: 10.1038/s41598-018-22437-z

Samala, R. K., Chan, H. P., Hadjiiski, L., Helvie, M. A., Wei, J., and Cha, K. (2016). Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography. *Med. Phys.* 43:6654. doi: 10.1118/1.4967345

Samala, R. K., Chan, H. P., Hadjiiski, L. M., Helvie, M. A., Richter, C., and Cha, K. (2018). Evolutionary pruning of transfer learned deep convolutional neural network for breast cancer diagnosis in digital breast tomosynthesis. *Phys. Med. Biol.* 63:e095005. doi: 10.1088/1361-6560/Aabb5b

Samala, R. K., Heang-Ping, C., Hadjiiski, L., Helvie, M. A., Richter, C. D., and Cha, K. H. (2019). Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets. *IEEE Trans. Med. Imaging* 38, 686–696. doi: 10.1109/TMI.2018.2870343

Skaane, P., Bandos, A. L., Niklason, L. T., Sebuodegard, S., Osteras, B. H., Gullien, R., et al. (2019). Digital mammography versus digital mammography plus tomosynthesis in breast cancer screening: the oslo tomosynthesis screening trial. *Radiology* 291, 22–29. doi: 10.1148/radiol.2019182394

Tagliafico, A. S., Calabrese, M., Bignotti, B., Signori, A., Fisci, E., Rossi, F., et al. (2017). Accuracy and reading time for six strategies using digital breast tomosynthesis in women with mammographically negative dense breasts. *Eur. Radiol.* 27, 5179–5184. doi: 10.1007/s00330-017-4918-5

Varela, C., Timp, S., and Karssemeijer, N. (2006). Use of border information in the classification of mammographic masses. *Phys. Med. Biol.* 51, 425–441. doi: 10.1088/0031-9155/51/2/016

Vourtsis, A., and Berg, W. A. (2019). Breast density implications and supplemental screening. *Eur. Radiol.* 29, 1762–1777. doi: 10.1007/s00330-018-5668-8

Wu, Y. T., Zhou, C., Chan, H. P., Paramagul, C., Hadjiiski, L. M., Daly, C. P., et al. (2010). Dynamic multiple thresholding breast boundary detection algorithm for mammograms. *Med. Phys.* 37, 391–401. doi: 10.1118/1.3273062

Yousefi, M., Krzyzak, A., and Suen, C. Y. (2018). Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning. *Comput. Biol. Med.* 96, 283–293. doi: 10.1016/j.compbiomed.2018.04.004

Zhang, C., and Li, L. (2014). *3D Segmentation of Masses in DCE-MRI Images Using FCM and Adaptive MRF*. San Diego, CA: SPIE.

Zhang, Y., Chan, H. P., Sahiner, B., Wei, J., Goodsitt, M. M., Hadjiiski, L. M., et al. (2006). A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis. *Med. Phys.* 33, 3781–3795. doi: 10.1118/1.2237543

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 Physiology

ORIGINAL RESEARCH
published: 27 November 2020
doi: 10.3389/fphys.2020.604764

Multi-Head Self-Attention Model for Classification of Temporal Lobe Epilepsy Subtypes

Peipei Gu[1], Ting Wu[2], Mingyang Zou[3†], Yijie Pan[4,5*†], Jiayang Guo[6]*, Jianbing Xiahou[7]*, Xueping Peng[8], Hailong Li[9], Junxia Ma[1] and Ling Zhang[3]**

[1] Software Engineering College, Zhengzhou University of Light Industry, Zhengzhou, China, [2] Department of Magnetoencephalography, Nanjing Brain Hospital Affiliated, Nanjing Medical University, Nanjing, China, [3] School of Biomedical Engineering, Hubei University of Science and Technology, Xianning, China, [4] Department of Computer Science and Technology, Tsinghua University, Beijing, China, [5] Ningbo Institute of Information Technology Application, Chinese Academy of Sciences, Ningbo, China, [6] Department of Electrical Engineering and Computer Science, University of Cincinnati, Cincinnati, OH, United States, [7] School of Informatics, Xiamen University, Xiamen, China, [8] Australian Artificial Intelligence Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW, Australia, [9] The Perinatal Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

As a long-standing chronic disease, Temporal Lobe Epilepsy (TLE), resulting from abnormal discharges of neurons and characterized by recurrent episodic central nervous system dysfunctions, has affected more than 70% of drug-resistant epilepsy patients across the world. As the etiology and clinical symptoms are complicated, differential diagnosis of TLE mainly relies on experienced clinicians, and specific diagnostic biomarkers remain unclear. Though great effort has been made regarding the genetics, pathology, and neuroimaging of TLE, an accurate and effective diagnosis of TLE, especially the TLE subtypes, remains an open problem. It is of a great importance to explore the brain network of TLE, since it can provide the basis for diagnoses and treatments of TLE. To this end, in this paper, we proposed a multi-head self-attention model (MSAM). By integrating the self-attention mechanism and multilayer perceptron method, the MSAM offers a promising tool to enhance the classification of TLE subtypes. In comparison with other approaches, including convolutional neural network (CNN), support vector machine (SVM), and random forest (RF), experimental results on our collected MEG dataset show that the MSAM achieves a supreme performance of 83.6% on accuracy, 90.9% on recall, 90.7% on precision, and 83.4% on F1-score, which outperforms its counterparts. Furthermore, effectiveness of varying head numbers of multi-head self-attention is assessed, which helps select the optimal number of multi-head. The self-attention aspect learns the weights of different signal locations which can effectively improve classification accuracy. In addition, the robustness of MSAM is extensively assessed with various ablation tests, which demonstrates the effectiveness and generalizability of the proposed approach.

Keywords: TLE diagnosis, self-attention model, epilepsy classification, temporal lobe epilepsy detection, multi-head self-attention

OPEN ACCESS

Edited by:
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Wei Chen,
Chinese Academy of Agricultural Sciences, China
Chen Yang,
Tsinghua University, China
Chongyang Shi,
Beijing Institute of Technology, China

***Correspondence:**
Ting Wu
lvwu123@163.com
Ling Zhang
lingzhang1212@hbust.edu.cn
Yijie Pan
pyj20@mails.tsinghua.edu.cn
Jiayang Guo
guojy@mail.uc.edu
Jianbing Xiahou
jbxiahou@xmu.edu.cn

† These authors have contributed equally to this work

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

Received: *10 September 2020*
Accepted: *22 October 2020*
Published: *27 November 2020*

Citation:
Gu P, Wu T, Zou M, Pan Y, Guo J, Xiahou J, Peng X, Li H, Ma J and Zhang L (2020) Multi-Head Self-Attention Model for Classification of Temporal Lobe Epilepsy Subtypes. Front. Physiol. 11:604764. doi: 10.3389/fphys.2020.604764

INTRODUCTION

Epilepsy, a chronic central nervous system disease, is typically caused by the repeated abnormal discharge of neurons and is characterized by symptoms that are sudden, periodic, and short-term. According to a recent survey (Beghi and Giussani, 2018), around 70 million people across the world are affected, of which, 90% are grouped in undeveloped areas. Epilepsy has a global occurrence rate of 6.38∼7.60% (Jette et al., 2017) and around 67.77 new cases for every hundred thousand people are found each year. The ones who fail to take control of epilepsy after drug treatments are known as drug-resistant epilepsy (DRE) (Wiebe, 2013), among which, temporal lobe epilepsy (TLE), a common type of epilepsy widely existing in young and elderly patients, accounts for around 70% (Mariani et al., 2019). With the growth of the population and advent of an aging society, it is inevitable that TLE will be of a great burden for human beings. Therefore, it is urgent to identify the subtype, cause, and inducement in the treatment of TLE. Though progress has been made through subjective analysis, traditional methods for imaging and clinical symptom assessment heavily rely on human experts, leading to a long diagnostic time. Moreover, subjective diagnostic results are often made from different experts, even for the same patient (Siuly and Li, 2015). Thus, it is hard to make medical decision using solely experts. Therefore, it is crucial to develop an efficient and objective TLE diagnosis method to support the treatment of TLE.

Typically, TLE can be categorized into three subtypes: simple partial seizure (SPS), complex partial seizure (), and these two types coexisting. The key difference between SPS and CPS lies in the disturbance of consciousness. In comparison with SPS, CPS is more likely to evolve into drug-resistant epilepsy, which denotes the ineffectiveness of drug treatments. On the contrary, taking antiepileptic drugs may result in side effects that affect the cognition and puberty development of the human brain, adding great emotional and economic burden to the patients and their families. An accurate diagnosis in the early stage of disease outbreak is fundamental to non-drug treatments, which avoids the dosage of drugs and further ensures a good quality of life for the patients. To this end, the medical community has put great effort into exploring the difference between the CPS and SPS brain networks and studying the treatments for different subtypes of brain network nodes.

Learning to classify these two subtypes accurately and objectively will benefit the clinical risk stratification and relieve the heavy dependency on human experts. In addition, by predicting the subtypes and taking active measures in advance, it also keeps the high-risk population with a conscious disorder from the risk of sudden death that results from the disturbance of consciousness after epileptic seizures.

In this study, we used collected magnetoencephalography (MEG) signals to classify the subtypes of temporal lobe epilepsy, as MEG has emerged as an non-invasive, reliable, fast, and easy-to-use technique to record functional activities of the brain (Englot et al., 2016; van Klink and Zijlmans, 2019; Shi et al., 2020). It has been observed that the spike-wave of epilepsy is indeed a time-dependent characteristic wave. MEG shows its great superiorities in acquiring the high-temporal resolution of data and spatial lateralization and localization (Liu et al., 2020). Therefore, compared with other tools, such as electroencephalogram (EEG) (Anastasiadou et al., 2019; Yao et al., 2019; Serna et al., 2020), MEG has been considered as an effective tool to diagnose epilepsy and find the location of cortical pathological activity or damage of epileptic foci (Burns et al., 2014).

Recent years have witnessed the emerging performance of deep learning in various research topics, including the study of epilepsy (Peng et al., 2019, 2020a,b; Niu et al., 2020). For example, Wu et al., 2018 proposed to deal with TLE lateralization based on MEG by transferring it into a series of binary classification problems. To that effect, the resting-state brain network is first employed to extract features of each participant, upon which the support vector machine (SVM) is built to achieve the classification of extracted features. Achilles et al., 2018 developed a non-invasive automatic system for monitoring epilepsies via resorting to the convolutional neural network (CNN) to deduce feature representations to distinctively detect seizures from the videos. The experimental results from different epileptic seizure types show a supreme performance of up to 78.33% AUC value, which demonstrates the promising prospect to utilize deep learning as a tool for curing epilepsy.

Fang et al., 2015 explored the anatomical connectivity differences underlying functional variance. Based on the constructed anatomical networks, multivariate pattern analysis is applied to extract the anatomical connectivity differences between the left and right TLE patients. Cantor-Rivera et al., 2015 derived an accuracy rate of more than 82% by using clinical parameters and extracting features of MRI and DTI images to identify TLE. Though a great TLE diagnosis rate of more than 80% has been made in most studies, a significant amount of misdiagnosis remains (around 10–20%). Many normal people are often mistakenly identified with a correct diagnoses on TLE disease. Moreover, these samples are in a small range and it is unclear whether they can be directly applied to other hospitals or not. To solve it, current state-of-the-art methods consider utilizing the toolbox of machine learning (deep learning) to achieve a high classification of epilepsy patients and normal persons (Zafar et al., 2017; Ahmedt-Aristizabal et al., 2018; Guo et al., 2020), and analyzing the changes of functional connectivity between enhanced and weakened brain regions (Rajpoot et al., 2015). Most experimental data of these methods are collected from EEG, fMRI, EEG fMRI, etc. (Pedreira et al., 2014; Sarraf and Tofighi, 2016; Vergun et al., 2016). While the combinations with neural networks further reduces the possibility of misdiagnoses, the limitation remains unsolved: Though off-the-shelf approaches can identify epilepsy patients and normal subjects, they fail to tell the specific epilepsy subtypes.

In this paper, we investigate the classification of TLE subtypes by integrating the self-attention mechanism and multilayer perceptron based method on our collected MEG dataset, aiming to find out the functional connection and pathogenesis of the brain network related to the seizure of these two subtypes.

This research is important to propose a more rapid, more accurate, and intelligent subtype recognition method. To that effect, our method, termed MSAM, builds a multi-head self-attention model to predict epileptic seizures, where the original MEG signal is fed as its input. The self-attention mechanism analyzes the influence of the position of the sampled signal, so as to set different weights for the classification algorithm. The pre-seizure and interictal periods are separated, and then the multilayer perceptron model is used to extract the information of frequency and time domain to realize the feature extraction and classification. We propose to construct a multi-head self-attention model and apply it to the temporal lobe epilepsy subtype recognition algorithm.

To summarize, our main contributions are:

- to investigate the characteristics of TLE subtypes;
- to propose an end-to-end multi-head self-attention model, called MSAM, that predicts TLE subtypes;
- to evaluate the proposed model on a real-world dataset with classification task, demonstrating that the MSAM is superior to all comparative methods.

MATERIALS AND METHODS

In this section, we introduce our multi-head self-attention model to classify subtypes of epilepsy since the classification of epilepsy is more important to epilepsy physicians than the position of epilepsy. On one hand, the same detected discharge location may cause different symptoms for different patients. On the other hand, even though the clinical symptoms are the same, the positioning results may be completely different. Thus, the clinical symptoms, locations, and subtypes of epilepsy patients are very complicated. To solve this, in this paper, we propose to make full use of the self-attention mechanism to distinguish different symptoms in the same location. Meanwhile, we further adopt multilayer perceptron to solve the obstacle that the same clinical symptom possesses varying positions. **Figure 1** displays the framework of our self-attention mechanism based deep learning network for epilepsy recognition.

We consider a set of N training dataset $D = \{(t_i{:}l_i), i = 1,\ldots,N\}$ where t_i is the $i - th$ sample and l_i is its corresponding label. In our settings, $l_i = 1$ if the $i - th$ sample is detected, and 0 if otherwise. Our network takes the $i - th$ sample as its input and forward it to predict the label l_i As can be seen from **Figure 1**, our framework consists of four components: data preprocessing level, which processes MEG data; feature extraction, multilayer perceptron layer, which is a feed forward neural network; multi-head self-attention layer, which analyzes the weights of locations; and the last layer, which classifies and detects TLE subtypes.

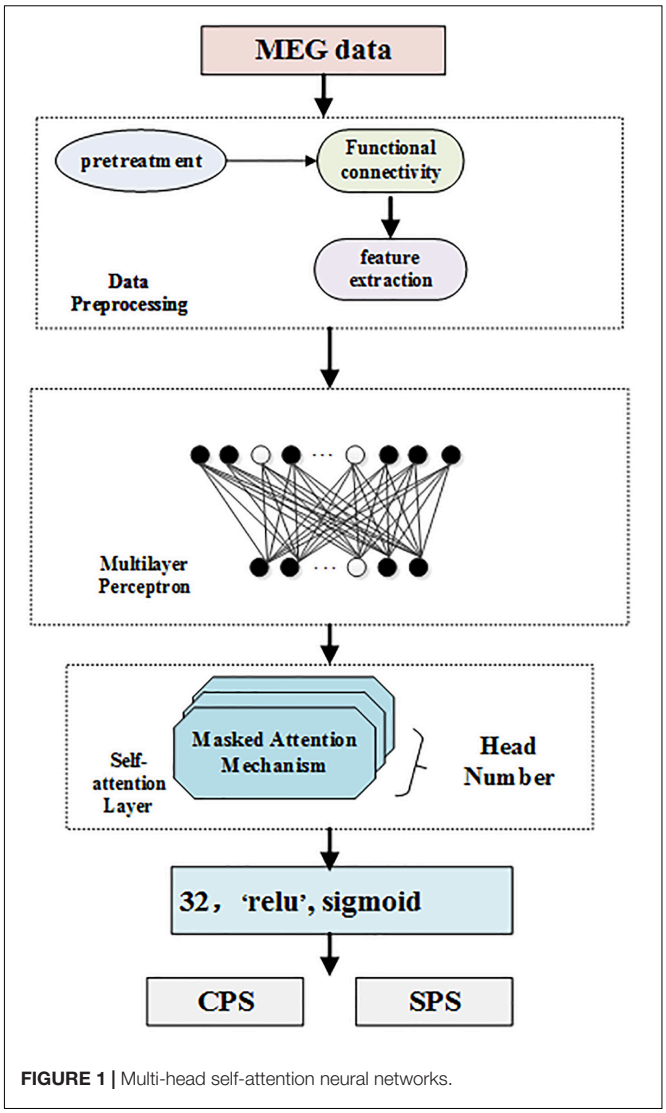


FIGURE 1 | Multi-head self-attention neural networks.

Multilayer Perceptron Layer

Multilayer perceptron (MLP) is a feed forward neural network model. MLP contains one dropout layer and four dense layers. The MLP module is shown in **Figure 1**. Each layer of the network is composed of multiple nodes. Except for the nodes in the output layer, each node is connected with all nodes in the next layer.

Dropout technology ensures that in the every iteration of the process for neural network training, dropout technology will randomly stop a certain number of neurons in the hidden layer, and use the mask process to set the output of these neurons in the hidden layer to 0, while the connection weights of the non-working neurons will not be updated in this iteration process. When the trained model is used in the test set, all nodes need to be used, and the neurons in the stopped-working state will return to work. Dropout technology effectively enhances the generalization ability of the deep neural network model and plays an important role in preventing over fitting of the deep learning model.

In **Figure 1**, the white nodes represent the neurons that will stop working according to a certain probability. After the dropout layer, there are four fully connected layers. The number of neurons in each hidden layer is 1024, 512, 128, and 32, respectively, corresponding to the activation functions of ELU, tanh, tanh, and relu.

Self-Attention Layer

Self-attention is an attention mechanism relating different positions of a single input sequence to compute a representation of the same sequence. In order to obtain these representations, every input is multiplied with a set of weights for Keys (denoted as K), a set of weights for queries (denoted as Q), and a set of weights for values (denoted as V). Then, the self-attention learns a function mapping query Q to a series of key-value pairs (K, V), as follows:

$$AttentionV = QK^T V \tag{1}$$

Attention essentially assigns a weight coefficient to each element in the sequence, which can also be understood as soft addressing. If each element is stored, the attention can calculate the similarity between Q and K. The similarity calculated by Q and K reflects the importance of the extracted V value, that is, the weight, and then the weighted summation obtains the attention value. The special point of the Self Attention mechanism in the K, Q, V model is that $Q = \mathrm{K} = \mathrm{V}$:

$$\text{Attention } (Q,\ \mathrm{K},\ \mathrm{V}) = \text{softmax}\left(\frac{\mathrm{QK^T}}{\sqrt{\mathrm{d_K}}}\right)\mathrm{V}. \tag{2}$$

The multi-head attention mechanism obtains h (i.e., one per head) different representations of (Q, K, V), computes self-attention for each representation, and concatenates the results. This can be expressed in the same notation as Eq. (4):

$$\text{head}_i = \text{Attention}\left(QW_i^Q, KW_i^k, VW_i^v\right) \tag{3}$$

$$\text{MultiHead}\,(Q, K, V) = \text{Contact}\left(head_1 \ldots., head_h\right) W^0 \tag{4}$$

where the projections are parameter matrices $W_i^Q \varepsilon \mathbb{R}^{d\times d_k}$, $W_i^k \varepsilon\ \mathbb{R}^{d\times d_k}$, $W_i^v \varepsilon\ \mathbb{R}^{d\times d_v}$, and $W^o \varepsilon\ \mathbb{R}^{hd_v\times d}$, $d_k = d_v = {}^{d}/_{h}$.

Classification and Detection Layer

To correctly predict if a sample is detected, we further deploy a softmax layer on top of the neural network. The basic process of the softmax layer is to map the output representation of the encoding layer into a probability interval (0, 1). In this paper, we regard the detection as a binary classification problem. Then, we forward input samples to the encoding network, outputs of which are further mapped into the probability interval (0, 1) by the softmax layer as below:

$$l_i = \mathrm{P}\,(t_i|S_i) = \frac{1}{1 + e^{-(W_c u + b_c)}} \varepsilon\ (0, 1) \tag{5}$$

W_c is the weight matrix and b_c is the bias term. Finally, we use cross-entropy between the ground truth visit y_i and the predicted visit $\widehat{Y}_i$ to calculate the loss for each patient from all the timestamps as below:

$$\mathrm{L}\,(\theta) = -\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left(y_i^{\mathrm{T}}\log\left(\widehat{y}_t\right) + \left(1 - y_t\right)^{\mathrm{T}}\log\left(1 - \widehat{y}_t\right)\right) \tag{6}$$

MEG Data

We collected our MEG data by recording 32 epilepsy patients from Brain Hospital Affiliated to Nanjing Medical University, China. To ensure the balance of data distribution, half of them were males and the other half were females. The age range of these patients varies from 20 to 32. The Institutional Review Board was approved and written consent was obtained from all subjects.

In more detail, the sampling frequency of our MEG data is 1200 Hz, and the signals have been filtered by the band-pass filter (0.03∼300 Hz), which is then digitized at 1000 Hz. We collected at least 20 groups of data for each subject and every group was observed for 2 min. That is to say, the total duration of the MEG raw data on each subject was no less than 40 min. The distance of head movement before and after MEG data collection was also measured, then those with a distance greater than 5 mm were discarded and re-measured to ensure the quality of collected data. In the process of data collection, both audio and video systems were used to monitor the subjects constantly. Moreover, the subjects were requested to be in a supine position with their eyes closed and to keep relaxed, such that the resting-state data could be observed.

Evaluation Index

To evaluate the performance index, we first built the confusion matrix, upon which we further calculated the number of true-positive samples (TP), true-negative samples (TN), false-positive samples (FP), and false-negative (FN) samples.

To deal with the task of recognizing TLE subtypes, including CPS and SPS, four evaluation metrics are considered: precision (denoted as P), accuracy (denoted as ACC), recall (denoted as R), and the F1 score. In more detail, precision can be defined as:

$$\mathrm{p} = \frac{TP}{TP + FP} \tag{7}$$

Accuracy is expressed as the ratio of the number of correctly classified samples and the total number of samples on the test data set:

$$\mathrm{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

Recall rate can be formulated as:

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{9}$$

F1 value is the harmonic mean of precision rate and recall rate, which can be rewritten as:

$$\frac{2}{F1} = \frac{1}{P} + \frac{1}{R}\mathrm{F1} = \frac{2TP}{2TP + FP + FN} \tag{10}$$

As can be seen, a higher accuracy metric will lead to a better F1 score. Generally, the accuracy indicates the correct number of positive predictions. Recall represents the number of prediction-correct positive cases, which is directly related to the true-positive (TP) samples and false-negative samples (FN).

Compared Methods

To show the effectiveness of our proposed MSAM, we compare our method with other models including Convolutional Neural

Network (CNN), Support Vector Machine (SVM), and Random Forest (RF). CNN is a conventional deep learning model used to classify TLE subtypes, and SVM and RF are traditional machine learning algorithms. More details about these models are provided below:

Convolutional Neural Network (CNN): CNN is a kind of feed forward neural network with a deep structure using convolution computation. It is one of the most representative algorithms in deep learning.

Support Vector Machine (SVM): SVM is a kind of generalized linear classifier that classifies data using the supervised labels. Its decision boundary is the maximum margin hyperplane.

Random Forest (RF): RF is a classifier which contains multiple decision trees, and the output category is determined by the category output of individual trees.

RESULTS

Performance Comparison

For fair comparison, all methods, including the proposed MSAM and the compared CNN, SVM, and RF, are trained with 300 epochs with a batch size of 32 and the results are calculated using cross-validation across the entire dataset. The head number in the self-attention of our method is set to 4.

We report the experimental results in **Table 1**. As can be seen from the table, the performance of our proposed MSAM takes a lead position in comparison with others. Specifically, it increases the performance of CNN by 0.4% on accuracy, 1.1% on recall, 0.7% on precision, and 1.1% on F1-score. Also, it outperforms SVM by 28.4, 75.5, 35.8, and 15.2% on accuracy, recall, precision, and F1-score, respectively. Besides, 1.0, 21.9, 13.0, and −1.8% gains are, respectively, obtained w.r.t accuracy, recall, precision, and F1-score on the basis of RF. The above experiments well demonstrate the capacity of our method in dealing with the MEG data classification.

Effect of Varying Head Number in Self-Attention

To explore the effect of head number in our self-attention, in **Figure 2** we perform experiments with different head numbers of 2, 4, 8, and 16. Similarly, the results are calculated using cross-validation across the entire dataset **Figures 2A–D**, respectively, display our performance of accuracy, recall, precision, and F1-score with different head numbers.

As can be seen, our MSAM obtains its best accuracy and F1-score performance when the head number is 8, while best recall and precision was obtained with a head number of 4. To reduce the model complexity, in our implementations we set the head number to 4.

TABLE 1 | Comparison of CNN, SVM, RF, and the proposed MSAM.

Method	Accuracy	Recall	Precision	F1-score
CNN	0.832	0.898	0.9	0.823
SVM	0.552	0.154	0.549	0.682
RF	0.826	0.69	0.777	0.852
MSAM	0.836	0.909	0.907	0.834

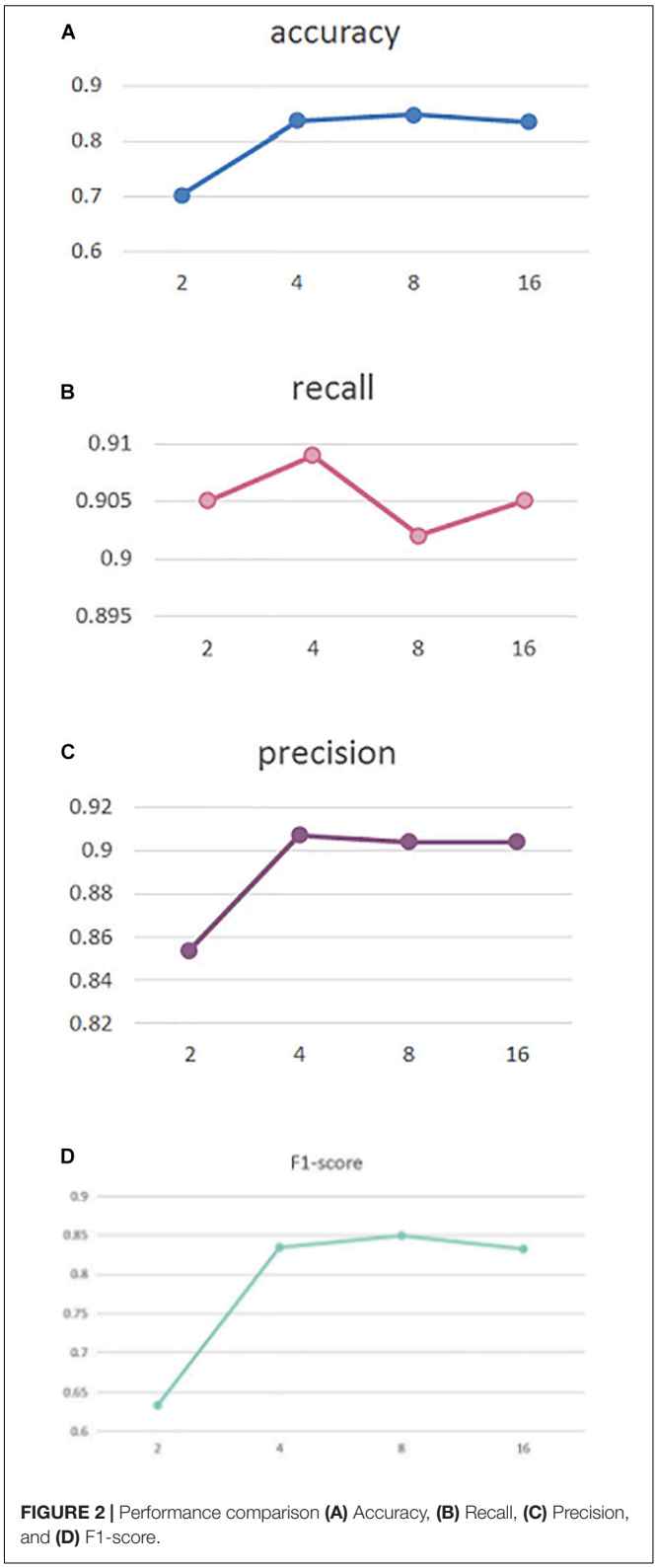


FIGURE 2 | Performance comparison **(A)** Accuracy, **(B)** Recall, **(C)** Precision, and **(D)** F1-score.

TABLE 2 | Comparison of ablated models.

Method	Accuracy	Recall	Precision	F1-score
attn_1	0.827	0.854	0.864	0.832
attn_2	0.825	0.845	0.859	0.831
MSAM	0.836	0.909	0.907	0.834

Ablation Study

We further conducted experiments to analyze the contributions of different components in our proposed method, including the multi-head self-attention layer and the self-attention layer. By, respectively, removing these two components, we have the following testing scenarios:

(1) Atten_1: The self-attention layer is removed;
(2) Atten_2: The multi-head self-attention layer is removed;
(3) MSAM: Both components are preserved, which makes up our method.

Following the above experimental settings, we train all models with 300 epochs and a batch size of 32. The head number is set to 4.

Table 2 shows the testing results. We can observe that our MSAM obtains the best performance, which demonstrates that both components of multi-head self-attention layer and the self-attention layer are crucial in boosting our classification performance.

DISCUSSION

To correctly classify the subtype of TLE, including CPS and SPS, is very important to the treatment of patients. Most existing studies focus on distinguishing if one person suffers from epilepsy while ignoring the importance of which type of epilepsy the patient is suffering from. To solve this, in this paper, we developed a deep learning-based classification model by integrating self-attention mechanism to enhance the classification of TLE subtypes.

To this end, our proposed MSAM model is performed on our collected MEG data from 32 patients, made up of 16 males and 16 females aged from 20 to 32. As shown in **Table 1**, the proposed MSAM significantly outperforms its counterparts, including CNN, SVM, and RF, with a supreme performance of 83.6% in accuracy, 90.9% on recall, 90.7% on precision, and 83.4% on F1-score. Thus, our method can be well applied to the problem of classifying the subtypes of TLE.

By setting head numbers to 2, 4, 8, and 16, we analyze the effect of different head numbers in **Figure 2**, which shows that head number of 4 and 8 rewards the best performance. In the experiments, to reduce the model complexity, we set it to 4.

We also conducted an ablation study to explore the efficacy of different components in our method. The experiments in **Table 2** show that the self-attention layer brings gains of 0.9, 5.5, 4.3, and 0.2% on accuracy, recall, precision, and F1-score, respectively. Besides, the multi-head self-attention layer also increased the performance by 1.1% on accuracy, 6.4% on recall, 4.8% on precision, and 0.3% on F1-score. Thus, both components of our MSAM play an important role in the classification of TLE subtypes.

Though significant contributions were made, limitations remain in this paper. First, the experiments were conducted on our collected single data source. More experiments on other datasets might be necessary to test the ability of the classification model. Second, epilepsy is a dynamically changed process. However, our classification focuses on patients from the same period, which may impede its practical applications. A long-term tracking experiment is needed. Third, this paper lacks studies on the prevention and early treatment of temporal lobe epilepsy. Due to the limited resources, we could not solve the limitations completely, which would be our focus in future work.

To summarize, this paper discusses and analyzes the classification of TLE subtypes. By integrating a self-attention mechanism, our MSAM is proposed to offer an effective classification model. The experimental results well demonstrate the effectiveness of our MSAM in classifying the TLE subtypes of CPS and SPS. Further works will be done to implement the limitations of this work as discussed above.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. Requests to access these datasets should be directed to TW, lvwu123@163.com.

ETHICS STATEMENT

The Institutional Review Board approved the study with human subjects and the written consents were obtained from all subjects.

AUTHOR CONTRIBUTIONS

PG, TW, LZ, YP, JG, and JX conceived and designed the experiments. LZ, YP, JM, XP, MZ, and JG performed the experiments. LZ, PG, HL, MZ, and JG wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by the Henan Key Science Research Project of Colleges and Universities (Grant No. 18B520041), the Nanjing Medical Science and Technique Development Foundation (Grant Nos. QRX17027 and YKK18118), Initial Scientific Research Fund of Ph.D. in Hubei University of Science and Technology (BK201802), Guidance Project of Hubei Educational Committee of China (B2019166), and the Undergraduate Innovation and Entrepreneurship Training Program of Hubei Province (S202010927039).

REFERENCES

Achilles, F., Tombari, F., Belagiannis, V., Loesch, A. M., Noachtar, S., and Navab, N. (2018). Convolutional neural networks for real-time epileptic seizure detection. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* 6, 264–269. doi: 10.1080/21681163.2016.1141062

Ahmedt-Aristizabal, D., Fookes, C., Nguyen, K., and Sridharan, S. (2018). "Deep classification of epileptic signals," in *Proceeding of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Honolulu, HI, 332–335. doi: 10.1109/embc.2018.8512249

Anastasiadou, M. N., Christodoulakis, M., Papathanasiou, E. S., Papacostas, S. S., Hadjipapas, A., and Mitsis, G. D. (2019). Graph theoretical characteristics of EEG-based functional brain networks in patients with epilepsy: the effect of reference choice and volume conduction. *Front. Neurosci.* 13:221. doi: 10.3389/fnins.2019.00221

Beghi, E., and Giussani, G. (2018). Aging and the epidemiology of epilepsy. *Neuroepidemiology* 51, 216–223. doi: 10.1159/000493484

Burns, S. P., Santaniello, S., Yaffe, R. B., Jouny, C. C., Crone, N. E., Bergey, G. K., et al. (2014). Network dynamics of the brain and influence of the epileptic seizure onset zone. *Proc. Natl Acad. Sci. U.S.A.* 111, E5321–E5330. doi: 10.1073/pnas.1401752111

Cantor-Rivera, D., Khan, A. R., Goubran, M., Mirsattari, S. M., and Peters, T. M. (2015). Detection of temporal lobe epilepsy using support vector machines in multi-parametric quantitative MR imaging. *Comput. Med. Imaging Graph.* 41, 14–28. doi: 10.1016/j.compmedimag.2014.07.002

Englot, D. J., Nagarajan, S. S., Wang, D. D., Rolston, J. D., Mizuiri, D., Honma, S. M., et al. (2016). The sensitivity and significance of lateralized interictal slow activity on magnetoencephalography in focal epilepsy. *Epilepsy Res.* 121, 21–28. doi: 10.1016/j.eplepsyres.2016.01.009

Fang, P., An, J., Zeng, L.-L., Shen, H., Chen, F., Wang, W., et al. (2015). Multivariate pattern analysis reveals anatomical connectivity differences between the left and right mesial temporal lobe epilepsy. *Neuroimage Clin.* 7, 555–561. doi: 10.1016/j.nicl.2014.12.018

Guo, J., Li, H., Pan, Y., Gao, Y., Sun, J., Wu, T., et al. (2020). Automatic and accurate epilepsy ripple and fast ripple detection via virtual sample generation and attention neural networks. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1710–1719. doi: 10.1109/tnsre.2020.3004368

Jette, N., Fiest, K. M., Sauro, K. M., Wiebe, S., and Patten, S. B. (2017). Author response: prevalence and incidence of epilepsy: a systematic review and meta-analysis of international studies. *Neurology* 89, 641–642. doi: 10.1212/wnl.0000000000004206

Liu, J., Sun, S., Liu, Y., Guo, J., Li, H., Gao, Y., et al. (2020). A novel megnet for classification of high-frequency oscillations in magnetoencephalography of epileptic patients. *Complexity* 2020:9237808. doi: 10.1155/2020/9237808

Mariani, V., Revay, M., D'Orio, P., Rizzi, M., Pelliccia, V., Nichelatti, M., et al. (2019). Prognostic factors of postoperative seizure outcome in patients with temporal lobe epilepsy and normal magnetic resonance imaging. *J. Neurol.* 266, 2144–2156. doi: 10.1007/s00415-019-09394-x

Niu, K., Guo, J., Pan, Y., Gao, X., Peng, X., Li, N., et al. (2020). Multichannel deep attention neural networks for the classification of autism spectrum disorder using neuroimaging and personal characteristic data. *Complexity* 2020:1357853. doi: 10.1155/2020/1357853

Pedreira, C., Vaudano, A. E., Thornton, R. C., Chaudhary, U. J., Vulliemoz, S., Laufs, H., et al. (2014). Classification of EEG abnormalities in partial epilepsy with simultaneous EEG-fMRI recordings. *Neuroimage* 99, 461–476. doi: 10.1016/j.neuroimage.2014.05.009

Peng, X., Long, G., Shen, T., Wang, S., and Jiang, J. (2020a). Self-attention enhanced patient journey understanding in healthcare system. *arXiv* [Preprint].

Peng, X., Long, G., Shen, T., Wang, S., Jiang, J., and Blumenstein, M. (2019). "Temporal self-attention network for medical concept embedding," in *Proceedings of the 19th IEEE International Conference on Data Mining, Beijing*, eds J. Wang, K. Shim, and X. Wu (Piscataway, NJ: IEEE), 498–507.

Peng, X., Long, G., Shen, T., Wang, S., Jiang, J., and Zhang, C. (2020b). BiteNet: bidirectional temporal encoder network to predict medical outcomes. *arXiv* [preprint].

Rajpoot, K., Riaz, A., Majeed, W., and Rajpoot, N. (2015). Functional connectivity alterations in epilepsy from resting-state functional MRI. *PLoS One* 10:e0134944. doi: 10.1371/journal.pone.0134944

Sarraf, S., and Tofighi, G. (2016). "Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data," in *Proceedings of the 2016 Future Technologies Conference (FTC)*, San Francisco, CA.

Serna, J.A.d.l.O, Arrieta Paternina, M. R., Zamora-Mendez, A., Tripathy, R. K., and Pachori, R. B. (2020). EEG-rhythm specific taylor-fourier filter bank implemented with o-splines for the detection of epilepsy using EEG signals. *IEEE Sens. J.* 20, 6542–6551. doi: 10.1109/jsen.2020.2976519

Shi, Q., Zhang, T., Miao, A., Sun, J., Sun, Y., Chen, Q., et al. (2020). Differences between interictal and ictal generalized spike-wave discharges in childhood absence epilepsy: a MEG study. *Front. Neurol.* 10:1359. doi: 10.3389/fneur.2019.01359

Siuly, S., and Li, Y. (2015). Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification. *Comput. Methods Programs Biomed.* 119, 29–42. doi: 10.1016/j.cmpb.2015.01.002

van Klink, N., and Zijlmans, M. (2019). High frequency oscillations in MEG: next steps in source imaging for focal epilepsy. *Brain* 142, 3318–3320. doi: 10.1093/brain/awz321

Vergun, S., Gaggle, W., Nair, V. A., Suhonen, J. I., Birn, R. M., Ahmed, A. S., et al. (2016). Classification and extraction of resting state networks using healthy and epilepsy fMRI data. *Front. Neurosci.* 10:440. doi: 10.3389/fnins.2016.00440

Wiebe, S. (2013). Definition of drug-resistant epilepsy: is it evidence based? *Epilepsia* 54, 9–12. doi: 10.1111/epi.12176

Wu, T., Chen, D., Chen, Q., Zhang, R., Zhang, W., Li, Y., et al. (2018). Automatic lateralization of temporal lobe epilepsy based on MEG network features using support vector machines. *Complexity* 2018, 1–10 doi: 10.1155/2018/4325096

Yao, J., Wang, H., and Xiao, Z. (2019). Correlation between EEG during AED withdrawal and epilepsy recurrence: a meta-analysis. *Neurol. Sci.* 40, 1637–1644. doi: 10.1007/s10072-019-03855-x

Zafar, R., Malik, A. S., Shuaibu, A. N., Rehman, M.J.u, and Dass, S. C. (2017). "Classification of fMRI data using support vector machine and convolutional neural network," in *Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications, Kuching*, (Piscataway, NJ: IEEE), 324–329.

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 17 December 2020
doi: 10.3389/fmolb.2020.613878

A Semi-automatic Diagnosis of Hip Dysplasia on X-Ray Films

Guangyao Yang[1†], Yaoxian Jiang[2†], Tong Liu[1], Xudong Zhao[1], Xiaodan Chang[2*] and Zhaowen Qiu[1,3*]*

[1] Department of Computer Science and Technology, College of Information and Computer Engineering, Northeast Forestry University, Harbin, China, [2] Department of Radiology, Affiliated Zhongshan Hosptial of Dalian University, Dalian, China, [3] Heilongjiang Tuomeng Technology Co. Ltd., Harbin, China

OPEN ACCESS

Edited by:
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Xingyu Liao,
Central South University, China
Renmin Han,
Shandong University, China

***Correspondence:**
Xudong Zhao
zhaoxudong@nefu.edu.cn
Xiaodan Chang
302647771@qq.com
Zhaowen Qiu
qiuzw@nefu.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 04 October 2020
Accepted: 25 November 2020
Published: 17 December 2020

Citation:
Yang G, Jiang Y, Liu T, Zhao X, Chang X and Qiu Z (2020) A Semi-automatic Diagnosis of Hip Dysplasia on X-Ray Films. Front. Mol. Biosci. 7:613878. doi: 10.3389/fmolb.2020.613878

Background: Diagnosis of hip joint plays an important role in early screening of hip diseases such as coxarthritis, heterotopic ossification, osteonecrosis of the femoral head, etc. Early detection of hip dysplasia on X-ray films may probably conduce to early treatment of patients, which can help to cure patients or relieve their pain as much as possible. There has been no method or tool for automatic diagnosis of hip dysplasia till now.

Results: A semi-automatic method for diagnosis of hip dysplasia is proposed. Considering the complexity of medical imaging, the contour of acetabulum, femoral head, and the upper side of thigh-bone are manually marked. Feature points are extracted according to marked contours. Traditional knowledge-driven diagnostic criteria is abandoned. Instead, a data-driven diagnostic model for hip dysplasia is presented. Angles including CE, sharp, and Tonnis angle which are commonly measured in clinical diagnosis, are automatically obtained. Samples, each of which consists of these three angle values, are used for clustering according to their densities in a descending order. A three-dimensional normal distribution derived from the cluster is built and regarded as the parametric model for diagnosis of hip dysplasia. Experiments on 143 X-ray films including 286 samples (i.e., 143 left and 143 right hip joints) demonstrate the effectiveness of our method. According to the method, a computer-aided diagnosis tool is developed for the convenience of clinicians, which can be downloaded at http://www.bio-nefu.com/HIPindex/. The data used to support the findings of this study are available from the corresponding authors upon request.

Conclusions: This data-driven method provides a more objective measurement of the angles. Besides, it provides a new criterion for diagnosis of hip dysplasia other than doctors' experience deriving from knowledge-driven clinical manual, which actually corresponds to very different way for clinical diagnosis of hip dysplasia.

Keywords: hip joint, dysplasia, x-ray, manual segmentation, automatic angle measurement, density descending clustering

1. INTRODUCTION

Hip is one of the largest joint in human body. Its normal structure maintains people's daily activities. Hip dysplasia, which is thought to be hereditary (Harsanyi et al., 2020), is the main cause of hip osteoarthritis (Ganz et al., 2008). If the surface of acetabular is too small or tilts for a long time, the femoral head cannot be completely covered. Therefore, it will lead to uneven pressure, which will develop into irreversible osteoarthritis in the end (Yasuda et al., 2020). Early screening of hip dysplasia for adults followed by proper clinical management can not only save medical resources but also keep patients away from the pain of operation (Gala et al., 2016).

Although three dimensional structure of hip joint can be derived from CT and MRI images, it has to be faced with relatively high fees of medical check and high radiation. Therefore, the radiograph of pelvis from a X-ray film becomes the main early diagnosis of hip dysplasia (Kayaalp et al., 2020; Powell et al., 2020). Inevitably, the ever-growing numbers of X-ray films increase the burden of radiologists. An automatic method or tool for them to make auxiliary measurement or even diagnosis is needed.

As to automatic measurement or diagnosis of hip dysplasia on X-ray films, there are three problems. Firstly, it is difficult to segment the hip automatically considering the inhomogeneous intensity derived from the image superimposition of acetabulum. Thus, most of the existing methods are aimed at femur segmentation. Xie et al. (2014) extracted shape features to segment the proximal femur. Wei et al. (2020) improved deep convolutional generative adversarial network (DCGAN) for segmentation of femur. Liu et al. (2020) proposed a Pyramid Nonlocal UNet (PN-UNet) for automatic misshapen landmark detection and neighboring patch segmentation. However, boundaries of acetabulum and femur were not clearly marked. Secondly, prevailing diagnosis of hip dysplasia mainly depends on the manual measurement of angles on X-ray films (Simone and Klaus, 2014). It is difficult to automatically calibrate feature angles including center-edge (CE) angle, sharp angle, Tonnis angle, etc., which are commonly measured in clinical diagnosis (Beltran et al., 2013). Thirdly, a diagnostic manual is consulted to test whether hip development is normal or not (Harper et al., 2020; Ömeroglu et al., 2020). Automatic and objective indicators for clinical diagnosis of hip dysplasia are needed.

In this paper, we propose an approach for semi-automatic diagnosis of hip dysplasia on X-ray films. The corresponding processing framework is shown in **Figure 1**. Due to the difficulty of accurate acetabulum segmentation, contours of hip joint including acetabulum, femoral head, and the upper side of thigh-bone are manually delineated. Then, feature angles including CE, sharp, and Tonnis are automatically extracted from the marked contour. Thereafter, a scatter point is obtained in three-dimensional space according to these feature angles. These procedure is repeated using 286 samples representing either left or right hip joints. Finally, previously proposed clustering method using density in a descending order is presented on these scatter points, and a model representing normal hip development is made for diagnosis of hip dysplasia.

2. METHOD

The dataset representing either normal development of hip joint or hip dysplasia is provided, which contains 143 X-ray films including 286 samples from 143 left and 143 right hip joints. That is, an X-ray film is considered as two parts, each of which contains either imaging of left hip joint or the right one. Manual delineation of acetabulum, femoral head, and the upper side of thigh-bone is made on each film. Actually, we follow the framework presented in **Figure 1** to establish the model representing normal hip development for diagnosis of hip dysplasia. More details can be seen in the following subsections.

2.1. Automatic Extraction of CE

According to the result of manual delineation, some feature points can be obtained, which help to form feature angles automatically. Sketch maps of these feature angles are illustrated in **Figure 2**. CE angle is commonly considered to be the first feature angle for clinical diagnosis. As shown in **Figure 2A**, it refers to the angle between two lines. One line is derived from the connection between the central point of femoral head (i.e., a_1 or a_2) and the external upper edge of acetabulum (i.e., b_1 or b_2). The other line refers to the vertical of the line after connecting the two central points of right and left femoral head which two are labeled as a_1 and a_2, respectively.

In order to automatically obtain CE angle, a_1, a_2, b_1, and b_2 have to be determined in advance. According to manual delineation, the external upper edge of acetabulum, i.e., b_1 and b_2, can be apparently labeled. As to the central point of the femoral head, it is regarded as the center of a circle which can cover the femoral head. Considering that three points which are not collinear can determine a circle in a plane, three feature points have to be automatically labeled. Here, the uppermost and outermost point on the contour of the femoral head are selected, as labeled with d_1, d_2, e_1, and e_2 shown in **Figure 3**. Taking the inflection points of femoral head and femoral neck as the boundary, e_1 and e_2 can be simply found. The third feature point can be obtained by connecting $e_{1,2}$ and $g_{1,2}$, which refer to the lower edge point of right and left acetabulum. That is the intersection of the connection line and the contour of femoral head denoted as f_1 and f_2 in **Figure 3**. Thus, the central point of the femoral head are obtained and labeled as a_1 and a_2 in **Figure 3**. Correspondingly, CE angle can be expressed as

$$\theta_{CE} = \arccos \frac{\overrightarrow{a_{1,2}c_{1,2}} \cdot \overrightarrow{a_{1,2}b_{1,2}}}{|\overrightarrow{a_{1,2}c_{1,2}}|\ |\overrightarrow{a_{1,2}b_{1,2}}|}. \quad (1)$$

2.2. Automatic Extraction of Sharp

Sharp angle is also regarded as a feature angle for clinical diagnosis. As shown in **Figure 2B**, it also refers to the angle between two lines. One line is derived from the connection between the lower edge point of the right acetabulum (i.e., a_1) and that of the left acetabulum (i.e., a_2). The other line refers to

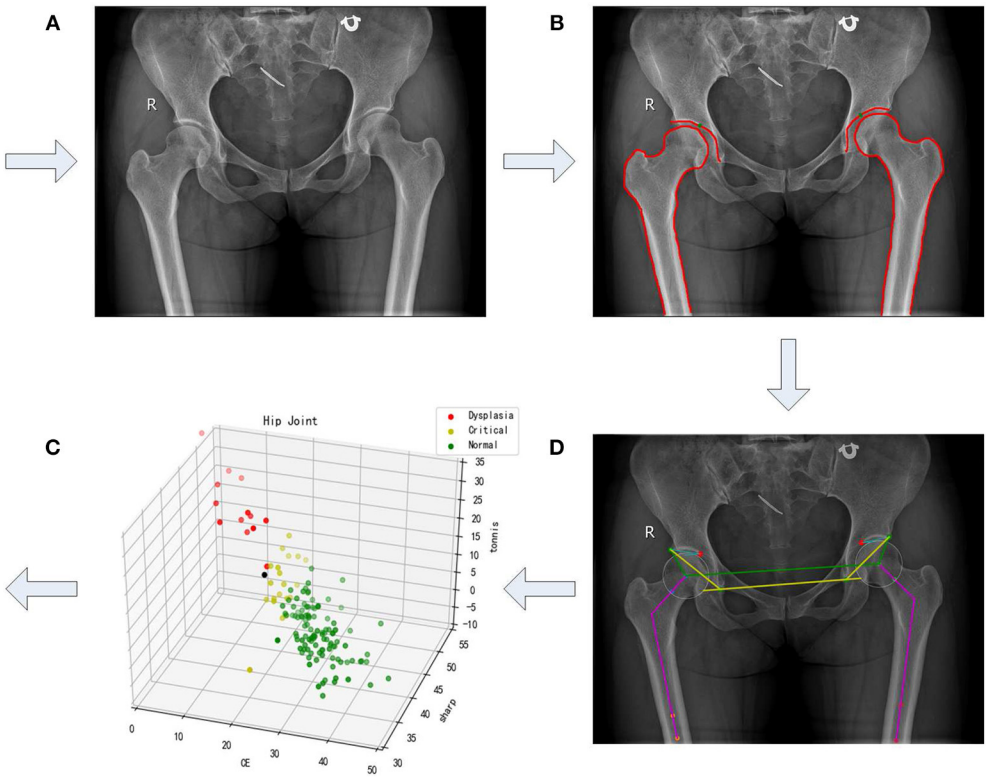


FIGURE 1 | A framework of semi-automatic method for diagnosis of hip dysplasia on X-ray films. **(A–D)** correspond to four steps of the framework. **(A)** refers to an original X-ray file of hip joint. **(B)** represents the manually marked results. **(C)** corresponds to the automatic angle measurement. **(D)** shows the corresponding scatter in the three-dimensional space with its three directions representing CE, sharp, and Tonnis angle.

the connection between the lower edge point of the acetabulum (i.e., a_1 or a_2) and the external upper edge of acetabulum (i.e., b_1 or b_2). Correspondingly, sharp angle is expressed as follows. That is,

$$\theta_{sharp} = \arccos \frac{\overrightarrow{a_{1,2}b_{1,2}} \cdot \overrightarrow{a_{2,1}a_{1,2}}}{|\overrightarrow{a_{1,2}b_{1,2}}|\,|\overrightarrow{a_{2,1}a_{1,2}}|}. \quad (2)$$

2.3. Automatic Extraction of Tonnis

Tonnis angle is also considered to be a feature angle for clinical diagnosis of hip dysplasia. As illustrated in **Figure 2C**, it also refers to the angle contained by two lines. One line is derived from the connection between the external upper edge of acetabulum (i.e., b_1 or b_2) and the lower edge of the weight-bearing area of acetabulum (i.e., a_1 or a_2). The other line refers to the connection between the lower edge point of the right acetabulum (i.e., g_1) and that of the left acetabulum (i.e., g_2).

In order to automatically obtain Tonnis angle, a_1 and a_2 have to be pointed out in advance. In fact, the lower edge of the weight-bearing area of acetabulum can be easily found in an X-ray film containing hip joint, for the weight-bearing area of acetabulum keeps an obvious contrast to its surrounding area (see **Figure 1A**). Therefore, these two points can be labeled during manual delineation (see green points in **Figure 1B**). Correspondingly, Tonnis angle is expressed as

$$\theta_{Tonnis} = \arccos \frac{\overrightarrow{a_{1,2}b_{1,2}} \cdot \overrightarrow{g_{2,1}g_{1,2}}}{|\overrightarrow{a_{1,2}b_{1,2}}|\,|\overrightarrow{g_{2,1}g_{1,2}}|}. \quad (3)$$

2.4. Clustering Using Sample Density in a Descending Order

As to each sample derived from an X-ray film, its CE, sharp, and Tonnis angle can be automatically calculated using Equations (1), (2), and (3). Correspondingly, a three-dimensional scatter point can be obtained associated with this sample. This procedure combining manual delineation of contours with automatic extraction of angles is repeated n times, where n represents sample size. Thus, we get a three-dimensional scatter plot as shown in **Figure 1D**. Using previously proposed clustering method (Liu et al., 2019), the cluster corresponding to normal hip development is obtained. If samples within the cluster are considered to obey three-dimensional normal distribution, then a statistical model can be established. The corresponding probability density function is expressed as

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{(3/2)}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}, \quad (4)$$

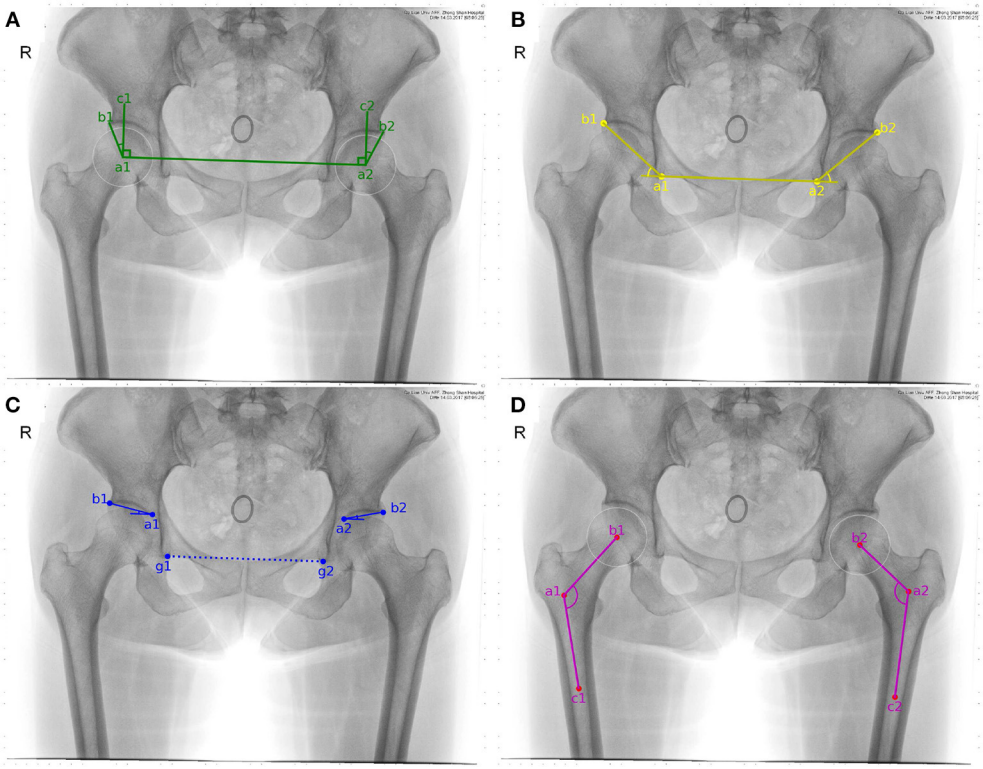


FIGURE 2 | Sketch maps of feature angles. **(A–D)** correspond to CE, sharp, Tonnis, and caput collum diaphysis (CCD) angle, respectively. Considering that CCD only measures coxa valga or coxa vara on femoral head, it is discarded in the following study.

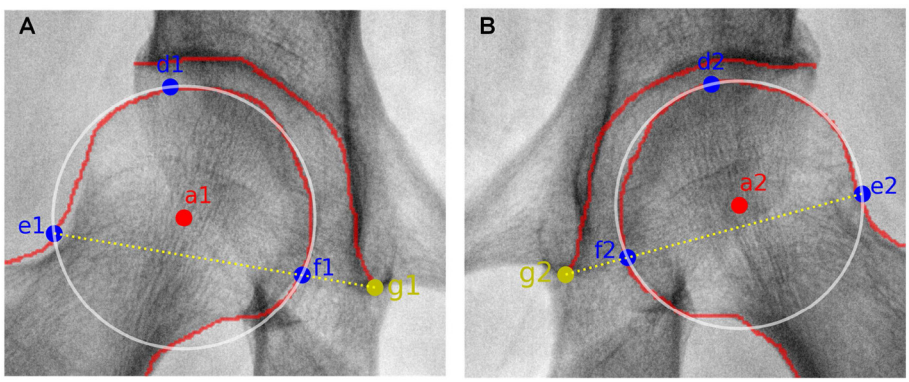


FIGURE 3 | Sketch maps of getting the circle representing the femoral head. **(A,B)** correspond to results from the right and left femoral head. $a_{1,2}$ refer to the central point of the right and left femoral head, respectively. $d_{1,2}$ and $e_{1,2}$ represent the uppermost and outermost points of the right and left femoral head, respectively. $g_{1,2}$ correspond to the lower edge points of the right and left acetabulum, respectively. As to $f_{1,2}$, they are the intersection points of the connection line (i.e., $e_{1,2}$ and $g_{1,2}$) and the contour of right and left femoral head, respectively.

where $\boldsymbol{x}$ represents the vector $(\theta_{CE}, \theta_{sharp}, \theta_{Tonnis})^T$. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote sample mean and covariance matrix, respectively.

3. RESULTS

Experiments were conducted on 286 samples derived from 143 X-ray films of hip joint which contained 143 left and 143 right hip joint. The data was digital bilateral hip x-rays retrospectively collected from the Affiliated Zhongshan Hospital of Dalian University from January 2017 to January 2018. This study was approved by the hospital's ethics committee. All patient information was de-identified before data analysis.

The procedure shown in **Figure 1** was accomplished using our own developed graphical user interface (GUI) listed in **Figure 4**.

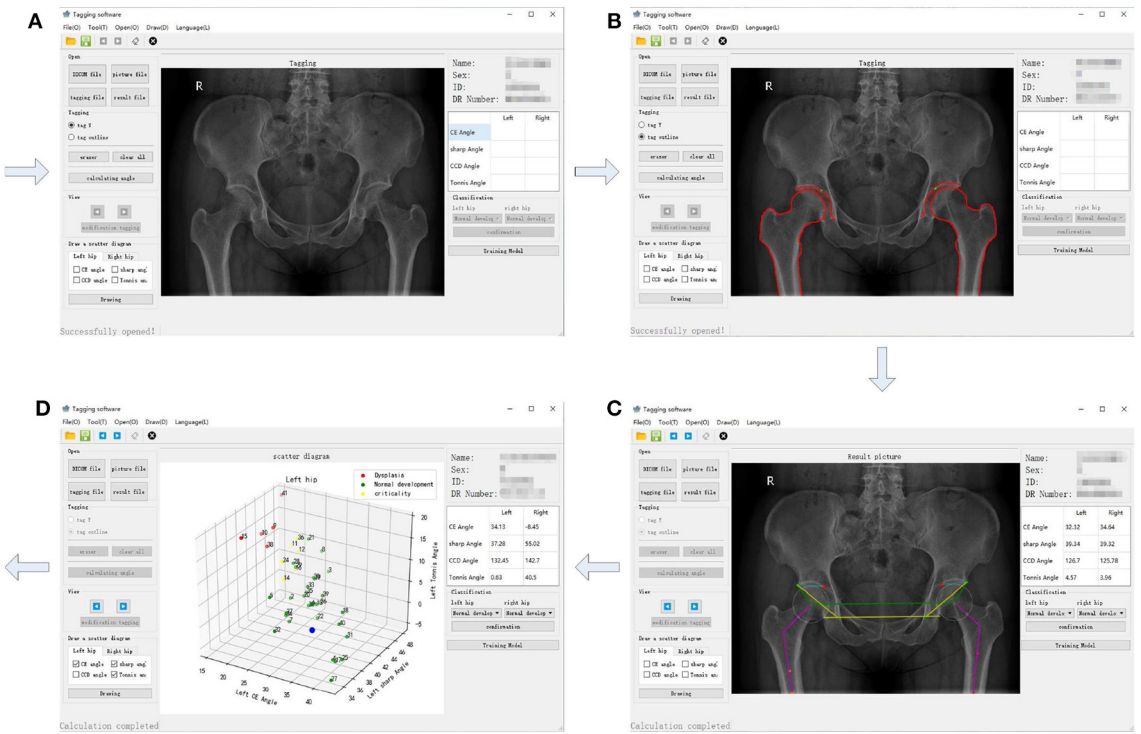


FIGURE 4 | The graphical user interface (GUI) of the semi-automatic method for diagnosis of hip dysplasia on X-ray films. **(A–D)** correspond to four steps of the GUI.

An X-ray film can be imported, as shown in **Figure 4A**. In **Figure 4B**, contours of acetabulum, femoral head, and the upper side of thigh-bone can be manually marked, together with the lower edge of the weight-bearing area of acetabulum. Then, feature angles including CE, sharp and Tonnis are automatically extracted, as illustrated in **Figure 4C**. Thereafter, a scatter point is projected into a three-dimensional space, which is composed of CE, sharp and Tonnis angle values calculated using Equations (1), (2), and (3), respectively. This procedure is repeated until enough scatter points have been got. Then, one can push the "Training Model" button shown in **Figure 4D** to establish the data-driven model for further clinical diagnosis of hip dysplasia.

Meanwhile, each CE, sharp, and Tonnis angle of the 286 samples were measured by a radiologist. The reference value of CE angle was considered as the knowledge-driven diagnostic criteria. Typically, a diagnosis of hip dysplasia was made, when $\theta_{CE} < 20^{\circ}$. On the contrary, it was thought to be normal development of hip joint, when $\theta_{CE} > 25^{\circ}$. Besides, it was considered as borderline dysplasia, when $20^{\circ} \leq \theta_{CE} \leq 25^{\circ}$. Therefore, 286 samples were labeled with color red, green and yellow, corresponding to abnormal, normal and borderline development of hip joint, respectively. Accordingly, the three-dimensional scatter plot and its two-dimensional projection results are listed in turn, as shown in **Figures 5A–D**. From these sub-figures, it can be seen that these samples obey normal distribution. However, plane or lines perpendicular to CE axis are considered to be the classification boundary using CE angle as the diagnostic criteria. Besides, it may be inappropriate even considering CE, sharp, and Tonnis angle at the same time, for the formed classification boundaries can be only perpendicular to coordinate axes. In fact, the appropriate classification boundary should be the plane perpendicular to the long axis of the ellipsoid derived from the three-dimensional normal distribution of samples.

Thus, previously proposed clustering method according to sample's density in a descending order (Liu et al., 2019) was utilized. In addition, samples within the cluster keeping the highest density peak, which represent normal development of hip joint, are considered to obey three-dimensional normal distribution. Using Equation (**4**), a model representing normal hip development is established for diagnosis of hip dysplasia, as shown in **Figure 5E**. Its two-dimensional projections are listed in turn, as shown in **Figures 5F–H**.

4. DISCUSSIONS

We intend to make some simple discussions as follows. Firstly, it needs to be considered whether automatic segmentation can be made. In fact, we have utilized U-Net (Ronneberger et al., 2015) for discovering contours of acetabulum, femoral head, and the upper side of thigh-bone. Due to the limited sample size, the segmentation results were not good. Besides, it was found X-ray films from different digital radiography (DR) machines

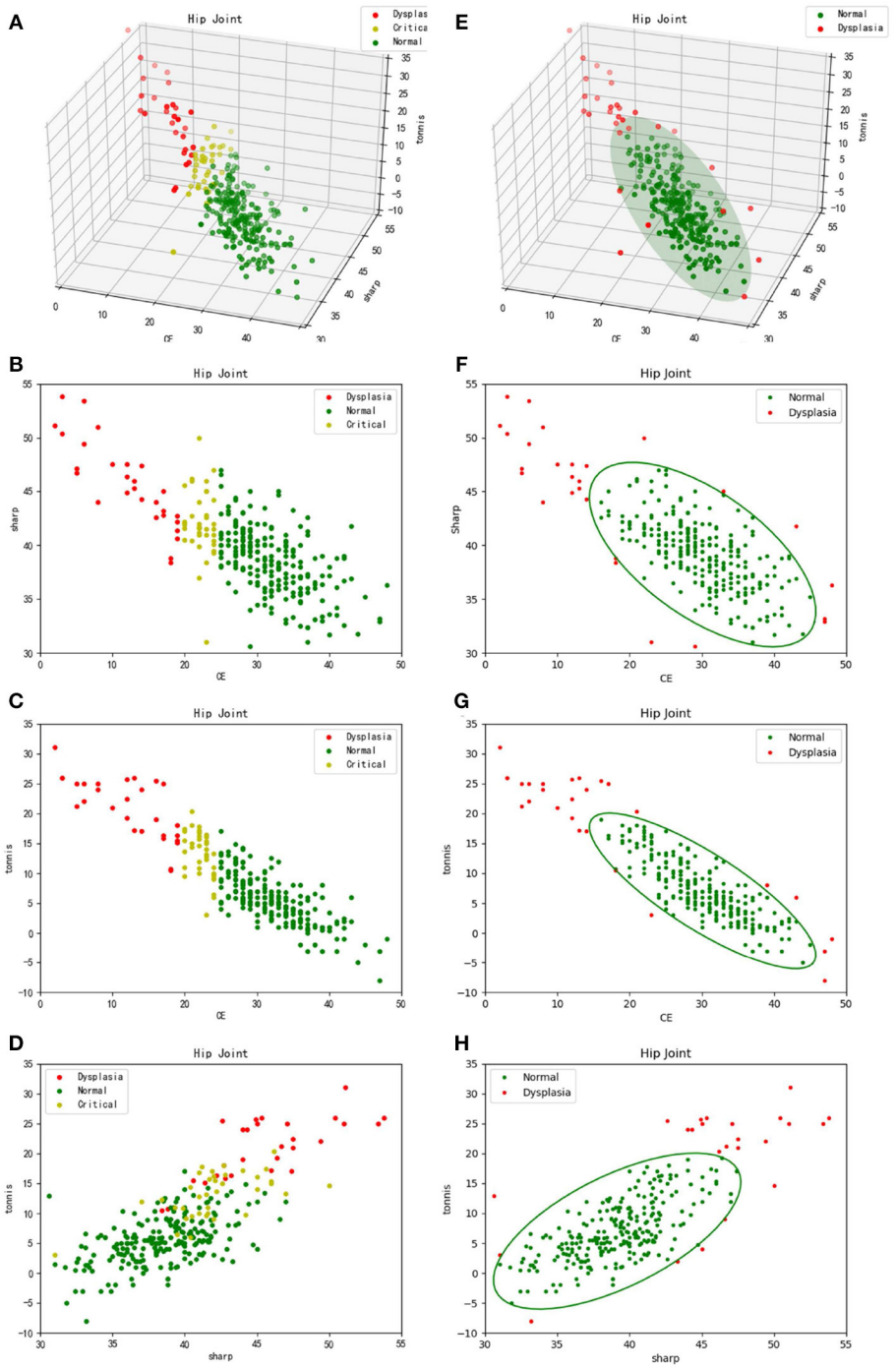


FIGURE 5 | Three-dimensional scatter plots and their two-dimensional projection results. **(A–D)** refer to the traditional diagnosis of hip dysplasia deriving from knowledge-driven clinical manual in a three-dimensional space and its two-dimensional projection subspaces, respectively. Therefore, θ_{CE} is to be considered. That is, samples labeled with different colors are derived from different θ_{CE} scopes. **(E–H)** correspond to the new data-driven criterion for diagnosis of hip dysplasia in a three-dimensional space and its two-dimensional projection subspaces, respectively. The ellipsoid and corresponding projection ellipses refer to the parametric model.

were quite different (see **Figures 1**, **2**). That makes automatic segmentation more hard.

Secondly, it needs to be discussed whether the knowledge-driven diagnostic criteria is effective or not. CE, sharp and Tonnis angle are commonly used as the measurements of hip joint development. The typical thresholds for diagnosis of dysplasia are $\theta_{CE} < 20^{\circ}$ and $\theta_{Tonnis} > 10^{\circ}$ (Kosuge et al., 2013), which empirically represent the over-shallow

acetabular and over-upward inclination of the weight-bearing area of acetabulum, respectively. In contrast, we proposed a data-driven diagnostic model for hip dysplasia, which took full account of sample distribution. After comparing the experimental results shown in **Figure 5**, it can be concluded that the data-driven criterion for diagnosis of hip dysplasia is more suitable, because it fits the sample distribution better.

Thirdly, outliers shown in **Figure 5E** need to be further considered. For those with small θ_{CE}s but high θ_{sharp}s and θ_{Tonnis}s, diagnosis of hip dysplasia can be made. However, other outliers need to be further discussed. The reason why they are different from the traditional samples with hip dysplasia and whether they belong to new subtypes of hip dysplasia or not needs to be explained. Therefore, these cases should be carefully selected. Except for X-ray films, other imaging diagnosis and clinical diagnosis should be provided to test whether these special outliers belong to hip dysplasia or not.

5. CONCLUSION

Diagnosis of hip dysplasia plays a vital role in early screening of hip diseases. In this study, we proposed a semi-automatic method for diagnosis of hip dysplasia on X-ray films. Due to the complex appearances of hip joint imaging on X-ray films, a manual delineation was made on contours of acetabulum, femoral head and the upper side of thigh-bone. Furthermore, feature points were automatically or semi-automatically extracted. Then, feature angles were automatically obtained. Samples derived from three feature angles (i.e., CE, sharp and Tonnis) were used to accomplish clustering, which helped to establish a criterion model based on three-dimensional normal distribution for diagnosis of hip dysplasia. Besides, a GUI was provided for the convenience of clinicians. In future work, automatic segmentation of hip joint will be considered.

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 the ethics committee of the hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XZ, XC, and ZQ conceived the general project and supervised it. XZ initiated the idea, conceived the whole process, and finalized the paper. GY and TL were the principal developers. YJ collected the cases and made the manual delineation. All authors read and approved the final manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research has been supported by the financial support of Major Projects of the Central Government to Guide the Development of Science and Technology for Local Government (No. ZY18C01), the Transverse Project of Northeast Forestry University (No. 20190424) and the financial support of Natural Science Foundation of Heilongjiang Province (No. LH2020F002).

REFERENCES

Beltran, L. S., Rosenberg, Z. S., Mayo, J. D., De Tuesta, M. D., Martin, O., Neto, L. P., et al. (2013). Imaging evaluation of developmental hip dysplasia in the young adult. *Am. J. Roentgenol.* 200, 1077–1088. doi: 10.2214/AJR.12.9360

Gala, L., Clohisy, J. C., and Beaule, P. E. (2016). Hip dysplasia in the young adult. *J. Bones Joint Surg.* 98, 63–73. doi: 10.2106/JBJS.O.00109

Ganz, R., Leunig, M., Leunig-Ganz, K., and Harris, W. H. (2008). The etiology of osteoarthritis of the hip. *Clin. Orthop. Relat. Res.* 466, 264–272. doi: 10.1007/s11999-007-0060-z

Harper, P., Joseph, B. M., Clarke, N. M., Herrera-Soto, J., Sankar, W. N., Schaeffer, E. K., et al. (2020). Even experts can be fooled: reliability of clinical examination for diagnosing hip dislocations in newborns. *J. Pediatr. Orthop.* 40, 408–412. doi: 10.1097/BPO.0000000000001602

Harsanyi, S., Zamborsky, R., Kokavec, M., and Danisovic, L. (2020). Genetics of developmental dysplasia of the hip. *Eur. J. Med. Genet.* 63:103990. doi: 10.1016/j.ejmg.2020.103990

Kayaalp, M. E., Can, A., Erdogan, F., Ozsahin, M. K., Aydingoz, O., and Kaynak, G. (2020). Clinical and radiological results of crowe type 3 or 4 dysplasia patients operated on with total hip arthroplasty using a cementless rectangular femoral component without fixating or grafting the transverse osteotomy site. *J. Arthroplasty* 35, 2537–2542. doi: 10.1016/j.arth.2020.04.045

Kosuge, D., Yamada, N., Azegami, S., Achan, P., and Ramachandran, M. (2013). Management of developmental dysplasia of the hip in young adults: current concepts. *Bone Joint J.* 95B, 732–737. doi: 10.1302/0301-620X.95B6.31286

Liu, C., Xie, H., Zhang, S., Mao, Z., Sun, J., and Zhang, Y. (2020). Misshapen pelvis landmark detection with local-global feature learning for diagnosing developmental dysplasia of the hip. *IEEE Trans. Med. Imaging* 39, 3944–3954. doi: 10.1109/TMI.2020.3008382

Liu, T., Li, H., and Zhao, X. (2019). Clustering by search in descending order and automatic find of density peaks. *IEEE Access* 7, 133772–133780. doi: 10.1109/ACCESS.2019.2939437

Ömeroglu, H., Tatlici, E., and Kose, N. (2020). Significance of asymmetry of groin and thigh skin creases in developmental dysplasia of the hip revisited: results of a comparative study. *J. Pediatr. Orthop.* 40, e761–e765. doi: 10.1097/BPO.0000000000001531

Powell, J., Gibly, R. F., Faulk, L. W., Carry, P., Mayer, S. W., and Selberg, C. M. (2020). Can eos imaging substitute for conventional radiography in measurement of acetabular morphology in the young dysplastic hip? *Pediatr. Orthop.* 40, 294–299. doi: 10.1097/BPO.0000000000001545

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science*, eds N. Navab, J. Homegger, W. Wells, and A. Frangi (Munich: Springer), 234–241.

Simone, W., and Klaus, W. (2014). Measurements and classifications in musculoskeletal radiology. *J. Magn. Reson. Imaging* 40, 248–249. doi: 10.1002/jmri.24618

Wei, Q., Han, J., Jia, Y., Zhu, L., Zhang, S., Lu, Y., et al. (2020). An approach for fully automatic femoral neck-shaft angle evaluation on radiographs. *Rev. Sci. Instrum.* 91:013706. doi: 10.1063/1.5089738

Xie, W., Franke, J., Chen, C., Gruetzner, P. A., Schumann, S., Nolte, L., et al. (2014). Statistical model-based segmentation of the proximal femur in digital antero-posterior (ap) pelvic radiographs. *Int. J. Comput. Assist. Radiol. Surg.* 9, 165–176. doi: 10.1007/s11548-013-0932-5

Yasuda, T., Matsunaga, K., Hashimura, T., Tsukamoto, Y., Sueyoshi, T., Ota, S., et al. (2020). Characterization of rapidly progressive osteoarthritis of the hip in its early stage. *Eur. J. Rheumatol.* 7, 130–134. doi: 10.5152/eurjrheum.2020.19159

Conflict of Interest: ZQ was employed by the company Heilongjiang Tuomeng Technology Co. Ltd., Harbin, China.

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.

frontiers in Physiology

REVIEW
published: 23 December 2020
doi: 10.3389/fphys.2020.611596

Tactile Perception Technologies and Their Applications in Minimally Invasive Surgery: A Review

Chao Huang[1,2], Qizhuo Wang[1], Mingfu Zhao[1], Chunyan Chen[2], Sinuo Pan[2] and Minjie Yuan[2]*

[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, [2]Ningbo Institute of Information Technology Application, Chinese Academy of Sciences, Ningbo, China

Minimally invasive surgery (MIS) has been the preferred surgery approach owing to its advantages over conventional open surgery. As a major limitation, the lack of tactile perception impairs the ability of surgeons in tissue distinction and maneuvers. Many studies have been reported on industrial robots to perceive various tactile information. However, only force data are widely used to restore part of the surgeon's sense of touch in MIS. In recent years, inspired by image classification technologies in computer vision, tactile data are represented as images, where a tactile element is treated as an image pixel. Processing raw data or features extracted from tactile images with artificial intelligence (AI) methods, including clustering, support vector machine (SVM), and deep learning, has been proven as effective methods in industrial robotic tactile perception tasks. This holds great promise for utilizing more tactile information in MIS. This review aims to provide potential tactile perception methods for MIS by reviewing literatures on tactile sensing in MIS and literatures on industrial robotic tactile perception technologies, especially AI methods on tactile images.

Keywords: tactile sensors, tactile perception, tactile images, minimally invasive surgery, robotic surgery, artificial intelligence

OPEN ACCESS

Edited by:
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Yan Wang,
Jilin University, China
Renmin Han,
Shandong University, Finland
Jiayang Guo,
University of Cincinnati, United States

****Correspondence:***
Chao Huang
chuang@ict.ac.cn

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

Received: *29 September 2020*
Accepted: *16 November 2020*
Published: *23 December 2020*

Citation:
Huang C, Wang Q, Zhao M, Chen C, Pan S and Yuan M (2020) Tactile Perception Technologies and Their Applications in Minimally Invasive Surgery: A Review. Front. Physiol. 11:611596. doi: 10.3389/fphys.2020.611596

INTRODUCTION

Minimally invasive surgery (MIS) is a surgery approach that provides indirect access to anatomy for surgeons by introducing specially designed surgical instruments or flexible catheters into a patient's body through minimally sized incisions (Verdura et al., 2000). Compared to conventional open surgery, MIS offers many advantages including reduced anesthesia and hospitalization time, mitigated tissue trauma and risk of postoperative infection, decreased intraoperative blood loss, and accelerated recovery (Puangmali et al., 2008). However, the indirect access to the anatomy brings two challenges: low degree of freedom (DOF) during manipulation and absence of tactile feedback during tool–tissue interactions (Abushagur et al., 2014). With the development of mechatronics, robot-assisted minimally invasive surgery (RMIS) systems, such as the ZEUS Surgical System (Uranues et al., 2002) and the da Vinci Surgical System (Guthar and Salisbury, 2000), have been developed to improve the dexterity of tools during manipulation, which partly resolve the motion constrain problem. Despite this, there are still limitations existing for MIS, including reduced hand-eye coordination, a narrowed field of vision, and

limited workspace of the tools (Bandari et al., 2020). More importantly, surgeons have little tactile information in MIS compared to the rich tactile feedback of human hands, which severely impairs the surgeon's ability to control the applied forces, thus causing extra tissue trauma or unintentional damage to healthy tissue (Ahmadi et al., 2012).

Tactile feelings, including but not limited to force, distributed pressure, temperature, vibrations, and texture, are complicated information that a human obtains through cutaneous receptors during physical interaction with environment. Depending on the sensing modalities, tactile sensors can be categorized into different kinds, including force sensors for measuring contact forces, slippage sensors for detecting slippage between tissue, and surgical instruments vibration sensors for measuring vibrations during contact. The goal of tactile technologies in MIS is to restore all the tactile information so that surgeons feel they are contacting that patients' anatomy directly with their own hands rather than operating a mechanism. Among this tactile information, force data are relatively easy to acquire, model, quantify, and display, so it is most widely used in MIS. The sensing principles, design requirements, specifications, developments of force sensors, and their applications in MIS have been thoroughly reviewed (Eltaib and Hewit, 2003; Puangmali et al., 2008; Schostek et al., 2009; Tiwana et al., 2012; Abushagur et al., 2014; Konstantinova et al., 2014; Saccomandi et al., 2014; Park et al., 2018; Al-Handarish et al., 2020; Bandari et al., 2020). In contrast, studies on utilizing other tactile information in MIS are very rare. Researchers have begun to realize the advantages of various tactile information in MIS, but challenges remain. Van der Putten et al. found that slippage and texture information can augment force information to prevent tissue trauma during manipulation but limited by cost and changes in instability; few studies were about texture information (Westebring-van der Putten et al., 2008). Okamura found some studies on tactile sensor arrays to perceive pressure distribution or deformation over a contact area, but it was challenging to acquire and display tactile data due to size and weight constraints (Okamura, 2009).

Tactile sensors are often categorized into single-point tactile sensor and the tactile array with respect to their spatial resolution. The single-point tactile sensor is usually embedded in the tip of the equipment to confirm the object-sensor contact and detect tactile signals at the contact point. The tactile array is composed of several single-point tactile sensors arranged according to certain rules. Compared with single-point tactile sensors, tactile array sensor can cover a wider area and can capture the tactile information of the object from multiple dimensions, so it can achieve high spatial resolution of touch.

In the field of industrial robots, tactile perception technologies have received considerable attention. Tactile perception is a procedure that obtains tactile information from tactile data sensed by tactile sensors. Many methods have been proposed to accomplish robot tactile perception tasks, including shape recognition, texture recognition, stiffness recognition, and sliding detection (Liu et al., 2017). In the early years, single-point tactile sensors were used to create point cloud models to finish tactile perception tasks. A current trend of tactile perception researches is to represent tactile data as images, where a tactile element is treated as an image pixel. From tactile images that tactile sensor arrays acquired, features are extracted, such as statistical features, vision feature descriptors, principal component analysis (PCA)-based features, and self-organizing features (Luo et al., 2017). These features are usually processed by AI methods like clustering, support vector machine (SVM), and deep neural networks, to obtain tactile information.

Robotically assisted surgery is a type of surgical procedure that is done using robotic systems. It was developed to overcome the limitations of pre-existing minimally invasive surgery and to enhance the capabilities of surgeons performing open surgery. According to their level of autonomy, surgical robotic systems are often classified into two categories: autonomous systems, which automatically execute tasks without interventions of the practitioner, and nonautonomous systems, which reproduce the surgeon's motion in either a master/slave teleoperated configuration or a hands-on configuration (Okamura, 2009). Due to the technical complications and high demanded reliability, most surgical robots belong to the second category. However, the development of robot tactile perception is promising for autonomous robotic systems. In the last decades, sensors have become smaller, cheaper, and more robust. Enormous studies on industrial robots aimed to perceive tactile in small areas like fingertips, on which sensors are tiny. Some studies accomplish tactile perception tasks with sensors made of soft material. In MIS, tactile information is usually displayed in the form of raw tactile data, which demands extra analysis. These tactile perception studies make it possible to provide more intuitive tactile information (e.g., stiffness distribution map) for surgeons utilizing nonautonomous surgical robotic systems and offer potential designs of autonomous surgical robotic systems.

In this paper, we review literatures on tactile perception technologies in industrial robots and MIS in the last decades to analyze the advantages and feasibility of applying tactile perception methods on MIS, especially the state-of-the-art AI methods on tactile images. Similarly, the features and advantages of tactile sensors varying in sensing modalities are analyzed, together with their applications in MIS.

The remainder of this paper is organized as follows: *Tactile Sensors and Their Applications in MIS* introduced tactile sensors and their applications in MIS. In *Tactile Perception Algorithms in MIS*, tactile perception algorithms in MIS are reviewed. In *Tactile Perception Applications in MIS*, the feasibility of applying tactile perception methods on MIS is analyzed. In *Conclusion*, a summary of the challenges and perspectives hoped for the future with tactile perception in MIS is presented.

TACTILE SENSORS AND THEIR APPLICATIONS IN MIS

Tactile sensors are used to collect tactile data at the contact point between the surgical equipment and tissues. Depending on modalities of tactile signal, various physical properties (e.g., softness and roughness) of a tissue can be extracted from tactile data. Tactile feedback is then provided for surgeons

based on these detected physical properties. In most of the literatures, force feedback is the main form of tactile feedback, and force sensors are the most widely used tactile sensors. Tactile sensors can be categorized into the single-point tactile sensor and the tactile array sensor. In this section, studies on providing force feedback with the above two kinds of tactile sensors are reviewed. Except for force sensors and force feedback, some novel tactile sensors and tactile feedback methods are investigated.

Single-Point Tactile Sensor and Force Feedback

A single-point tactile sensor is usually embedded on the tip of the surgical equipment to confirm the object–sensor contact and detect tactile signals at the contact point. In MIS, force feedback is extremely important to doctors in the consideration of the various consistency of the tissue. The force feedback implies the active force applied to the operators' hands directly where the active force is usually related to the reactive force from the tissue to the tools. Many studies investigated the different application scenarios of force feedback in MIS. We summarized related cases into knotting, insertion, and incision, which will be described in the later paragraphs. After that, we will expound on the importance of force feedback in the abovementioned cases by a series of relevant studies, while the comparison with visual force feedback will also be referred to. We also investigated the development of force feedback in a famous minimally invasive surgical robotic system named da Vinci robot. Finally, the limitation of force feedback in minimally invasive surgery has been given out.

Knotting

In the knotting situation through the laparoscopic procedures, the force feedback indicating the tension of the thread from the tip of the tools is extremely important to guarantee the firmness of the knots but prevent damage to the tissue. To sense the point force feedback from tool tips, load cells are commonly used, as the case in (Song et al., 2009). Moreover, in (Song et al., 2011), a load cell with fiber Bragg grating (FBG) sensors was applied to measure the tension of the thread, where FGB sensors are optical fiber sensors improving the accuracy by encoding the wavelength. Richards et al. utilized the force/torque at the grasper-side to calculate the grasping force (Richards et al., 2000). Fazal and Karsiti decomposed the reactive force happened during the insertion process into three types by a piezoelectric type one-dimensional sensor and mathematical statistics, which were the force generated due to the stiffness of the tissue, the friction force, and the cutting force, thus enabling us to analyze each type of force separately (Fazal and Karsiti, 2009). Mohareri et al. creatively passed the reactive force produced by one hand to another hand and improved the knotting accuracy to 98% (Mohareri et al., 2014).

Incision

Apart from the suture scenario, the incision situation is another indispensable part of minimally invasive surgery. Callaghan and McGrath designed a force-feedback scissor with button load cells attached to the scissor blades to measure the interforce between the blades and the tissue (Callaghan and McGrath, 2007). However, a load cell normally could only sense force from one axial and two moments; therefore, the more complicated design is considered in the later researches. In (Valdastri et al., 2006), an integrated triaxial force sensor was developed and attached to the cutting tool for fetal surgery. A similar design of 3-DOF sensors could be found in (Berkelman et al., 2003). In (Kim et al., 2015), a type of surgical instrument with force sensors of 4 DOFs was developed, which could be applied to measure the normal force and the tangential force from the tip of the tools by capacitive transduction principle.

Palpation

Another conspicuous application scenario of force feedback in minimally invasive surgery is palpation. Since the tumor is always stiffer than the surrounding skin, the pressure intensity on the tumor tends to be obviously larger; therefore, sensors with a single point of contact can detect tumors by palpation (Talasaz and Patel, 2013). Similarly, to detect abnormal masses in the breast, a tactile sensing instrument (TSI) was designed in (Hosseini et al., 2010) and applied in a simulated scenario with a certain detecting route, which was the transverse scan mode. By combining the stress variation curves of each line, users could determine the x- and y-axis coordinates of the abnormal masses. The stress variations of the sensor in the two cases that were operating manually and by a robot showed a similar pattern. Besides, a new tactile sensory system was developed in (Afshari et al., 2011) by combining the displacement sensor and the force sensor to determine the existence and detect the location of kidney stones during laparoscopy. Since the surface stiffness was proportional to the result of the force sensor as well as the displacement sensor, the stiffness could be presented by these two values and depicted by a curving line through the path on the surface of model. Besides, Yip et al. first developed a miniature uniaxial force sensor to do endocardial measurements (Yip et al., 2010). In the research of (Munawar and Fischer, 2016), an elastic spherical proxy regions was designed to sense the forces from various directions.

Necessity of Force Feedback

To support the multiple cases in the above paragraphs, we also investigated the necessity of force feedback in minimally invasive surgery, which was proved and explained in (Morimoto et al., 1997) and (Tholey et al., 2005). The accuracy of the force applied seemed to be improved with the increasing force feedback in (Wagner and Howe, 2007) and (Bell et al., 2007). The reasons could be generalized into two points as in (Mohareri et al., 2014). One is sensing the invisible property such as the stiffness and the texture. Another one is preventing the undesired damage of tissue. Many studies investigated the cases of providing force feedback, visual feedback, visual force feedback (force feedback in the forms of image, sometimes like color bars), and no feedback.

Mahvash et al. put forward the result that providing force feedback generated less error than other cases in the cardiac palpation (Mahvash et al., 2008). A similar result could be found in (Kitagawa et al., 2005). Mohareri et al. found out that the tightening degree in the knitting situation tended to be less uniform with visual feedback and summed up that the visual feedback could compensate part of the force feedback but was entirely not enough while applying the needles and thread (Mohareri et al., 2014). Reiley et al. investigated the practicability of the visual force feedback and concluded that operators without robotic experience could benefit from visual force feedback while practitioners do not as much as their counterparts (Reiley et al., 2008). Similar results were also shown in (Gwilliam et al., 2009) and summarized in (Okamura, 2009). However, visual force feedback could be the better solution in knot-tightening tasks as demonstrated in (Talasaz et al., 2012) and (Talasaz et al., 2017). Later, after (Talasaz et al., 2012), Talasaz and Patel first operated the system with an MIS tactile sensing probe remotely and viewed the feedback through a camera display (Talasaz and Patel, 2013). Besides, Guo et al. applied visual force feedback in vascular interventional surgery and showed great conformity (Guo et al., 2012).

Based on the aforementioned techniques, many operation platforms for minimally invasive surgery have been developed, including Robodoc, Probot, Zeus, and the most recent one named da Vinci (Marohn and Hanly, 2004; Puangmali et al., 2008; Munawar and Fischer, 2016). The da Vinci operation system solved several major limitations in recent minimally invasive surgeries, including the need for hand motion feedback, hand–eye coordination, feeling hands inside the body, expanding the DOF, elimination of surgeon tremor, and variable motion scaling (Guthar and Salisbury, 2000). Many pieces of research were applying based on the da Vinci operation system; however, the force feedback has been added to this system only recently. The examples could be found in (Mahvash et al., 2008) and (Reiley et al., 2008).

Challenges

Force feedback has been very promising for a long while; however, it has also faced some unsolved problems. Although the force feedback provides better tumor localization performance and more precise suture and incision operation with straightforward quantitative measures, it can be somehow time consuming since the measurement from one point to another is low effective in the algorithm level as shown in (Talasaz and Patel, 2013). Apart from this problem, another problem from the force feedback method is the attenuation of the force signal since the surgical tools are always long and stiff. To solve this problem, force amplification is considered as shown in (Song et al., 2009). However, due to the unpredictable disturbance of the tissues, the small disturbing force might also be amplified leading to fatal maloperation. To solve this problem, several actions in the laparoscopic cholecystectomy procedure are described and modeled in spatial coordinates in (Pitakwatchara et al., 2006) to amplify the operation reactive force but remain the disturbing force. Nevertheless, the research remained on the theoretical level without any real model.

Tactile Array Sensor

A tactile array sensor is composed of several single-point tactile sensors arranged according to certain rules. It is usually a flat cuboid with M × N tactile sensing units, where M and N indicate the number of rows and columns of sensing units. In the last decades, tactile data sensed by tactile array sensor was generally displayed as a wave diagram with M × N waveforms, each of which indicates a time-dependent physical quantity obtained by a sensing unit. Recently, with the development of computer version, the methods for processing images have been faster and more accurate, inspired by which tactile data are represented as image sequences, where each sequence represent tactile data over time sensed by a tactile array sensor, and each image pixel represents tactile data sensed by a sensing unit in a certain time. **Figure 1** shows a comparison between a wave diagram and tactile images.

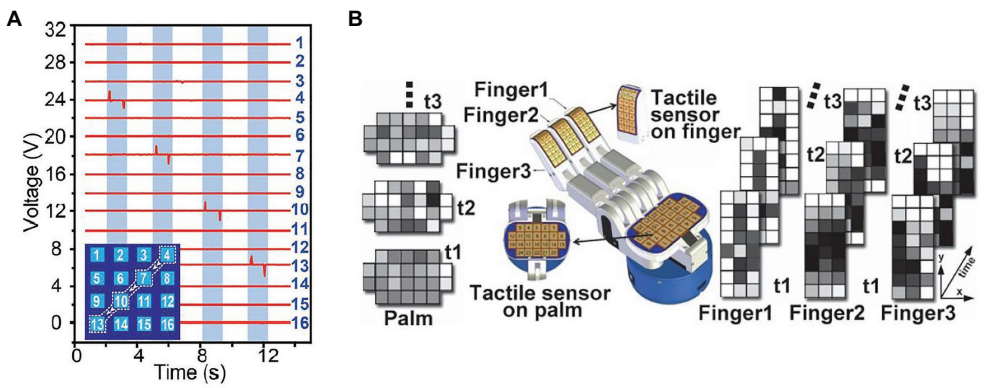


FIGURE 1 | (A) An example of tactile wave diagram, where each waveform indicates voltage sensed by a sensing unit. This diagram shows the sensing result of a case that a capacitive stylus touched the surface of a 4 × 4 tactile array sensor along a path: 4 → 7 → 10 → 13 (Wang et al., 2016). **(B)** An example of tactile image sequences, where each sequence represent tactile data over time sensed by an tactile array sensor, and each image pixel represents tactile data sensed by a sensing unit in a certain time (Cao et al., 2016).

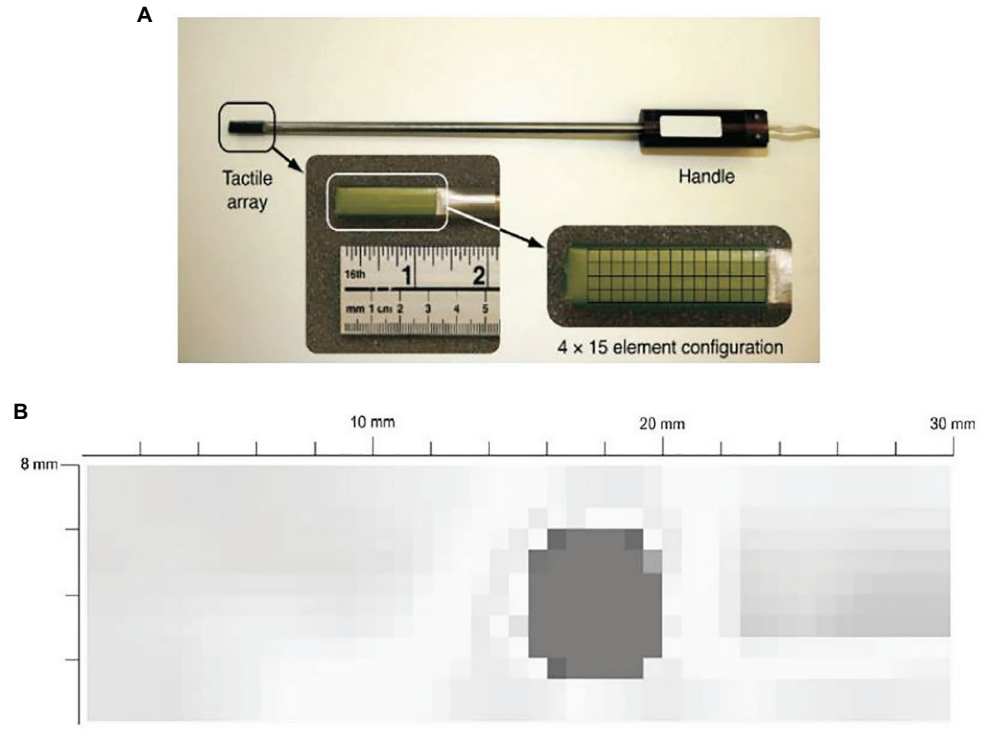


FIGURE 2 | (A) A tactile sensor array with 4 × 15 sensing elements. **(B)** a typical contour map of a tumor obtained from the visualization software (Trejos et al., 2009).

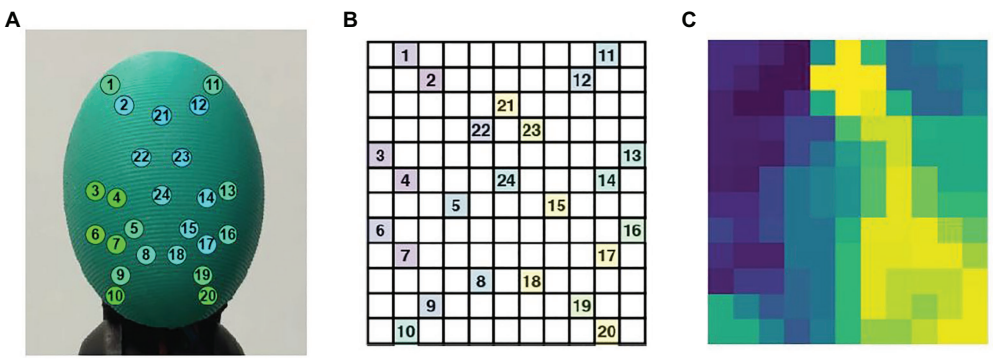


FIGURE 3 | (A) The BioTac sensor with 24 electrodes distributions, **(B)** distribution of the BioTac SP electrodes in a 12 × 11 tactile image, **(C)** result of filling the gaps in the tactile image with the mean value of the eight-closest neighbors (Zapata-Impata et al., 2019).

For example, Trejos et al. developed a TSI that uses a commercially available pressure pad (Trejos et al., 2009). The TSI is shown in **Figure 2A**. The TSI industrial TactArray on this instrument consists of an array with 15 rows and 4 columns of electrodes, which are oriented orthogonally to each other. Each overlapping area created by the row and column electrodes forms a distinct capacitor. The results from the tactile sensor as shown in **Figure 2B** converts the measured voltage values from the capacitive sensor to pressure measurements and displays these results in a color contour map of pressure distributions.

Zapata-Impata et al. used the BioTac SP tactile sensor manufactured by Syntouch (Zapata-Impata et al., 2019). **Figure 3A** shows a representation of the location of the electrodes in the sensor. A tactile image can be created for this 2D array in which the 24 electrodes values e_i are spatially distributed to occupy the image pixels at certain coordinates (i,j). Basically, the tactile image consists of a 12 × 11 matrix in which the 24 electrodes are distributed as shown in **Figure 3B**. **Figure 3C** shows the final tactile image; all the gaps (cells without assigned values) are then filled using the mean value of the eight-closest neighbors.

Wang et al. reported a self-powered, high-resolution, and pressure-sensitive triboelectric sensor matrix (TESM) based on single-electrode triboelectric generators that enable real-time tactile mapping (Wang et al., 2016). **Figure 4A** shows a flexible 16 × 16 pixelated TESM with a resolution of 5 dpi can map single and multipoint tactile stimuli in real time *via* the multichannel data acquisition method while maintaining an excellent pressure sensitivity of 0.06 kPa^{-1} and long-term durability. **Figure 4B** is a schematic of how the sensor matrix images the pressure distribution when a mold in the shape of a "6" is pressed against the top of the TESM.

Compared with single-point tactile sensors, a tactile array sensor can cover a wider area and can capture the tactile information of the target from multiple dimensions, so it can achieve high spatial resolution of touch. Therefore, it is applied to minimally invasive surgery now. For vascular interventional surgery, Guo et al. reported a novel catheter sidewall force tactile sensor array, which is based on a developed robotic catheter operating system with a master–slave structure (Guo et al., 2013). It can detect the force information between the sidewall of the catheter and the blood vessel in detail and transmit the detected force information to the surgeon through the robot catheter system. Besides, to reduce the postoperative pains, Li et al. proposed an original miniature three-dimensional force sensor that can detect the interaction forces during tissue palpation in minimally invasive surgery (Li et al., 2015). In addition, to detect and locate tissue abnormalities, Li et al. presented a novel and high-sensitivity optical tactile sensor array based on fiber Bragg grating (FBG) (Li et al., 2018). Each tactile unit is mainly composed of a spiral elastomer, a suspended optical fiber engraved with an FBG element, and a contact connected with elastomers with threads. Moreover, for tissue palpation, Xie et al. proposed a new type of optical fiber tactile probe, which consists of 3 × 4 tactile sensors (Xie et al., 2013). In this paper, one single camera is employed to capture and detect the light intensity changes of all sensing elements and convert to force information. Finally, for tissue palpation, Roesthuis et al. proposed an experimental bench, which includes a tendon-driven manipulator. A kind of nitinol FBG wire is fabricated, on which 12 FBG sensor arrays are integrated and distributed over four different groups. In closed-loop control, the minimum average tracking error of circular trajectory is 0.67 mm (Roesthuis et al., 2013).

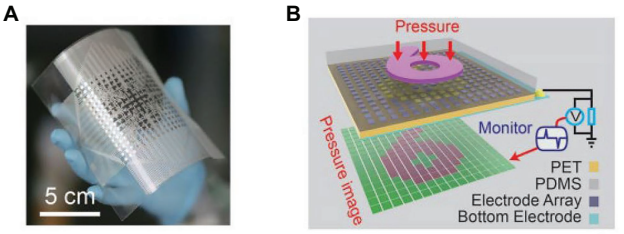


FIGURE 4 | (A) Photograph of a fabricated 16 × 16 TESM with good flexibility. **(B)** Demonstration of the mapping output voltage of the sensor matrix under the pressure of a module in the shape of a "6" (Wang et al., 2016).

Novel Tactile Sensor and Tactile Feedback

Novel Tactile Sensor

In conventional open surgery, surgeons make ample use of their cutaneous senses to differentiate tissue qualities, which can hardly be achieved with force sensors alone, motivating some researchers to expend effort on enabling other sensing modality in MIS. In the last decades, researchers have made an attempt to use other tactile signals to measure properties of tissues. Eklund et al. developed an *in vitro* tissue hardness measurement method using a catheter-type version of piezoelectric vibration sensors (Eklund et al., 1999). Eltaib and Hewit proposed a tactile sensor by attaching a pressure sensor to the end of a sinusoidally driven rod of the tactile probe. The sensor measured both vibration and contact force to detect differences between soft and hard tissues and assist surgeons in detecting abnormal tissues (Eltaib and Hewit, 2000). Baumann et al. presented a method of measuring mechanical tissue impedance by determining resonance frequency with an electromechanic vibrotactile sensor integrated into an operating instrument (Baumann et al., 2001). Chuang et al. reported a miniature piezoelectric hardness sensor mounting on an endoscope to detect submucosal tumors (Chuang et al., 2015). Kim et al. fabricated sensorized surgical forceps with five-degree-of-freedom (5-DOF) force/torque (FIT) sensing capability (Kim et al., 2018). A summary of the representative tactile sensors and their applications is presented in **Table 1**.

TABLE 1 | Representative tactile sensors and their applications.

Literature	Application	Sensor type
Eklund et al., 1999	*In vitro* tissue hardness measurement	Catheter type version of piezoelectric vibration sensor
Baumann et al., 2001	Measuring mechanical tissue impedance	Electromechanic vibrotactile sensor
Song et al., 2009	Measuring the tension of the thread	Single-point FBG force sensor
Trejos et al., 2009	Assessing the feasibility of using the tactile sensing instrument under robotic control to locate underlying tumors.	Six-DOF force/torque sensor array
Hosseini et al., 2010	Detecting abnormal masses in the breast	Tactile probe
Afshari et al., 2011	Detecting the location of kidney stones	Force sensor and displacement sensor
Mohareri et al., 2014	Passing the reactive force produced by one hand to another hand for bimanual robot-assisted surgery	Single-point force sensor
Chuang et al., 2015	Detecting submucosal tumors	Piezoelectric hardness sensor
Kim et al., 2015	Measuring the normal force and the tangential force from the tip of the tools in the incision situation	Four-DOF force sensor
Wang et al., 2016	Real-time tactile mapping	Triboelectric sensor array
Li et al., 2018	Detecting and locating tissue abnormalities	Optical tactile sensor array

Tactile Feedback

Some novel tactile display systems were developed to provide feedback for surgeons based on various tactile information. Schostek et al. proposed a tactile sensor, integrated into a laparoscopic grasper jaw, to obtain information about shape and consistency of tissue structures (Schostek et al., 2006). The tactile data were wirelessly transferred *via* Bluetooth and graphically displayed to the surgeon. However, tissue exploration time was longer compared to a conventional grasper. Prasad et al. presented an audio display system to relay force information to the surgeon, but continual noise in an operating room setting remained a problem (Prasad et al., 2003). Fischer et al. developed a system that displayed oxygenation values to surgeons. They simultaneously used force sensors and oxygenation sensors to measure tissue interaction forces and tissue oxygenation next to translational forces, when tissue oxygenation decreases below a certain value, trauma will occur (Fischer et al., 2006). Pacchierotti et al. reported a cutaneous feedback solution on an da Vinci surgical robot. They proposed a model-free algorithm based on look-up tables to map the contact deformations, dc pressure, and vibrations to input commands for the cutaneous device's motors. A custom cutaneous display was attached to the master controller to reproduce the tactile sensations by continually moving, tilting, and vibrating a flat plate at the operator's fingertip (Pacchierotti et al., 2016).

TACTILE PERCEPTION ALGORITHMS IN MIS

Recent researches on tactile perception algorithms are focused on the tactile array sensor. With the tactile array sensor, we can collect an $M \times N$ tactile image, where each tactile element is treated as an image pixel. From tactile images that tactile sensor arrays acquired, features are extracted, such as statistical features, vision feature descriptors, PCA-based features, and self-organizing features. These features are usually processed by AI methods like clustering, SVM, and deep neural networks to obtain tactile information. After training the algorithm, we can use it to assist doctors in minimally invasive surgery. There are a lot of scenarios where the algorithm can be used. We summarized related cases into wall following, shape recognition, stable scraping, and hardness detection.

Wall Following

To perform wall following, Fagogenis et al. designed an image classifier, which is based on machine learning and can distinguish between blood (no contact) or ventricular wall tissue and the bioprosthetic aortic valve (Fagogenis et al., 2019). The algorithm used the bag-of-words method to group tactile images, which is based on the number of occurrences of specific features of interest. During training, the algorithm can select features that were of interest and the relationship between their number and the tactile image class. For training, they used OpenCV to detect features in a set of training images based on manually labeled. Then, the detected features are mathematically encoded with LUCID descriptors to achieve efficient online computation. In order to reduce the number of features, they used clustering (k-mean) to identify the optimal feature representatives. The resulting cluster centers were the representative features used for the rest of the training and for runtime image classification. After determining the representative feature set, they traversed the training data for the second time and constructed the feature histogram for each image by calculating the number of times each representative feature appeared in the image. The last step was to train an SVM classifier, which learned the relationship between the feature histogram and the corresponding classes. Using the trained algorithm, we first detected the features and calculated the corresponding LUCID descriptors and then classified the images. Then, these features were matched to the nearest representative features, and the resulting feature histogram was constructed. Based on the histogram, the SVM classifier is used to predict the tissue-based contact state. They used a small group of training tactile images (~2,000 images) with training taking just a few minutes (~4 min) and achieved good results. Because image classification took 1 ms, our haptic vision system estimated contact state based on the camera's frame rate (~45 frames/s). The accuracy of contact classification algorithm is 97% (tested on 7,000 images not used for training) with type I error (false positive) of 3.7% and type II (false negative) of 2.3%.

Shape Recognition

To recognize the shape of an object, Liu et al. proposed a new algorithm to identify the shape of an object by tactile pressure images, which can distinguish the contact shapes between a group of low-resolution pressure maps (Liu et al., 2012). The algorithm can be divided into four steps. The first step of the algorithm is "Data extraction." Data extraction normalizes the strongly correlated tactile images sequence into a single map to save computational cost and reduce the disturbances from signal noise. The second step is "Preprocessing." It consists of several subalgorithms to prepare the information for its latter "Feature extraction." Preprocessing is essential not only to prepare the information for further steps but also to enhance the interests of tactile images. The third step is "Feature extraction." In this step, the tactile image is transformed into a 512-feature vector, and the extracted features are used to train the developed neural network for object shape recognition. These features are not affected by occlusion, position, or scale, as well as image size, resolution, and number of frames. All these characteristics make the algorithm robust and effective. Finally, a three-layer neural network is developed to train, validate, and test the efficiency and success rate of the algorithm. It is trained to use the features extracted at the previous stage as a classifier. **Figure 5** shows a diagram of the three-layer neural network for object classification. Through the experimental study, it was found that using four different contact shapes to test, the average classification accuracy reached 91.07%. The shape recognition algorithm based on the feature extraction has strong robustness and effectiveness in distinguishing different target shapes. It can be directly applied to minimally invasive surgery to identify the shape of the contact site and determine whether the tissue is abnormal, which is convenient for doctors to detect abnormal tissues with abnormal shapes in time.

Stable Scraping

To judge the stability of the grip, Zapata-Impata et al. proposed a spatiotemporal tactile features learning method based on convolutional long short-term memory (ConvLSTM) (Zapata-Impata et al., 2019). This method preprocessed the tactile readings and fed to a ConvLSTM that learns to detect multiple types of slip directions with just 50 ms of data. The architecture of the ConvLSTM network is shown in **Figure 6**. For preprocessing data, this method used a sensor with 24-electrode distributions to obtain tactile images. In more detail, the sensor uses these electrodes to record signals from four emitters and measure the impedance in the fluid between them and the elastic skin of the sensor. The fluid moves while contact is experienced by these sensors, thus affecting the measurements made by the electrodes. The whole method used four object sets, containing a total of 11 different objects, and was used to capture a new tactile dataset, recording seven different types of slip directions: north, slip south, slip east, slip west, slip clockwise, slip anticlockwise, or touch. Basically, the method created the ConvLSTM learns spatial features from pictures while simultaneously learning temporal ones. In the process of creating ConvLSTM, this method studied how the performance of the ConvLSTM changes depending on several parameters: the number of ConvLSTM layers, the size of the convolutional filters, and the number of filters inside each ConvLSTM layer. Finally, according to the experimental results, the network structures of five ConvLSTM layers, 3 × 3 filters, and ConvLSTM layers with 32 filters are selected to focus more attention on the low-level details in the tactile image and get better accuracy. For feature learning in time series, this method only needs three to five continuous tactile images, and the network can accurately learn to detect the sliding direction. In the task of detecting these seven states on seen objects, the system achieved an accuracy rate of 99%. Even if the ConvLSTM network was sensitive to new objects, during the robustness experiments, its performance dropped to an accuracy rate of 82.56% in the case of new objects with familiar properties (solids set) and an accuracy rate of 73.54 and 70.94% for stranger sets like the textures and small sets. The spatiotemporal tactile features learning method can be directly applied to minimally invasive surgery to improve the stability of tissue detect/mass grasp. However, at present, the single-point sensor is used to judge the grasping stability, and the judgment of the slip direction is only based on the single-point tactile characteristics. If the array tactile map is used, the regional feature information can be considered in the process of judgment to improve the stability of grasping. Therefore, it is very promising to apply this algorithm to minimally invasive surgery.

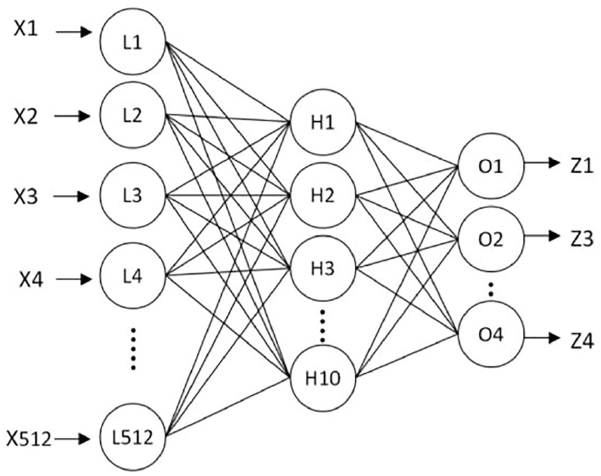


FIGURE 5 | The three-layer neural network for object classification (Liu et al., 2012).

Hardness Detection

To detect hardness, Yuan et al. designed a deep learning method that can estimate shape-independent hardness (Yuan et al., 2017). The algorithm, the convolutional neural network (CNN) feature layer of VGG network, is used to extract the physical signs of the tactile image, and a feature sequence is generated and input to LSTM to evaluate the softness and hardness of the sample. The algorithm can estimate the hardness of objects with different shapes and hardness ranging from 8 to 87 in Shore 00 scale.

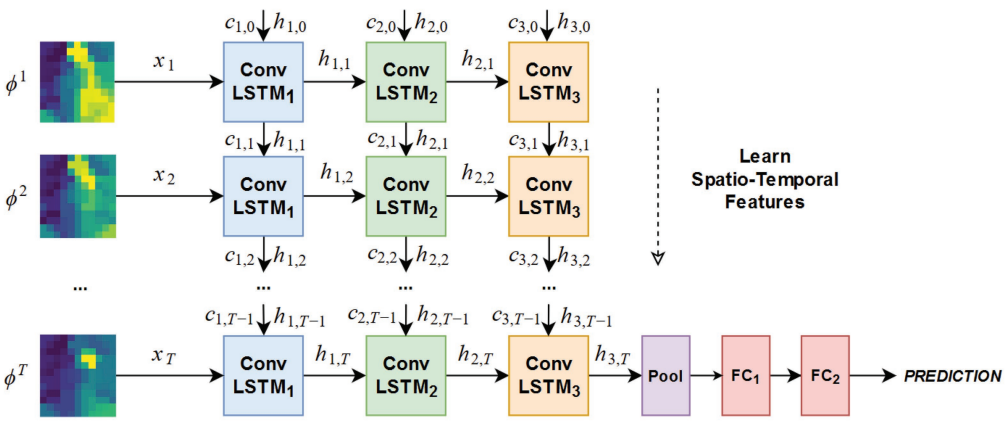


FIGURE 6 | Architecture of the ConvLSTM network tested in the experimentation (Zapata-Impata et al., 2019).

In minimally invasive surgery, tissue hardness detection is very important. Tumor detection is a very good example: some solid tumors are harder than the surrounding tissue, and their existence can also be obtained through tactile feedback to determine the location of resection and increase the success rate of surgery. Xie et al. proposed a method based on pixel calculation to measure the normal force and its distribution in the sensor area, to judge the hardness range of the area and determine the abnormal structure (Xie et al., 2013). As shown in **Figure 7**, in this method, the tactile image data of different brightness under different forces are captured by an optical fiber tactile sensor containing 3 × 4 sensing elements, and the tactile image is divided into 12 different regions in turn. By calculating the pixel values of each sensing region, and according to the predetermined linear relationship, the magnitude of the force applied in the region is obtained. The sensor outputs responses after palpation in two different areas; **Figure 7C** shows areas including the nodule, while **Figure 7D** shows areas that do not. From the result, outputs of each sensing element in the nodule-free area vary mostly in the range of 0–0.4 N. while in the nodule-embedded area, outputs of the sensing elements in contact with the nodule exceed the value of 0.8 N. The location of the nodules can be seen more clearly by subtracting **Figure 7C** of **Figure 7D** from **Figure 7E**. This method determines the tissue lump according to the pressure distribution map; although it is effective in some ways, the method based on linear fitting has a risk of producing large errors, and the definition range of hardness is single. If the image processing algorithm is used, the above two defects can be improved. Therefore, in the tissue hardness detection method, the method based on array tactile image processing is worth studying.

TACTILE PERCEPTION APPLICATIONS IN MIS

In the minimally invasive surgery, given the very small holes for the tools to operate, the capacity to feel tends to be limited as discussed in (Eltaib and Hewit, 2003). In line with that, Dargahi and Najarian summarized four categories of properties that were usually considered to be important in minimally invasive surgery including the force, position and size, hardness/softness, and roughness and texture (Dargahi and Najarian, 2004). However, with the help of tactile perception, the general performance is liable to be improved. To show the feasibility of the assistance provided by the tactile perception in the minimally invasive surgery, we will discuss the assistance provided by tactile perception catering for each property, respectively, after which the general discussion will be given.

Obtaining Tactile Properties of Tissues

Measuring the acting and reactive force from tissues could be applied in many cases such as controlling the surgical knife on the liver tissue (Chanthasopeephan et al., 2003), measuring

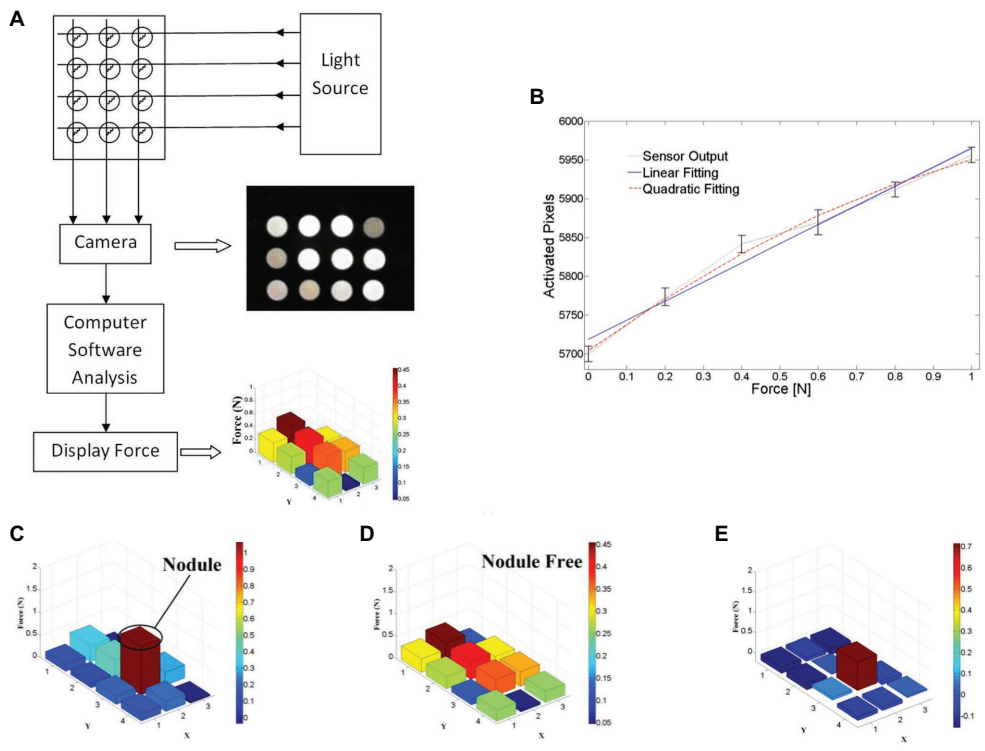


FIGURE 7 | (A) Schematic design of proposed tactile sensor using camera, **(B)** measured output responses of sensing element 1 to the normal force applied, **(C)** test results on nodule area, **(D)** test results on nodule-free area, and **(E)** effective stiffness distribution map by compare test result on two areas (Xie et al., 2013).

the tension of the thread while knotting (Song et al., 2009), modeling needle insertion force (Fazal and Karsiti, 2009), and differentiating between tissue samples in the scissoring process (Callaghan and McGrath, 2007). The doctors usually rely on the magnitude of force feedback to estimate when to stop every single shearing or insertion operation. For example, He et al. designed a 3-DOF force sensing pick instrument applied in the retinal microsurgery with fiber optic sensors placed at the distal tip of the surgical instrument. To realize multiple degrees of freedom, a linear model and second-order Bernstein polynomial were used to distinct forces in different directions (He et al., 2014). Besides, many previous studies like (Mohareri et al., 2014) show that, with the force-feedback data, doctors tend to make each separate operation more uniform, such as knots with similar thread tension. Nevertheless, the sensing directions and the accuracy are still of the top concern when researchers strive to improve the overall performance of the assistance of force.

The assistance in position and size is usually for tumor localization (Hosseini et al., 2006; Perri et al., 2010). For example, Liu et al. measured the indentation depth to detect the abnormal part of the tissue (Liu et al., 2010). Lederman et al. used rigid fingertip sheaths to locate the 3D mass (Lederman and Klatzky, 1999). Afshari et al. utilized the stress distribution to determine the stone inside the kidney (Afshari et al., 2010). Most of the time, the position and size of mass are reflected by the surging magnitude of rigidity sensed from the tools during palpation or the image constructed by an array of force sensors. In (Xie et al., 2013), a 3 × 4 sensing array was designed to detect the force distribution, and the doctors could be provided with the visualized data on which area tended to be stiffer. In the evaluation test on a lamb kidney with nodules embedded, the design presented a very effective performance. The doctors are more likely to rely on their experience to estimate the position and size based on the force feedback. **Figure 8** shows an example of the mass localization utilizing an array of force sensors.

When calculating the hardness/softness, the sensors are usually placed on endoscopic graspers (Najarian et al., 2006). In many cases, tissue hardness could also be utilized to locate the mass in palpation. For instance, Ju et al. relied on the sensors on the catheter robot to locate the mass (Ju et al., 2018, 2019). Moreover, Kalantari et al. measured the relative hardness/softness of the tissue to sense various types of cardiac tissues while performing mitral valve annuloplasty (Kalantari et al., 2011). In (Yip et al., 2010), a creative uniaxial force sensor based on fiber optic transduction was developed, which could detect very small forces but show few root mean square (RMS) errors. In the designing process, properties of waterproof, electrical passivity, and material constraints were especially considered so that the instrument could perfectly meet the requirement of operating in the cardiac environment. However, there is no certain threshold to determine hard or soft by machine, which means the subjective judgment from doctors is indispensable.

Roughness and texture are the fourth groups of properties that can assist doctors in MIS. To measure roughness and texture, the sensors are usually placed on endoscopic graspers (Bonakdar and Narayanan, 2010), like laparoscopic graspers (Dargahi and Payandeh, 1998; Dargahi, 2002; Zemiti et al., 2006;

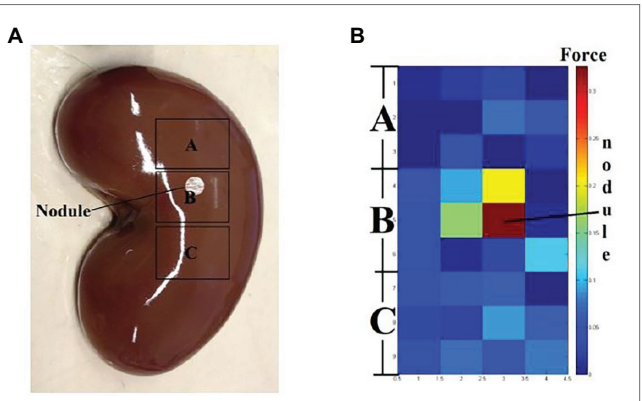


FIGURE 8 | (A) A kidney with invisible nodule buried in area B. **(B)** The sensing result by palpating on three areas. Various color blocks indicating different force values (Xie et al., 2013).

Lee et al., 2016), which could be applied to cholecystectomy (Richards et al., 2000) and Nissen fundoplication (Rosen et al., 2001), and could also measure viscoelastic properties of tissues (Narayanan et al., 2006). In (Bicchi et al., 1996), tissue elastic properties were measured. For instance, in (Dargahi, 2002), a polyvinylidene fluoride tactile sensor was designed to measure the compliance and roughness of tissues. The principle was to measure the relative deformation when the tissue contacted with the sensor surface. However, it is never the best choice to detect all the tissue surface points by point, which indicates the large time consumed. For wiser utilization, subjective judgment by humans should also be taken into account.

AI-Based Tactile Perception Applications in MIS

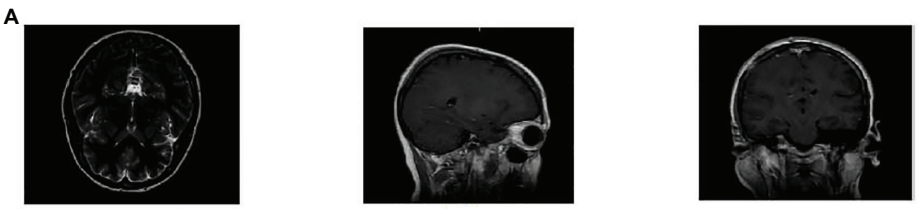
In line with the aforementioned cases, some researchers have devoted to the intelligent algorithms that can diminish the participation of subjectivity. Many of them have been described in *Tactile Perception Algorithms in MIS*. In this subsection, we will introduce more AI-based tactile perception technologies that were proven to be effective in MIS. Beasley and Howe used the pulsatile pressure variation from force sensors to find the artery through a signal processing algorithm and applied an adaptive extrapolation algorithm to generate the ultimate position prediction. The rough idea of adaptive extrapolation was applying 15 sensing samples and linear regression to fit the predicted arteries. It has been tested on the ZEUS Surgical Robot System and resulted in a < 2-nm mean error (Beasley and Howe, 2002).

Sadeghi-Goughari et al. introduced a new minimally invasive diagnosis technique named intraoperative thermal imaging (ITT) based on artificial tactile sensing (ATS) technology and artificial neural networks (ANNs) (Sadeghi-Goughari et al., 2016). In this study, a forward analysis and an inverse analysis based on ANN were proposed to estimate features including temperature and depth of a tumor using a brain thermogram. The brain thermogram is shown in **Figure 9A**. This work involved the forward analysis of heat conduction in cancerous brain tissue by employing a finite element method (FEM). Then, a

three-layer feed-forward neural network (FFNN) with back propagation learning algorithm was developed to estimate related features of a tumor. Parameters of the proposed FFNN are shown in **Figure 9B**. The inputs of FFNN are thermal parameters extracted from tissue surface temperature profiles. Training of the ANN was performed by a backpropagation algorithm. By comparing estimated values of tumor features and expected values, potential brain tissue abnormalities were detected, which greatly facilitate the task of the neurosurgeon during MIS.

Zhao et al. proposed a tracking-by-detection framework of surgical instruments in MIS (Zhao et al., 2017). As shown in **Figure 10A**, the operation of conventional MIS instruments can be subdivided into shaft portions and end effector. In the proposed method, the shaft portion was described by line features through the random sample consensus (RANSAC) scheme, and the end effector was depicted by some special image features based on deep learning through a well-trained CNN. With camera parameters and insertion points, a tracking method was proposed to estimate the 3D position and orientation of the instruments. As shown in **Figure 10B**, the scanning range was restricted to a sector area with the symmetry axis $\tilde{L}^{(i-1)}$, where $\tilde{I}$ is the image and s is an arbitrary scale. For the current frame i, the bounding box $(\tilde{p},s)$ slid along the symmetry axis $\tilde{L}^{(i)}$ obtained by shaft detection. The parameter $\tilde{p}$ is the center of the bounding box. The image in the bounding box at every sliding step with scale s was resized to 101 × 101 and then used as an input for the trained CNN. The highest score of the CNN positive output corresponds to the bounding box $\left(\tilde{p}^i,s^i\right)$, where $\tilde{p}^i$ is treated as the imaged tip position direction of the current frame i. **Figure 10C** shows the selected frames from the tracking procedure of the proposed method. However, compared with those in the *ex vivo* test, the 2D measurement error in the *in vivo* test was at least 2.5 pixels. When the respective 2D tracking by the proposed method was applied to each frame with the CNN-based detection of instruments, the insufficient illumination of the image part (end effector) accounted for drifted tracking results in some frames (see **Figure 10D**), which is the main reason why the *in vivo* test has higher 2D measurement errors. This issue can be resolved by adding samples of *in vivo* sequences into the training database.

Lee and Won presented a novel method to estimate the stiffness and geometric information of a tissue inclusion (Lee and Won, 2013). The estimation was performed based on the tactile data obtained on the tissue surface. To obtain the tactile data, the author developed an optical tactile sensation imaging system (TSIS). The TSIS obtained tactile images with maximum pixel values, total pixel values, and deformation areas. These parameters were used to estimate the size, depth, and elasticity of the embedded lesions. The proposed method consisted of a forward algorithm and an inversion algorithm. The forward algorithm was designed to predict maximum deformation, total deformation, and deformation areas based on the parameters including size, depth, and modulus of the tissue inclusion. In the inversion algorithm, tactile parameters obtained from the TSIS and simulated values from the forward algorithm were used to estimate the size, depth, and modulus of the embedded lesion. **Figure 11A** describes a cross-section of an idealized breast mode. **Figure 11B** shows the sensing probe of TSIS modeled on top of the breast tissue. When the TSIS compressed against the tissue surface containing a stiff tissue inclusion, it produced different parameters: size d, depth h, and Young's modulus E. The FEM in the forward algorithm quantified deformation as the maximum deformation O^1_{FEM} (the largest vertical displacement of FEM elements of sensing probe from the nondeformed position), the total deformation O^2_{FEM} (displacement summation of FEM elements of sensing probe from the nondeformed position), and the deformation area O^3_{FEM} (the projected area of the deformed surface of the sensing probe), as shown in **Figure 11C**. The tactile data are necessary to relate FEM tactile data $\left(O^1_{\mathrm{FEM}},O^2_{\mathrm{FEM}},O^3_{\mathrm{FEM}}\right)$ and TSIS tactile data $\left(O^1_{\mathrm{TSIS}},O^2_{\mathrm{TSIS}},O^3_{\mathrm{TSIS}}\right)$. The definitions of TSIS tactile data are as follows: The maximum pixel value O^1_{TSIS} is defined as the pixel value in the centroid of the tactile data. The total pixel value O^2_{TSIS} is defined as the summation of pixel values in the tactile data. The deformation area of pixel O^3_{TSIS} is defined as the number of pixels greater than the specific threshold value in the tactile data. The inversion algorithm was used to estimate $\left(\hat{d},\hat{h},\hat{E}\right)$ using the determined $\left(O^1_{\mathrm{TSIS}},O^2_{\mathrm{TSIS}},O^3_{\mathrm{TSIS}}\right)$. In this method, the multilayered ANN was considered as an inversion algorithm, as shown in **Figure 11D**.



Training function	Performance function	Number of inputs	Number of out puts	Transfer Function	Number of neurons in 1st layer	Number of neurons in 2nd layer	Number of neurons in 3rd layer	Number of epochs
Trainbr	MSE	5	2	Tansig	3	5	2	1000

FIGURE 9 | (A) Examples of the brain thermogram. **(B)** Parameters of artificial neural network (ANN; Sadeghi-Goughari et al., 2016).

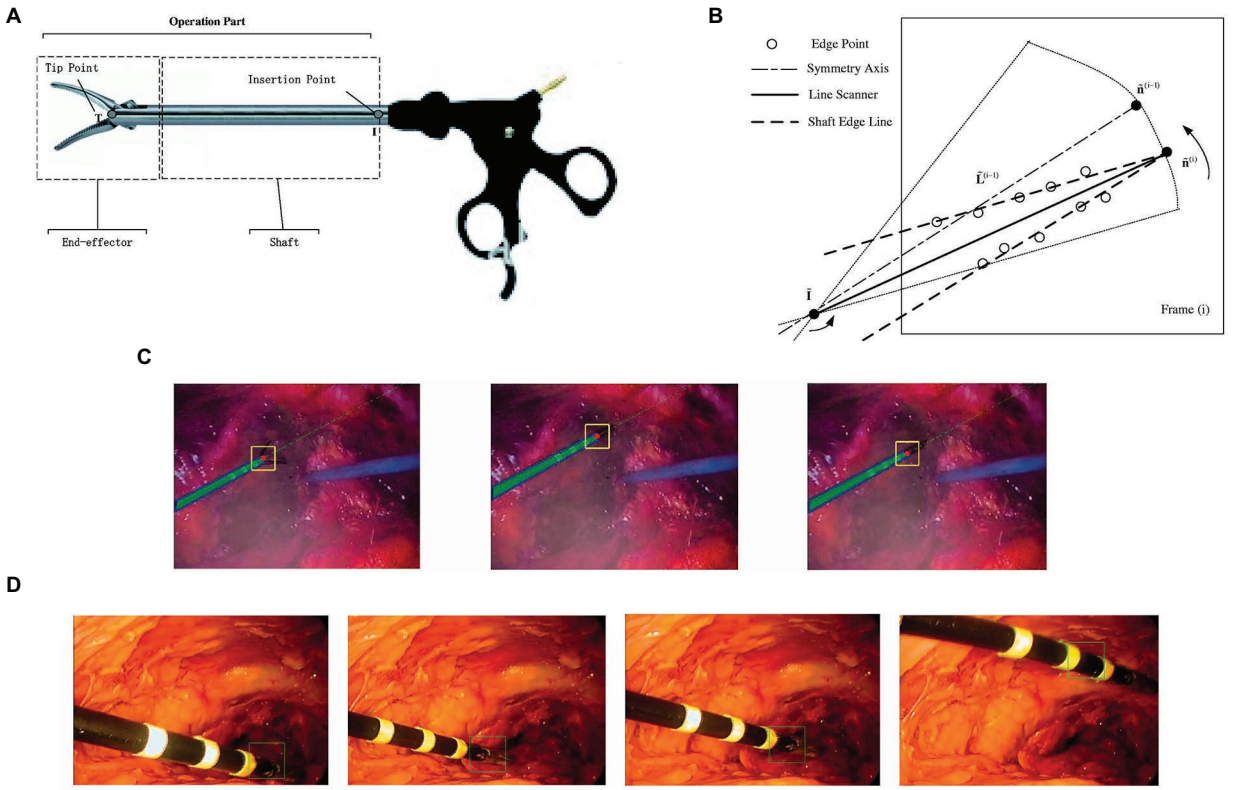


FIGURE 10 | (A) The operation part of minimally invasive surgery (MIS) instrument: end effector and shaft portion; **(B)** line scanner application for detection of shaft edge lines and shaft image direction estimation; **(C)** selected frames of the instrument tracking and detection: the red circles are the tracked end-effector tip position, and the green dashed line is the shaft symmetry axis; **(D)** example frames of *in vivo* sequences with the end-effector positions shown by squares (Zhao et al., 2017).

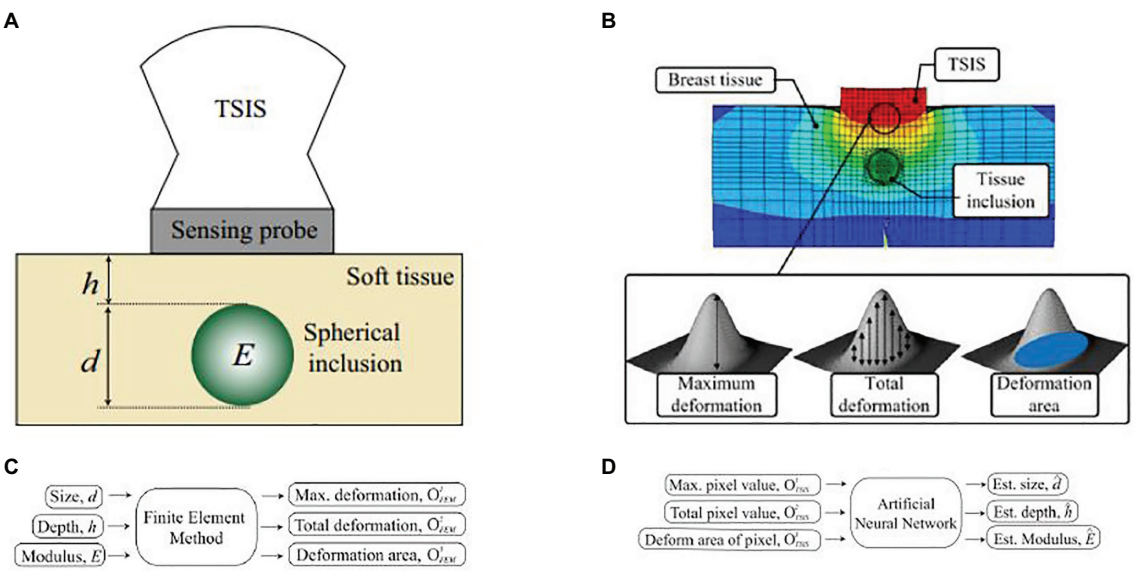


FIGURE 11 | (A) A cross-section of an idealized breast model for estimating inclusion parameters. The tissue inclusion has three parameters: size *d*, depth *h*, and Young's modulus *E*. **(B)** Finite element method (FEM) model of an idealized breast tissue model. The sensing probe of tactile sensation imaging system (TSIS) is modeled on top of the breast tissue model. **(C)** The forward algorithm. **(D)** The inversion algorithm (Lee and Won, 2013).

Despite these methods, algorithms design of tactile perception in the minimally invasive surgery is still a new subject, with related little research, although we hold the opinion that this field could be considerably promising due to the application need.

CONCLUSION

Minimally invasive surgery has been the preferred surgery approach owing to its advantages over conventional open surgery. Tactile information has been proven effective to improve surgeons' performance, while most reviews for MIS were only focusing on force sensors and force feedback, neglecting other tactile information. In this paper, we reported tactile sensors, tactile perception algorithms, and tactile perception applications for MIS. These include a description of various tactile sensors and feedbacks not limited to force sensors and force feedback, the state-of-the-art and novel machine learning algorithms in tactile images for tactile perception in MIS, and potential tactile perception applications for MIS, especially for detecting tissue properties. Finally, this review contains some of the limitations and challenges of each technical aspect.

An emerging research and development trend in the literature is the fusion of various tactile information. Utilizing force information alone has met challenges, including low effectiveness in the algorithm level, the attenuation of the force signal, and amplified disturbing force. Therefore, some studies aimed to develop hybrid sensors employing more than merely one sensing principle to measure one or multiple physical stimuli (e.g., force, slippage, stiffness, etc.) to obtain more robust measurements of physical stimuli and cover wider working environments. With the development of tactile sensors of various sensing modalities, some novel tactile feedback systems were reported (e.g., graphical display system, audio display system, etc.). Some researchers attempted to obtain more tactile information at the algorithm level. Inspired by computer vision technologies, some researchers reported machine learning algorithms for obtaining more than merely one kind of tactile information from tactile images, where a tactile element is treated as an image pixel. Tactile perception algorithms design in MIS is still a new subject, with related little research; while considering the high accuracy, high robustness, and excellent real-time performance of machine learning algorithms, we hold the opinion that this field could be considerably promising due to the application need.

AUTHOR CONTRIBUTIONS

All authors researched the literature, drafted, and wrote the review article, and approved the submitted version.

FUNDING

This review was supported by grants from the Science and Technology Bureau of Ningbo National High-Tech Zone (Program for Development and Industrialization of Intelligent Assistant Robot), the Industrial Internet Innovation and Development Project in 2018 (Program for Industrial Internet platform Testbed for Network Collaborative Manufacturing), and the Industrial Internet Innovation and Development Project in 2020 (Program for Industrial Internet Platform Application Innovation Promotion Center).

REFERENCES

Abushagur, A. A. G., Arsad, N., Reaz, M. I., and Bakar, A. A. A. (2014). Advances in bio-tactile sensors for minimally invasive surgery using the fibre bragg grating force sensor technique: a survey. *Sensors* 14, 6633–6665. doi: 10.3390/s140406633

Afshari, E., Najarian, S., and Simforoosh, N. (2010). Application of artificial tactile sensing approach in kidney-stone-removal laparoscopy. *Biomed. Mater. Eng.* 20, 261–267. doi: 10.3233/BME-2010-0640

Afshari, E., Najarian, S., Simforoosh, N., and Farkoush, S. H. (2011). Design and fabrication of a novel tactile sensory system applicable in artificial palpation. *Minim. Invasive Ther. Allied Technol.* 20, 22–29. doi: 10.3109/13645706.2010.518739

Ahmadi, R., Packirisamy, M., Dargahi, J., and Cecere, R. (2012). Discretely loaded beam-type optical fiber tactile sensor for tissue manipulation and palpation in minimally invasive robotic surgery. *IEEE Sens. J.* 12, 22–32. doi: 10.1109/jsen.2011.2113394

Al-Handarish, Y., Omisore, O. M., Igbe, T., Han, S., Li, H., Du, W., et al. (2020). A survey of tactile-sensing systems and their applications in biomedical engineering. *Adv. Mater. Sci. Eng.* 2020, 1–17. doi: 10.1155/2020/4047937

Bandari, N., Dargahi, J., and Packirisamy, M. (2020). Tactile sensors for minimally invasive surgery: a review of the state-of-the-art, applications, and perspectives. *IEEE Access* 8, 7682–7708. doi: 10.1109/access.2019.2962636

Baumann, I., Plinkert, P. K., Kunert, W., and Buess, G. F. (2001). Vibrotactile characteristics of different tissues in endoscopic otolaryngologic surgery - in vivo and ex vivo measurements. *Minim. Invasive Ther. Allied Technol.* 10, 323–327. doi: 10.1080/136457001753337627

Beasley, R. A., and Howe, R. D. (2002). "Tactile tracking of arteries in robotic surgery" in *2002 IEEE International Conference on Robotics and Automation, Vols I-Iv, Proceedings*; May 11-15, 2002.

Bell, A. K., and Cao, C. G. L. (2007). "Effects of artificial force feedback in laparoscopic surgery training simulators" in *2007 Ieee International Conference on Systems, Man and Cybernetics, Vols 1–8*; October 07-10, 2007; 564–568.

Berkelman, P. J., Whitcomb, L. L., Taylor, R. H., and Jensen, P. (2003). A miniature microsurgical instrument tip force sensor for enhanced force feedback during robot-assisted manipulation. *IEEE Trans. Robot. Autom.* 19, 917–922. doi: 10.1109/tra.2003.817526

Bicchi, A., Canepa, G., de rossi, D., Iacconi, P., and Scilingo, E. P. (1996). "A sensor-based minimally invasive surgery tool for detecting tissue elastic properties" in *Proceedings of IEEE International Conference on Robotics and Automation*; April 22-28, 1996; Minneapolis, MN, USA, USA: IEEE.

Bonakdar, A., and Narayanan, N. (2010). Determination of tissue properties using microfabricated piezoelectric tactile sensor during minimally invasive surgery. *Sens. Rev.* 30, 233–241. doi: 10.1108/02602281011051425

Callaghan, D. J., and McGrath, M. M. (2007). "A Force measurement evaluation tool for telerobotic cutting applications: development of an effective characterization platform," in *Proceedings of World Academy of Science, Engineering and Technology, Vol 25*, ed. C. Ardil, 274–280.

Cao, L., Kotagiri, R., Sun, F., Li, H., Huang, W., Aye, Z. M. M., et al. (2016). "Efficient spatio-temporal tactile object recognition with randomized tiling convolutional networks in a hierarchical fusion strategy" in *Thirtieth Aaai Conference on Artificial Intelligence*; Febraury 12-17, 2016; 3337–3345.

Chanthasopeephan, T., Desai, J. P., and Lau, A. C. W. (2003). Measuring forces in liver cutting: new equipment and experimental results. *Ann. Biomed. Eng.* 31, 1372–1382. doi: 10.1114/1.1624601

Chuang, C. H., Li, T. H., Chou, I. C., and Teng, Y. J. (2015). "Piezoelectric tactile sensor for submucosal tumor hardness detection in endoscopy" in *2015 Transducers –2015 18th International Conference on Solid-State Sensors, Actuators and Microsystems*; June 21-25, 2015.

Dargahi, J. (2002). An endoscopic and robotic tooth-like compliance and roughness tactile sensor. *J. Mech. Des.* 124, 576–582. doi: 10.1115/1.1471531

Dargahi, J., and Najarian, S. (2004). Human tactile perception as a standard for artificial tactile sensing—a review. *Int. J. Med. Robot.* 1, 23–35. doi: 10.1002/rcs.3

Dargahi, J., and Payandeh, S. (1998). "Surface texture measurement by combining signals from two sensing elements of a piezoelectric tactile sensor" in *Sensor fusion: Architectures, algorithms, and applications ii.* ed. B. V. Dasarathy (Orlando, FL, USA). 122–128.

Eklund, A., Bergh, A., and Lindahl, O. A. (1999). A catheter tactile sensor for measuring hardness of soft tissue: measurement in a silicone model and in an in vitro human prostate model. *Med. Biol. Eng. Comput.* 37, 618–624. doi: 10.1007/bf02513357

Eltaib, M., and Hewit, J. R. (2000). A tactile sensor for minimal access surgery applications. *IFAC Proc. Vol.* 33, 505–508. doi: 10.1016/S1474-6670(17)39194-2

Eltaib, M. E. H., and Hewit, J. R. (2003). Tactile sensing technology for minimal access surgery - a review. *Mechatronics* 13, 1163–1177. doi: 10.1016/S0957-4158(03)00048-5

Fagogenis, G., Mencattelli, M., Machaidze, Z., Rosa, B., Price, K., Wu, F., et al. (2019). Autonomous robotic intracardiac catheter navigation using haptic vision. *Sci. Robot.* 4:eaaw1977. doi: 10.1126/scirobotics.aaw1977

Fazal, I., and Karsiti, M. N. (2009). "Needle insertion simulation forces v/s experimental forces for haptic feedback device" in *2009 IEEE Student Conference on Research and Development: Scored 2009, Proceedings*; November 16-18, 2009.

Fischer, G. S., Akinbiyi, T., Saha, S., Zand, J., Talamini, M., Marohn, M., et al. (2006). "Ischemia and force sensing surgical instruments for augmenting available surgeon information" in *2006 1st IEEE Ras-Embs International Conference on Biomedical Robotics and Biomechatronics, Vols 1–3*; Febraury 20-22, 2006; 989.

Guo, J., Guo, S., Wang, P., Wei, W., and Wang, Y. (2013). "A Novel type of catheter sidewall tactile sensor array for vascular interventional surgery" in *2013 Icme International Conference on Complex Medical Engineering*; May 25-28, 2013.

Guo, J., Guo, S., Xiao, N., Ma, X., Yoshida, S., Tamiya, T., et al. (2012). A novel robotic catheter system with force and visual feedback for vascular interventional surgery. *Int. J. Mech. Autom.* 2, 15–24. doi: 10.1504/IJMA.2012.046583

Guthar, G. S., and Salisbury, J. K. (2000). "The IntuitiveTM telesurgery system: overview and application" in *Proceedings-IEEE International Conference on Robotics and Automation* 1; April 24-28, 2000; 618–621.

Gwilliam, J. C., Mahvash, M., Vagvolgyi, B., Vacharat, A., Yuh, D. D., Okamura, A. M., et al. (2009). "Effects of haptic and graphical force feedback on teleoperated palpation" in *Icra: 2009 IEEE International Conference on Robotics and Automation, Vols 1–7*; May 12-17, 2009; 3315.

He, X., Handa, J., Gehlbach, P., Taylor, R., and Iordachita, I. (2014). A Submillimetric 3-DOF force sensing instrument with integrated fiber Bragg grating for retinal microsurgery. *IEEE Trans. Biomed. Eng.* 61, 522–534. doi: 10.1109/tbme.2013.2283501

Hosseini, S. M., Kashani, S. M. T., Najarian, S., Panahi, F., Naeini, S. M. M., and Mojra, A. (2010). A medical tactile sensing instrument for detecting embedded objects, with specific application for breast examination. *Int. J. Med. Robot.* 6, 73–82. doi: 10.1002/rcs.291

Hosseini, M., Najarian, S., Motaghinasab, S., and Dargahi, J. (2006). Detection of tumours using a computational tactile sensing approach. *Int. J. Med. Robot.* 2, 333–340. doi: 10.1002/rcs.112

Ju, F., Wang, Y., Zhang, Z., Wang, Y., Yun, Y., Guo, H., et al. (2019). A miniature piezoelectric spiral tactile sensor for tissue hardness palpation with catheter robot in minimally invasive surgery. *Smart Mater. Struct.* 28:025033. doi: 10.1088/1361-665X/aafc8d

Ju, F., Yun, Y., Zhang, Z., Wang, Y., Wang, Y., Zhang, L., et al. (2018). A variable-impedance piezoelectric tactile sensor with tunable sensing performance for tissue hardness sensing in robotic tumor palpation. *Smart Mater. Struct.* 27:115039. doi: 10.1088/1361-665X/aae54f

Kalantari, M., Ramezanifard, M., Ahmadi, R., Dargahi, J., and Koevecses, J. (2011). A piezoresistive tactile sensor for tissue characterization during catheter-based cardiac surgery. *Int. J. Med. Robot.* 7, 431–440. doi: 10.1002/rcs.413

Kim, U., Kim, Y. B., So, J., Seok, D. -Y., and Choi, H. R. (2018). Sensorized surgical forceps for robotic-assisted minimally invasive surgery. *IEEE Trans. Ind. Electron.* 65, 9604–9613. doi: 10.1109/tie.2018.2821626

Kim, U., Lee, D. -H., Yoon, W. J., Hannaford, B., and Choi, H. R. (2015). Force sensor integrated surgical forceps for minimally invasive robotic surgery. *IEEE Trans. Robot.* 31, 1214–1224. doi: 10.1109/tro.2015.2473515

Kitagawa, M., Dokko, D., Okamura, A. M., and Yuh, D. D. (2005). Effect of sensory substitution on suture-manipulation forces for robotic surgical systems. *J. Thorac. Cardiovasc. Surg.* 129, 151–158. doi: 10.1016/j.jtcvs.2004.05.029

Konstantinova, J., Jiang, A., Althoefer, K., Dasgupta, P., and Nanayakkara, T. (2014). Implementation of tactile sensing for palpation in robot-assisted minimally invasive surgery: a review. *IEEE Sens. J.* 14, 2490–2501. doi: 10.1109/jsen.2014.2325794

Lederman, S. J., and Klatzky, R. L. (1999). Sensing and displaying spatially distributed fingertip forces in haptic interfaces for teleoperator and virtual environment systems. *Presence Teleoperat. Virt. Environ.* 8, 86–103. doi: 10.1162/105474699566062

Lee, D. -H., Kim, U., Gulrez, T., Yoon, W. J., Hannaford, B., and Choi, H. R. (2016). A laparoscopic grasping tool with force sensing capability. *IEEE-Asme Trans. Mech.* 21, 130–141. doi: 10.1109/tmech.2015.2442591

Lee, J. -H., and Won, C. -H. (2013). The tactile sensation imaging system for embedded lesion characterization. *IEEE J. Biomed. Health Inform.* 17, 452–458. doi: 10.1109/jbhi.2013.2245142

Li, K., Pan, B., Zhan, J., Gao, W., Fu, Y., and Wang, S. (2015). Design and performance evaluation of a 3-axis force sensor for MIS palpation. *Sens. Rev.* 35, 219–228. doi: 10.1108/sr-04-2014-632

Li, T., Shi, C., and Ren, H. (2018). A high-sensitivity tactile sensor array based on fiber bragg grating sensing for tissue palpation in minimally invasive surgery. *IEEE-Asme Trans. Mech.* 23, 2306–2315. doi: 10.1109/tmech.2018.2856897

Liu, H., Greco, J., Song, X., Bimbo, J., Seneviratne, L., and Althoefer, K. (2012). "Tactile image based contact shape recognition using neural network" in *2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*; September 13-15, 2012; 138–143.

Liu, H., Puangmali, P., Zbyszewski, D., Elhage, O., Dasgupta, P., Dai, J. S., et al. (2010). An indentation depth-force sensing wheeled probe for abnormality identification during minimally invasive surgery. *Proc. Inst. Mech. Eng. H* 224, 751–763. doi: 10.1243/09544119jeim682

Liu, H., Wu, Y., Sun, F., and Guo, D. (2017). Recent progress on tactile object recognition. *Int. J. Adv. Robot. Syst.* 14, 1–12. doi: 10.1177/1729881417717056

Luo, S., Bimbo, J., Dahiya, R., and Liu, H. (2017). Robotic tactile perception of object properties: a review. *Mechatronics* 48, 54–67. doi: 10.1016/j.mechatronics.2017.11.002

Mahvash, M., Gwilliam, J., Agarwal, R., Vagvolgyi, B., Su, L. -M., Yuh, D. D., et al. (2008). "Force-feedback surgical teleoperator: controller design and palpation experiments." in: *Symposium on Haptics Interfaces for Virtual Environment and Teleoperator Systems 2008, Proceedings*. eds. J. Weisenberger, A. Okamura and K. MacLean.

Marohn, M. R., and Hanly, E. J. (2004). Twenty-first century surgery using twenty-first century technology: surgical robotics. *Curr. Surg.* 61, 466–473. doi: 10.1016/j.cursur.2004.03.009

Mohareri, O., Schneider, C., and Salcudean, S. (2014). "Bimanual telerobotic surgery with asymmetric force feedback: A daVinci (R) surgical system implementation" in *2014 Ieee/Rsj International Conference on Intelligent Robots and Systems*; September 14-18, 2014; 4272–4277.

Morimoto, A. K., Foral, R. D., Kuhlman, J. L., Zucker, K. A., Curet, M. J., Bocklage, T., et al. (1997). Force sensor for laparoscopic babcock. *Stud. Health Technol. Inform.* 39, 354–361.

Munawar, A., and Fischer, G. (2016). "Towards a haptic feedback framework for multi-DOF robotic laparoscopic surgery platforms" in *2016 IEEE/Rsj International Conference on Intelligent Robots and Systems*; October 9-14, 2016.

Najarian, S., Dargahi, J., and Zheng, X. Z. (2006). A novel method in measuring the stiffness of sensed objects with applications for biomedical robotic systems. *Int. J. Med. Robot.* 2, 84–90. doi: 10.1002/rcs.75

Narayanan, N. B., Bonakdar, A., Dargahi, J., Packirisamy, M., and Bhat, R. (2006). Design and analysis of a micromachined piezoelectric sensor for measuring the viscoelastic properties of tissues in minimally invasive surgery. *Smart Mater. Struct.* 15, 1684–1690. doi: 10.1088/0964-1726/15/6/021

Okamura, A. M. (2009). Haptic feedback in robot-assisted minimally invasive surgery. *Curr. Opin. Urol.* 19, 102–107. doi: 10.1097/MOU.0b013e32831a478c

Pacchierotti, C., Prattichizzo, D., and Kuchenbecker, K. J. (2016). Cutaneous feedback of fingertip deformation and vibration for palpation in robotic surgery. *IEEE Trans. Biomed. Eng.* 63, 278–287. doi: 10.1109/tbme.2015.2455932

Park, M., Bok, B. -G., Ahn, J. -H., and Kim, M. -S. (2018). Recent advances in tactile sensing technology. *Micromachines* 9:321. doi: 10.3390/mi9070321

Perri, M. T., Trejos, A. L., Naish, M. D., Patel, R. V., and Malthaner, R. A. (2010). New tactile sensing system for minimally invasive surgical tumour localization. *Int. J. Med. Robot.* 6, 211–220. doi: 10.1002/rcs.308

Pitakwatchara, P., Warisawa, S. I., and Mitsuishi, M. (2006). "Analysis of the surgery task for the force feedback amplification in the minimally invasive surgical system" in *Conference proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 30-September 3, 2006;* 829–832.

Prasad, S. K., Kitagawa, M., Fischer, G. S., Zand, J., Talamini, M. A., Taylor, R. H., et al. (2003). "A modular 2-DOF force-sensing instrument for laparoscopic surgery," in *Medical Image Computing and Computer-Assisted Intervention-Miccai 2003, Pt 1*. eds. R. E. Ellis and T. M. Peters, 279–286.

Puangmali, P., Althoefer, K., Seneviratne, L. D., Murphy, D., and Dasgupta, P. (2008). State-of-the-art in force and tactile sensing for minimally invasive surgery. *IEEE Sensors J.* 8, 371–381. doi: 10.1109/jsen.2008.917481

Reiley, C. E., Akinbiyi, T., Burschka, D., Chang, D. C., Okamura, A. M., and Yuh, D. D. (2008). Effects of visual force feedback on robot-assisted surgical task performance. *J. Thorac. Cardiovasc. Surg.* 135, 196–202. doi: 10.1016/j.jtcvs.2007.08.043

Richards, C., Rosen, J., Hannaford, B., Pellegrini, C., and Sinanan, M. (2000). Skills evaluation in minimally invasive surgery using force/torque signatures. *Surg. Endos.* 14, 791–798. doi: 10.1007/s004640000230

Roesthuis, R. J., Janssen, S., and Misra, S. (2013). "On using an array of fiber bragg grating sensors for closed-loop control of flexible minimally invasive surgical instruments" in *2013 IEEE/Rsj International Conference on Intelligent Robots and Systems*. ed. N. Amato. November 3-7, 2013; 2545–2551.

Rosen, J., Solazzo, M., Hannaford, B., and Sinanan, M. (2001). "Objective laparoscopic skills assessments of surgical residents using hidden Markov models based on haptic information and tool/tissue interactions" in *Medicine meets virtual reality 2001: Outer space, inner space, virtual space*. eds. J. D. Westwood, H. M. Hoffman, G. T. Mogel, D. Stredney and R. A. Robb (Amsterdam, BG: IOS Press). 417–423.

Saccomandi, P., Schena, E., Oddo, C. M., Zollo, L., Silvestri, S., and Guglielmelli, E. (2014). Microfabricated tactile sensors for biomedical applications: a review. *Biosensors* 4, 422–448. doi: 10.3390/bios4040422

Sadeghi-Goughari, M., Mojra, A., and Sadeghi, S. (2016). Parameter estimation of brain tumors using intraoperative thermal imaging based on artificial tactile sensing in conjunction with artificial neural network. *J. Phys. D. Appl. Phys.* 49:075404. doi: 10.1088/0022-3727/49/7/075404

Schostek, S., Ho, C. -N., Kalanovic, D., and Schurr, M. O. (2006). Artificial tactile sensing in minimally invasive surgery - a new technical approach. *Minim. Invasive Ther. Allied Technol.* 15, 296–304. doi: 10.1080/13645700600836299

Schostek, S., Schurr, M. O., and Buess, G. F. (2009). Review on aspects of artificial tactile feedback in laparoscopic surgery. *Med. Eng. Phys.* 31, 887–898. doi: 10.1016/j.medengphy.2009.06.003

Song, H. -S., Kim, H., Jeong, J., and Lee, J. -J. (2011). "Development of FBG sensor system for force-feedback in minimally invasive robotic surgery" in *2011 Fifth International Conference on Sensing Technology*; November 28-December 1, 2011.

Song, H. -S., Kim, K. -Y., and Lee, J. -J. (2009). "Development of the dexterous manipulator and the force sensor for minimally invasive surgery" in *Proceedings of the Fourth International Conference on Autonomous Robots and Agents*; Febraury 10-12, 2009.

Talasaz, A., and Patel, R. V. (2013). Integration of force reflection with tactile sensing for minimally invasive robotics-assisted tumor localization. *IEEE Trans. Haptics.* 6, 217–228. doi: 10.1109/ToH.2012.64

Talasaz, A., Trejos, A. L., and Patel, R. V. (2012). "Effect of force feedback on performance of robotics-assisted suturing." in *2012 4th IEEE Ras and Embs International Conference on Biomedical Robotics and Biomechatronics*. eds. J. P. Desai, L. P. S. Jay and L. Zollo. June 24-27, 2012; 823–828.

Talasaz, A., Trejos, A. L., and Patel, R. V. (2017). The role of direct and visual force feedback in suturing using a 7-DOF dual-arm Teleoperated system. *IEEE Trans. Haptics.* 10, 276–287. doi: 10.1109/toh.2016.2616874

Tholey, G., Desai, J. P., and Castellanos, A. E. (2005). Force feedback plays a significant role in minimally invasive surgery - results and analysis. *Ann. Surg.* 241, 102–109. doi: 10.1097/01.sla.0000149301.60553.1e

Tiwana, M. I., Redmond, S. J., and Lovell, N. H. (2012). A review of tactile sensing technologies with applications in biomedical engineering. *Sens. Actuat. Phys.* 179, 17–31. doi: 10.1016/j.sna.2012.02.051

Trejos, A. L., Jayender, J., Perri, M. T., Naish, M. D., Patel, R. V., and Malthaner, R. A. (2009). Robot-assisted tactile sensing for minimally invasive tumor localization. *Int. J. Robot. Res.* 28, 1118–1133. doi: 10.1177/0278364909101136

Uranues, S., Maechler, H., Bergmann, P., Huber, S., Hoebarth, G., Pfeifer, J., et al. (2002). Early experience with telemanipulative abdominal and cardiac surgery with the Zeus™ robotic system. *Eur. Surg.* 34, 190–193. doi: 10.1046/j.1563-2563.2002.t01-1-02049.x

Valdastri, P., Harada, K., Menciassi, A., Beccai, L., Stefanini, C., Fujie, M., et al. (2006). Integration of a miniaturised triaxial force sensor in a minimally invasive surgical tool. *IEEE Trans. Biomed. Eng.* 53, 2397–2400. doi: 10.1109/tbme.2006.883618

Verdura, J., Carroll, M. E., Beane, R., Ek, S., and Callery, M. P. (2000). *Systems methods and instruments for minimally invasive surgery*. United States patent application 6165184.

Wagner, C. R., and Howe, R. D. (2007). Force feedback benefit depends on experience in multiple degree of freedom robotic surgery task. *IEEE Trans. Robot.* 23, 1235–1240. doi: 10.1109/tro.2007.904891

Wang, X., Zhang, H., Dong, L., Han, X., Du, W., Zhai, J., et al. (2016). Self-powered high-resolution and pressure-sensitive triboelectric sensor matrix for real-time tactile mapping. *Adv. Mater.* 28, 2896–2903. doi: 10.1002/adma.201503407

Westebring-van der Putten, E. P., Goossens, R. H. M., Jakimowicz, J. J., and Dankelman, J. (2008). Haptics in minimally invasive surgery - a review. *Minim. Invasive Ther. Allied Technol.* 17, 3–16. doi: 10.1080/13645700701820242

Xie, H., Liu, H., Luo, S., Seneviratne, L. D., and Althoefer, K. (2013). "Fiber optics tactile array probe for tissue palpation during minimally invasive surgery" in *2013 IEEE/Rsj International Conference on Intelligent Robots and Systems*. ed. N. Amato. November 3-7, 2013; 2539–2544.

Yip, M. C., Yuen, S. G., and Howe, R. D. (2010). A robust uniaxial force sensor for minimally invasive surgery. *I.E.E.E. Trans. Biomed. Eng.* 57, 1008–1011. doi: 10.1109/tbme.2009.2039570

Yuan, W., Zhu, C., Owens, A., Srinivasan, M., and Adelson, E. (2017). "Shape-independent hardness estimation using deep learning and a GelSight tactile sensor" in *2017 IEEE International Conference on Robotics and Automation (ICRA)*; May 29-June 3, 2017; Singapore, Singapore: IEEE.

Zapata-Impata, B. S., Gil, P., and Torres, F. (2019). Learning spatio temporal tactile features with a ConvLSTM for the direction of slip detection. *Sensors* 19:523. doi: 10.3390/s19030523

Zemiti, N., Ortmaier, T., Vitrani, M. A., and Morel, G. (2006). "A force controlled laparoscopic surgical robot without distal force sensing" in *Experimental robotics ix*. eds. M. H. Ang and O. Khatib (Heidelberger, Berlin: Springer), 153.

Zhao, Z., Voros, S., Weng, Y., Chang, F., and Li, R. (2017). Tracking-by-detection of surgical instruments in minimally invasive surgery via the convolutional neural network deep learning-based method. *Comput. Assist. Surg.* 22, 26–35. doi: 10.1080/24699322.2017.1378777

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 Physiology

ORIGINAL RESEARCH
published: 23 December 2020
doi: 10.3389/fphys.2020.612928

3D Compressed Convolutional Neural Network Differentiates Neuromyelitis Optical Spectrum Disorders From Multiple Sclerosis Using Automated White Matter Hyperintensities Segmentations

Zhuo Wang[1,2], Zhezhou Yu[1], Yao Wang[1], Huimao Zhang[2,3], Yishan Luo[4], Lin Shi[4,5], Yan Wang[1] and Chunjie Guo[2,3]**

[1] Key Laboratory of Symbol Computation & Knowledge Engineering, Ministry of Education, College of Computer Science & Technology, Jilin University, Changchun, China, [2] Department of Radiology, the First Hospital of Jilin University, Changchun, China, [3] Jilin Provincial Key Laboratory for Medical imaging, Changchun, China, [4] BrainNow Research Institute, Hong Kong, China, [5] Department of Imaging and Interventional Radiology, Chinese University of Hong Kong, Hong Kong, China

OPEN ACCESS

Edited by:
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Qi Liu,
Tongji University, China
Wen Qin,
Tianjin Medical University General Hospital, China

***Correspondence:**
Yan Wang
wy6868@jlu.edu.cn
Chunjie Guo
guocj@jlu.edu.cn

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

Received: 01 October 2020
Accepted: 07 December 2020
Published: 23 December 2020

Citation:
Wang Z, Yu Z, Wang Y, Zhang H, Luo Y, Shi L, Wang Y and Guo C (2020) 3D Compressed Convolutional Neural Network Differentiates Neuromyelitis Optical Spectrum Disorders From Multiple Sclerosis Using Automated White Matter Hyperintensities Segmentations. Front. Physiol. 11:612928. doi: 10.3389/fphys.2020.612928

Background: Magnetic resonance imaging (MRI) has a wide range of applications in medical imaging. Recently, studies based on deep learning algorithms have demonstrated powerful processing capabilities for medical imaging data. Previous studies have mostly focused on common diseases that usually have large scales of datasets and centralized the lesions in the brain. In this paper, we used deep learning models to process MRI images to differentiate the rare neuromyelitis optical spectrum disorder (NMOSD) from multiple sclerosis (MS) automatically, which are characterized by scattered and overlapping lesions.

Methods: We proposed a novel model structure to capture 3D MRI images' essential information and converted them into lower dimensions. To empirically prove the efficiency of our model, firstly, we used a conventional 3-dimensional (3D) model to classify the T2-weighted fluid-attenuated inversion recovery (T2-FLAIR) images and proved that the traditional 3D convolutional neural network (CNN) models lack the learning capacity to distinguish between NMOSD and MS. Then, we compressed the 3D T2-FLAIR images by a two-view compression block to apply two different depths (18 and 34 layers) of 2D models for disease diagnosis and also applied transfer learning by pre-training our model on ImageNet dataset.

Results: We found that our models possess superior performance when our models were pre-trained on ImageNet dataset, in which the models' average accuracies of 34 layers model and 18 layers model were 0.75 and 0.725, sensitivities were 0.707 and 0.708, and specificities were 0.759 and 0.719, respectively. Meanwhile, the traditional 3D CNN models lacked the learning capacity to distinguish between NMOSD and MS.

Conclusion: The novel CNN model we proposed could automatically differentiate the rare NMOSD from MS, especially, our model showed better performance than traditional3D CNN models. It indicated that our 3D compressed CNN models are applicable in handling diseases with small-scale datasets and possess overlapping and scattered lesions.

Keywords: Neuromyelitis optical spectrum disorder (NMOSD), multiple sclerosis (MS), magnetic resonance imaging (MRI), deep learning, convolutional neural networks (CNNs)

INTRODUCTION

Neuromyelitis optical spectrum disorder (NMOSD) is a rare aquaporin-4 immunoglobin G antibody (AQP4-IgG) mediated chronic disorder of the brain and spinal cord (Wingerchuk et al., 2007, 2015). Traditionally considered a subtype of multiple sclerosis (MS), NMOSD has been recognized as a distinct clinical entity based on unique immunologic features in recent years (Wingerchuk et al., 2015). Up to 70% of NMOSD patients have brain lesions visible on magnetic resonance imaging (MRI) (Kim et al., 2015). But only about half of NMOSD patients have typical brain lesions, and their distributions of NMOSD and MS are overlapped (Cacciaguerra et al., 2019). Furthermore, compared to MS and stroke, it is challenging to segment and quantify white matter lesions (WMLs) on T2-weighted fluid-attenuated inversion recovery (T2-FLAIR) images in NMOSD, as its lesions are often located very close to the ventricles. However, it is vital to differentiate NMOSD from MS. Some MS treatments such as β-interferon can worsen NMOSD (Jacob et al., 2012; Kim et al., 2012), but distinguishing between the two disease entities is challenging. Studies based on machine learning to discriminate NMOSD from MS are limited.

Machine learning algorithms that precede human observation have shown application potential in medical image processing (Wernick et al., 2010). These algorithms handle a large number of features extracted from patients and lack inconsistencies (Eshaghi et al., 2016). Therefore, machine learning algorithms can build decision systems to support the diagnostic process. Previous studies have proved the efficiency and robustness of machine learning algorithms for many common diseases, such as breast cancer (Rastghalam and Pourghassem, 2016), brain tumors (Zacharaki et al., 2009), etc. However, NMOSD is a rare disease, which is a lack of large-scale public datasets for scientific research, and its similar phenotypes with MS bring challenges to build high-performance machine learning models. Laura Cacciaguerra et al. used typical/atypical brain and spinal cord lesions to construct a possible evidence-based diagnostic machine learning algorithm to discriminate NMOSD from MS with the sensitivity of 0.92, 0.82, and specificity of 0.91,0.91 in training and validation samples separately. However, the blinded machine learning approaches were not conducted yet (Cacciaguerra et al., 2019). Eshaghi et al. built a machine learning classifier using brain gray matter (GM) imaging measures to distinguish patients with MS from those with NMOSD with an average accuracy of 74%. When they used thalamic volume together with the white matter lesion volume, the classifier achieved an average accuracy of 80% (Eshaghi et al., 2016). Machine learning-based models indeed show applicational potential for medical image processing; however, there are some issues to resolve. First, medical images can't be the model's input directly; all features have to be extracted from the raw images by radiologists, which means the radiologists have already studied the raw data's intrinsic information. Feature extraction processing, which is the bottleneck of the models' performances, is highly influenced by the radiologists' subjective judgments. Second, manual intervention is indispensable for both the training and testing phases, which means radiologists must process all images.

As a subfield of machine learning, deep learning can solve the problems as mentioned above. Deep learning algorithms can efficiently extract raw images' features through convolutional neural networks (CNNs), and they are also widely applied to the medical images' classification, segmentation, and detection tasks (Litjens et al., 2017). Researchers have conducted on the deep CNNs and have achieved better results than with other machine learning algorithms. Shen et al. (2015) proposed M-CNN to distinguish between malignant and benign nodules without nodule segmentation. Nie et al. (2016) proposed a three-dimensional (3D) CNN model to predict the overall survival (OS) for brain glioma patients. Payan et al. used 3D-CNN to process brain MRI data of Alzheimer's disease (AD), and the classification of AD reached an accuracy rate of 89.47% (Payan and Montana, 2015). Wang et al. built an ensemble 3D-DenseNet to predict AD (Wang H. et al., 2019). U-net, proposed by Ronneberger et al. (2015), had a good effect on two-dimensional (2D) medical image segmentation. Based on this, Milletari et al. (2016) combined U-net and ResNet to propose V-net to solve the image segmentation problems of 3D data. Previous work based on 3D deep learning models focused on diseases with large data sets and concentrated lesions, such as AD and tumors. Multiple 3D-CNNs were used to extract the features and demonstrated the effectiveness of dealing with 3D images. 3D medical images are the input of 3D CNN models, which can reflect the whole lesion. However, there is still a lack of 3D CNN models to handle rare diseases with small-scale datasets, such as NMOSD. Significantly, the lesion distributions of NMOSD and MS are overlapped, which brings more challenges to build a high-performance 3D CNN model. 2D CNN model with less learnable parameters is more easily trained with small-scale datasets. Our experiment created a new model combining the advantages of 3D and 2D CNN models to differentiate NMOSD from MS in terms of WMLs segmentation with less learnable parameters and achieved better performance than the 3D ResNet baseline.

In this paper, we (i) investigated the traditional 3D CNN model for a 3D MRI data classification task and found that the

conventional approach lacks generalization ability for NMOSD and MS classification; (ii) presented a two-view 2D model to boost the classification performance, comparing models that were pre-trained on an ImageNet dataset with models not pre-trained; (iii) set up experiments to analyze the primary factors influencing the experiments' results.

MATERIALS AND METHODS

Data Description

Participants

A retrospective sample of 41 NMOSD patients diagnosed according to the revised 2015 diagnostic criteria (Wingerchuk et al., 2015) was recruited in this work. 47 MS patients who had received their diagnosis according to the 2010 McDonald Criteria (Polman et al., 2011) were enrolled, and they also fulfilled the recently revised diagnostic criteria (Thompson et al., 2018). The MS group matched for age, sex, disease duration, and Expanded Disability Status Scale (EDSS) (Kurtzke, 1983) scores to the NMOSD group. All patients in the acute disease phase with brain MRI lesions were included in this study. Clinical characteristics, including EDSS scores of all patients, were performed within 48 hours from the MRI acquisition (**Table 1**). All the patients undergoing high-dose corticosteroid treatment or with a medical condition that could result in hyperintensity on T2-weighted and T2-FLAIR images were excluded. Besides, neurological comorbidities, a history of head trauma or surgery, and low-quality images or severe motion artifacts were excluded. This study was approved by the local ethics committee, and written informed consent was obtained from all participants.

MRI Acquisition

The MRI protocol included 3D T2-FLAIR and 3D T1-weighted (T1W) fast field echo (FFE) sequences were obtained from the same 3.0T Philips Ingenia scanner (Philips Healthcare, Best, Netherland) between July 2015 and April 2018. There were some acquisition parameter variations over the years, where images were acquired axially or sagittally with parameter ranges. Sagittal/Axial T2-FLAIR: repetition time (TR) = 4800 m sec, echo time (TE) = 279–324 m sec, flip angle = 90°, number of slices = 160–192, field of view (FOV) = 220 mm, acquisition matrix = 224*224, section thickness = 2 mm. Sagittal T1W: repetition time (TR) = 7.0–7.8 m sec, echo time (TE) = 3.2–3.6 m sec, flip angle = 7°, number of sections = 160–192, field of view (FOV) = 220 mm, acquisition matrix = 240*240, section thickness = 1 mm.

TABLE 1 | Demographic, clinical characteristics and brain WMH volume measurement.

	MS (n = 47)	NMOSD (n = 41)	P-value
Age (years)	40.0 ± 11.1	44 ± 13	0.0865
Gender	14 M/33F	8 M/33 F	0.2720
Disease Duration (months)	57.1 ± 73.4	35.0 ± 42.9	0.1133
EDSS	3.6 ± 1.8	4.2 ± 2.4	0.1949
ICV (ml)	1407.2 ± 171.2	1367.3 ± 120.3	0.2165
WMH (ml)	9.0 ± 8.8	3.5 ± 4.8	0.0007
Normalized WMH (% of ICV)	0.66 ± 0.65	0.26 ± 0.37	0.0008

Automated Segmentation

The automated segmentation software, AccuBrain IV1.0® (Luo, 2017; Abrigo et al., 2019; Guo et al., 2019; Wang C. et al., 2019) (BrainNow Medical Technology Limited, Hong Kong, China), was used in the WML segmentation. Using AccuBrain, WML segmentation was automatically performed on T1W and T2-FLAIR images. Segmentation was first performed on T1 images, where brain structure masks and tissue masks were generated. Then, T1W and T2-FLAIR images were co-registered, and the structure and tissue masks were transformed into the T2-FLAIR space. Using a coarse-to-fine white matter hyperintensities (WMH) segmentation process, which utilizes mathematical morphological operations including binary dilation, grayscale closing, binary reconstruction, and grayscale reconstruction (Shi et al., 2013), WMH is extracted on T2-FLAIR images and is refined using the transformed brain structure mask from T1WI. The intracranial volume (ICV) is also calculated automatically, then the normalized WMH (% of ICV) is available. **Figure 1** shows examples of the results of automated segmentation. The performance of WMH has been validated to have good accuracy and reproducibility (Guo et al., 2019).

Data Processing

Since NMOSD is a rare disease, it hard to find large-scale MRI datasets for research. We have collected the MRI datasets with a wide time range, so there are small differences in the size of images. Hence, we unified and centralized the raw images for the first time. The diagram of our data preprocessing procedure is shown in **Figure 2**. The method is carried out in three steps: i) we crop the raw MRI T2-FLAIR images to remove the black background and get the fixed crop kernel; ii) we crop the WMH brain lesion image with the fixed crop kernel that obtained at pervious step. iii) we use the OpenCV package to resize each section of peer MRI data to unify the data shape to 100*100*100 pixels; (iv) we apply data augmentation by shifting and flipping.

Data and Code Availability Statement

The reconstruction algorithms to support the findings of this study are still under early-stage development. For the datasets, we do not share them directly due to the ethics of clinical research. Both the codes and data can only be acquired via a special request to the corresponding author.

MODEL STRUCTURE

Previous studies have pointed out that the deeper the network, the stronger its learning ability (Bengio and Lecun, 2007; Bianchini and Scarselli, 2014; Montufar et al., 2014). But information loss as a common problem of traditional

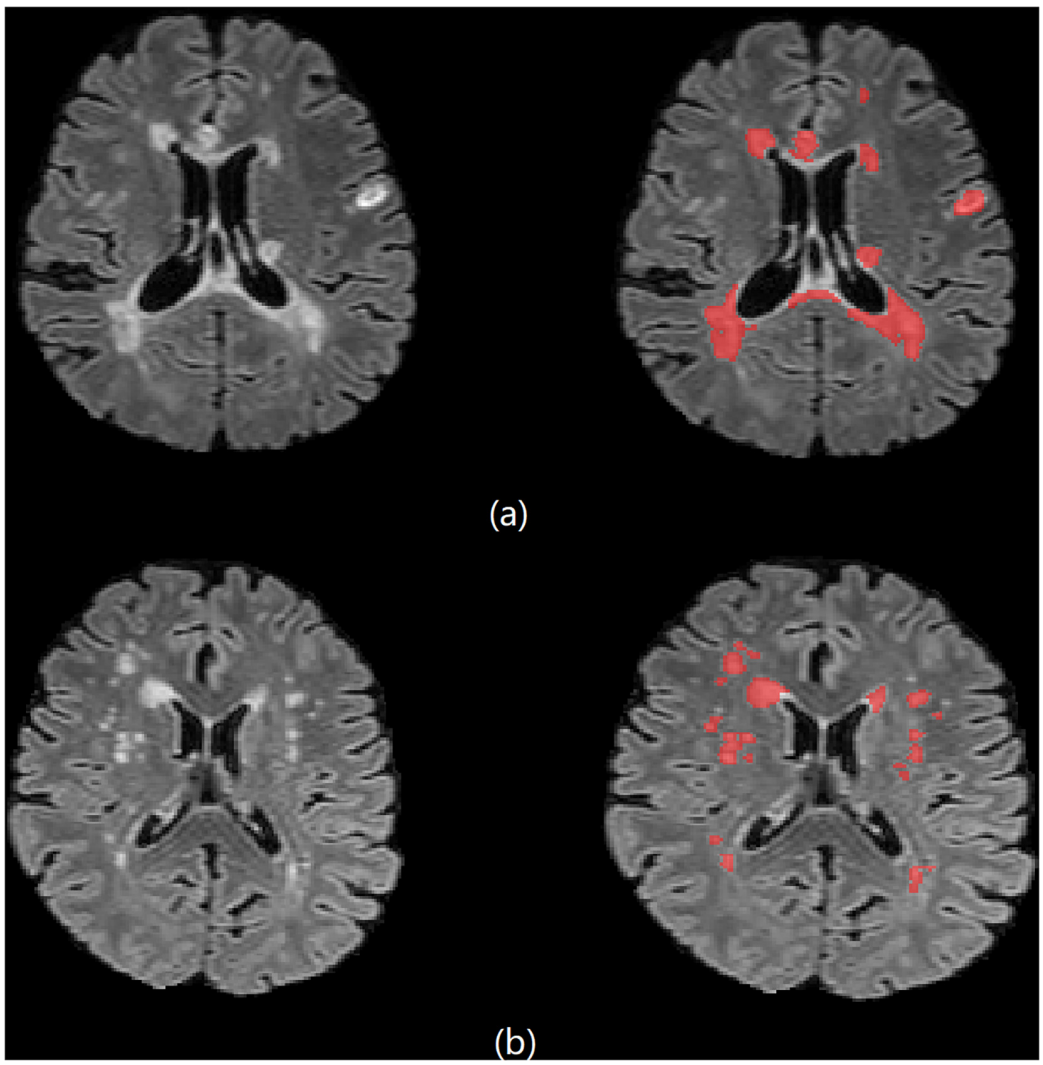


FIGURE 1 | Brain 3D T2-FLAIR MR images of **(a)** one 33-year-old female MS patient and **(b)** one 61-year-old female NMOSD patient.

deep networks often appears in the process of information transmission. At the same time, the model is hard to train because of the gradient vanishing and the gradient explosion. He K. et al. proposed the idea of residual learning to solve this problem: To maintain the integrity of the information, He K. et al. proposed a model structure to bypass the input information to the output, which makes the model deeper (He et al., 2016). Formally, we consider a ResNet block defined as:

$$y = F(x, \theta) + x \tag{1}$$

where x and y are the input and output of the ResNet blocks considered. θ is a set of learning parameters of function $F(x)$.

Batch normalization (BN) is added to help the network increase generalization ability and accelerate the training process (Ioffe and Szegedy, 2015). ReLU is added to the network as a non-linear activation function.

3D ResNet for Image Classification

In this section, inspired by the work of Hara, K. et al. and ResNet (He et al., 2016), and compared with the previous studies of 3D CNN models (Payan and Montana, 2015; Hara et al., 2017, 2018; Liang et al., 2018; Wang H. et al., 2019), we used 3D ResNet as baseline models to solve the classification problem of NMOSD and MS MRI T2-FLAIR images. For grayscale MRI T2-FLAIR images, the shape of input data for the network is $(1, h, w, l)$, and (h, w, l) representing the shape size (height, width, length) of each MRI T2-FLAIR image. The 3d convolution layer with input size$(C_{in},\ h_{in},\ w_{in},\ l_{in})$ and output $(C_{out},\ h_{out},\ w_{out},\ l_{out})$ can be precisely described as:

$$out\left(C_{out}\right) = \sum_{k=0}^{C_{in}-1} weight\left(k\right) input(k) + bais \tag{2}$$

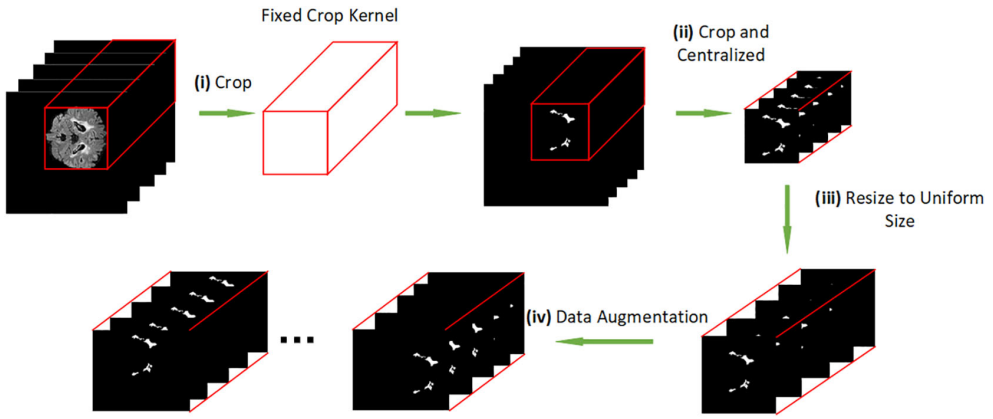


FIGURE 2 | (i) cropping the raw MRI T2-FLAIR images to remove the background. **(ii)** cropping the brain lesion image with the same shape as the resized brain MRI T2-FLAIR images. **(iii)** resizing each slice of peer MRI T2-FLAIR images to 100*100*100 pixels. **(iv)** data augmentation.

where is the valid 3D cross-correlation operator, and 2d convolution calculates similarly.

2D ResNet for Image Classification

The large sample space and small data sets bring challenges to the 3D deep learning model's training. Transfer learning can improve the model (Pan and Yang, 2010; Tajbakhsh et al., 2016), but there is no available 3D pre-training model for image classification. On the other hand, it's impossible to distinguish NMOSD from MS by a single 2D slice because of the highly overlapping brain lesions. A compression block proposed to extract the most critical features each 2D slice by one view and map the 3D input $(1, h, w, l)$ to a 2D form (c, h', w'), where c represents the image's channel. The convolutional kernels calculate each channel and sum them up to new channels that merge each 2D slice's internal relationship, making the mapping operation lose structural information. Therefore, we proposed a two-view structure to extract the essential features from two different axes. The internal structure information is retained by concatenating the compression blocks' output, which effectively reduces the sample space and can apply the compressed data to transfer learning. **Figure 3** shows our model architecture, and the compression block used in our model. In the compression block, the first convolution layer is designed to extract information from input image. After many attempts with different parameters, we set the convolution layer with a kernel size of 7 and stride of 3, and the output channel of this layer is 32, which may get the best performance. The compression operation can be precisely described as:

$$x(h, w, l) \overset{h \rightrightarrows c}{\Rightarrow} x_1(c, w, l) \tag{3}$$

$$x(h, w, l) \overset{w \rightrightarrows c}{\Rightarrow} x_2(h, c, l) \tag{4}$$

$$y_1 = \sigma(BN\left(conv2d\left(x_1\right) + b_1\right)) \tag{5}$$

$$y_2 = \sigma(BN\left(conv2d\left(x_2\right) + b_2\right)) \tag{6}$$

$$y = concatenate\left(y_1, y_2\right) \tag{7}$$

where BN is the batch normalization layer, and σ denotes ReLU. Equation (3), (4) demonstrates that we change one axis of 3D data to a channel, which means the compression block extract features from one view.

Implementation Details

Our model is based on python and PyTorch. The optimizer used in training is SGD, the initial learning rate is 0.01, and the learning rate is reduced by 10% every ten steps. We added dropout before the final fully connected layer (Srivastava et al., 2014) to prevent model overfitting on the training data, with a dropout rate of 0.15. We trained our model on a server with one NVIDIA 1080ti. We also applied data augmentation for T2-FLAIR MRI images by shifting. To reduce the accidental factors, we conducted five-fold cross-validations to ensure that each image was tested at least once and also repeated the cross-validations process for 15 times to average the results. T2-FLAIR MRI images of 41 NMOSD subjects and 47 MS subjects were used for the experiments. The original data set was randomly divided into five equal-sized groups to utilize a five-fold cross-validation method to evaluate the model's performance, which means four groups were used for model training, and one group was regarded as the validation dataset.

RESULTS

3D and 2D ResNet Model for Image Classification

In this experiment, we applied a 3D RseNet model to process NMOSD and MS MRI T2-FLAIR images. The 3D model we used was based on the models proposed by Payan and Montana (2015), Hara et al. (2017, 2018), Wang H. et al. (2019), which were applied traditional 3D ResNet for classification tasks and achieved desired results. However, the overlapping lesions and the limited data set's size of NMOSD and MS restrict the learning efficiency of 3D ResNet. Considering the above problems, we

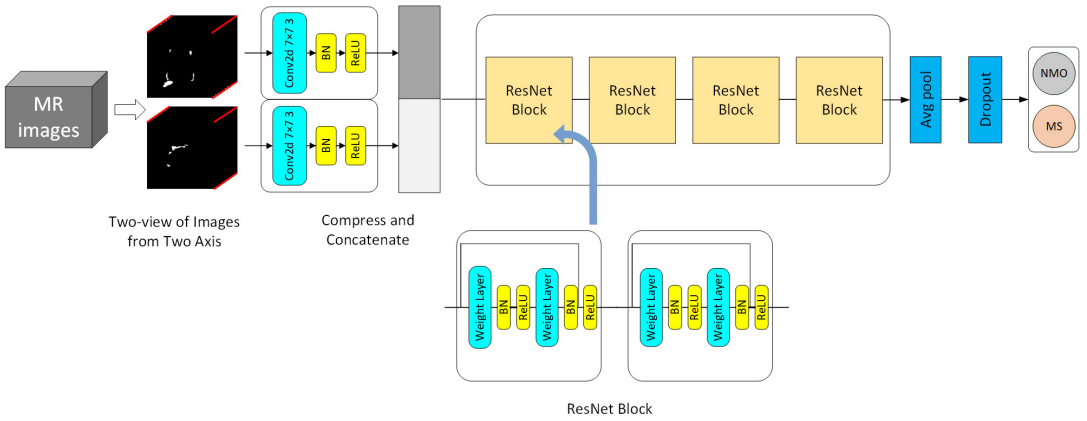


FIGURE 3 | 2D model architecture for classification of NMOSD and MS.

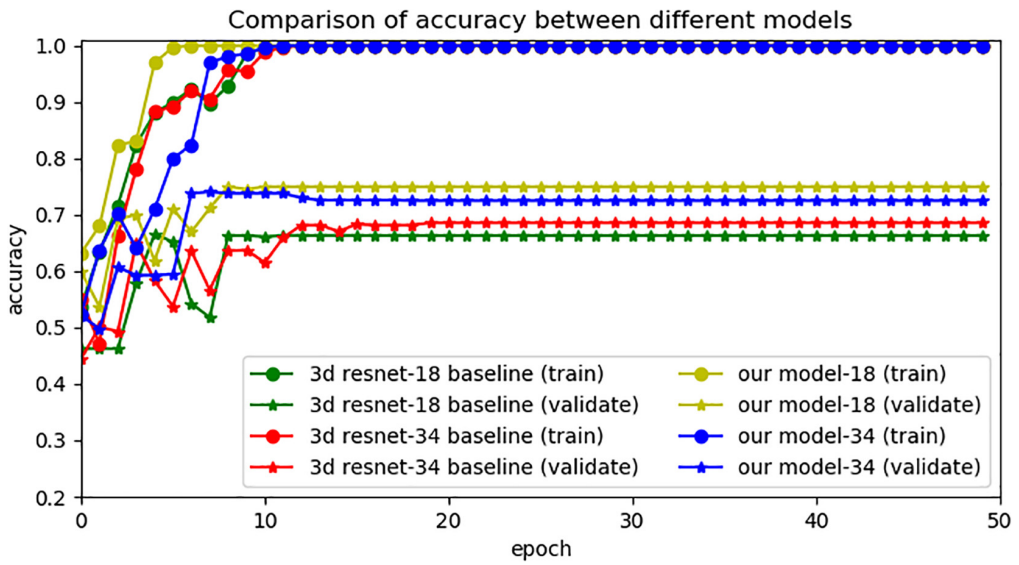


FIGURE 4 | Comparison of accuracy between different models.

proposed a compression block to map the high dimensional 3D data into a lower dimension. It can also extract the information from the overlapping lesion locations by learning each single MRI section and building long-range dependence at the same time. After compressing the 3D data, we applied 2D ResNet for model processing and used transfer learning to improve the model's generalization ability effectively. **Figure 4** shows the accuracy of different models. We set 3D ResNet-18 (18 layers) and 3D ResNet-34 (34 layers) to the baseline models (Payan and Montana, 2015; Hara et al., 2017, 2018; Wang H. et al., 2019). This experiment demonstrates the limitations of traditional 3D CNN models' learning capacity on NMOSD and MS datasets. The resulting diagram also showed that the accuracy of models changed rapidly, indicating the convergences in the training phase. In contrast, the validation phase's unsatisfactory validation performances of baseline model showed that the generalization ability of the 3D ResNet was not ideal.

Overall, the experimental results indicated that our models have higher prediction performance, as embodied in higher validation accuracy. It also reflected that our model needs less time to train. The best performance was achieved when the model was pretrained on ImageNet datasets.

Comparing Different Models

To compare the differences in performance between different models, we repeated the experiment 15 times. **Table 2** shows the statistical test results. **Table 3** shows the model complexity statistics.

The experimental results show that our model was better than traditional 3D CNNs and non-pretrained models, as the best accuracy, sensitivity and specificity of our models on the validation dataset was 0.75, 0.707 and 0.759 when the depth of our model was 18 layers. The performance of our pretrained model was improved comparing with the 3D baseline and

TABLE 2 | Average performance comparison between different models.

Model	Accuracy (mean ± std)	Sensitivity (mean ± std)	Specificity (mean ± std)
Pretrained ResNet-18(ours)	**0.750 ± 0.02**	0.707 ± 0.09	**0.759 ± 0.06**
Pretrained ResNet-34(ours)	0.725 ± 0.01	**0.708 ± 0.08**	0.719 ± 0.11
No-pretrained ResNet-18	0.696 ± 0.01	0.689 ± 0.09	0.707 ± 0.07
No-pretrained ResNet-34	0.690 ± 0.04	0.653 ± 0.15	0.706 ± 0.09
3D ResNet-18	0.669 ± 0.02	0.694 ± 0.06	0.661 ± 0.06
3D ResNet-34	0.689 ± 0.04	0.701 ± 0.07	0.657 ± 0.07

no-pretrained models, it indicates that both the compression block and pretrained datasets were beneficial for performance and generalization ability improvement. When we applied the compression block, the sample space of the image was effectively reduced, and the intrinsic features of images were easier to extract. **Table 2** demonstrates that our model needs fewer parameters and less training time.

DISCUSSION

In this paper, we used deep learning models to automatically process WMH segmentation T2-FLAIR images to distinguish NMOSD from MS automatically, and none of the approaches had manual interventions. Previous studies based on 3D deep learning models focused on diseases with large data sets and concentrated lesions, for example, AD and tumors. Using 3D-CNN to process 3D T2-FLAIR images is indeed in line with people's intuition. However, since 3D-CNN increases the number of parameters, the model will be low learning efficiency on the training data when the data sets is limited. Especially, the overlapping lesion increase the difficulty for training the model, which makes the generalization ability of the traditional 3D models is weak. The experimental result also shows that the traditional 3D CNN model lacks the learning capacity for NMOSD and MS MRI T2-FLAIR images. There are some reasons for those results:i) The data set size was limited because of the disease's rareness, which means the model did not feed in abundant data to be generalized. ii) The sample space was too large due to the wide distribution of lesions. iii) Only about half of the NMOSD patients have typical brain lesions, and the brain lesion distribution of NMOSD and MS have overlap in lesion localization (Cacciaguerra et al., 2019). Transfer learning is a widely used technique that allows the model to have a better initial parameter, which will enhance the performance of the model, especially for dealing with small sample data. Currently, there is no 3D pre-trained model available to apply to NMOSD and MS T2-FLAIR images. In Experiment 2, our model compressed the 3D data and then did fine-tuning to effectively improve the performance, which also proved that models pre-trained on ImageNet datasets have better generalization ability for medical images. Our model effectively reduced the sample space dimension and reduced parameter amount, making the model much more easily to train and increase the generalization ability of our model.

This study has some limitations. First, distinguishing NMOSD from MS by automatic WMLs segmentation is a great challenge because of their variability and scattered spatial distribution. In this study, three high error rate cases disagreed with the necessary diagnostic categories due to imprecise segmentation. The outliers had lesions that were part of a confluence of lesions or were located very close to the ventricles, which are extremely difficult for automatic quantification, and semimanual lesion outline correction will be conducted in our next study. Second, the AQP4-IgG seronegative NMOSD patients were not excluded from this study to avoid introducing demographic information bias. However, it is a clinical challenge to distinguish AQP4-IgG seronegative NMOSD from atypical MS. The mean proportion of agreement for diagnosing the two diseases was low among expert clinicians (Po = 0.5) (Wingerchuk et al., 2015), which means the AQP4-IgG seronegative NMOSD subjects might increase the difficulties in the classification of these two diseases. Besides, MS subjects with lesions mimicking stroke were also likely to be mistaken as NMOSD. Third, the prevalence of NMOSD is approximately 0.52–4.4 per 100,000 individuals (Wingerchuk et al., 2007); thus, only a small sample size was recruited in this pilot study, bringing challenges for deep learning models.

TABLE 3 | Complexity comparison between different models.

Model	Parameters	Training Time(s/epoch)
ResNet-18(ours)	**11,475,330**	**5**
ResNet-34(ours)	21,583,490	6
3D ResNet-18	33,161,986	16
3D ResNet-34	63,471,618	23

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because we do not share them directly due to the ethics of the clinical study. Requests to access the datasets should be directed to CG, guocj@jlu.edu.cn.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the local ethics committee of the First Hospital of Jilin University, China (2015-133). The

patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZW, ZY, and CG performed study conception and design. ZW, YaoW, and YanW performed the study. HZ, YL, and LS analyzed the data. YaoW and CG drafting of the manuscript. YanW and CG performed critical revision. 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. 62072212, 61772227, and 81600923), the Development Project of Jilin Province of China (Nos. 20200401083GX, 2020C003, 20190902012TC, and 20190304129YY). Foundation of Health and Family Planning Commission of Jilin Province (2017J073). This work was also supported by the Jilin Provincial Key Laboratory of Big Data Intelligent Computing (No. 20180622002JC).

REFERENCES

Abrigo, J., Shi, L., Luo, Y., Chen, Q., Chu, W. C. W., Mok, V. C. T., et al. (2019). Standardization of hippocampus volumetry using automated brain structure volumetry tool for an initial Alzheimer's disease imaging biomarker. *Acta Radiol.* 60, 769–776. doi: 10.1177/0284185118795327

Bengio, Y., and Lecun, Y. (2007). "Scaling learning algorithms towards AI," in *Large-Scale Kernel Machines*, Vol. 34, eds L. Bottou, O. Chapelle, D. DeCoste, and J. Weston (Cambridge, MA: MIT Press), 1–41.

Bianchini, M., and Scarselli, F. (2014). On the complexity of neural network classifiers: a comparison between shallow and deep architectures. *IEEE Trans. Neural Netw. Learn. Syst.* 25, 1553–1565. doi: 10.1109/tnnls.2013.2293637

Cacciaguerra, L., Meani, A., Mesaros, S., Radaelli, M., Palace, J., Dujmovic-Basuroski, I., et al. (2019). Brain and cord imaging features in neuromyelitis optica spectrum disorders. *Ann. Neurol.* 85, 371–384.

Eshaghi, A., Wottschel, V., and Cortese, R. (2016). Gray matter MRI differentiates neuromyelitis optica from multiple sclerosis using random forest. *Neurology* 87, 2463–2470. doi: 10.1212/wnl.0000000000003395

Guo, C., Niu, K., Luo, Y., Shi, L., Wang, Z., Zhao, M., et al. (2019). Intra-scanner and inter-scanner reproducibility of automatic white matter hyperintensities quantification. *Front. Neurosci.* 13:679. doi: 10.3389/fnins.2019.00679

Hara, K., Kataoka, H., and Satoh, Y. (2017). "Learning spatio-temporal features with 3d residual networks for action recognition," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, Nashville, TN, 3154–3160.

Hara, K., Kataoka, H., and Satoh, Y. (2018). "Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and ImageNet," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 6546–6555.

He, K., Zhang, X., and Ren, S. (2016). "Deep residual learning for image recognition," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 770–778.

Ioffe, S., and Szegedy, C. (2015). "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the International Conference on Machine Learning*, Vienna, 448–456.

Jacob, A., Hutchinson, M., and Elsone, L. (2012). Does natalizumab therapy worsen neuromyelitis optica? *Neurology* 79, 1065–1066. doi: 10.1212/wnl.0b013e31826845fe

Kim, H. J., Paul, F., and Lana-Peixoto, M. A. (2015). MRI characteristics of neuromyelitis optica spectrum disorder. *Neurology* 84, 1165–1173.

Kim, S.-H., Kim, W., Li, X. F., Jung, I. J., and Kim, H. J. (2012). Does interferon beta treatment exacerbate neuromyelitis optica spectrum disorder? *Mult. Scler. J.* 18, 1480–1483. doi: 10.1177/1352458512439439

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 33, 1444–1452. doi: 10.1212/wnl.33.11.1444

Liang, S., Zhang, R., Liang, D., Song, T., Ai, T., Xia, C., et al. (2018). Multimodal 3D densenet for IDH genotype prediction in gliomas. *Genes* 9:382. doi: 10.3390/genes9080382

Litjens, G., Kooi, T., and Bejnordi, B. E. (2017). A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88.

Luo, Y. (2017). Automate the quantitative calculation method of subregion brain atrophy. China Patent NO CN107103612B. Shanghai: China Patent and Trademark Office.

Milletari, F., Navab, N., and Ahmadi, S. (2016). "V-net: fully convolutional neural networks for volumetric medical image segmentation," in *Proceeding of the International Conference on 3D Vision (3DV)*, Stanford, CA, 565–571.

Montufar, G., Pascanu, R., and Cho, K. (2014). "On the number of linear regions of deep neural networks," in *Proceedings of the International Conference on Neural Information Processing Systems*, Tallinn, 2924–2932.

Nie, D., Zhang, H., Adeli, E., Liu, L., and Shen, D. (2016). 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. *Med. Image Comput. Comput. Assist. Interv.* 9901, 212–220. doi: 10.1007/978-3-319-46723-8_25

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359.

Payan, A., and Montana, G. (2015). "Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks," in *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, Prague, 355–362.

Polman, C. H., Reingold, S. C., and Banwell, B. (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann. Neurol.* 69, 292–302.

Rastghalam, R., and Pourghassem, H. (2016). Breast cancer detection using MRF-based probable texture feature and decision-level fusion-based classification using HMM on thermography images. *Pattern Recognit.* 51, 176–186. doi: 10.1016/j.patcog.2015.09.009

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, 234–241.

Shen, W., Zhou, M., Yang, F., Yang, C., and Tian, J. (2015). Multi-scale convolutional neural networks for lung nodule classification. *Inf. Process. Med. Imaging* 24, 588–599. doi: 10.1007/978-3-319-19992-4_46

Shi, L., Wang, D., Liu, S., Pu, Y., Wang, Y., Chu, W. C. W., et al. (2013). Automated quantification of white matter lesion in magnetic resonance imaging of patients with acute infarction. *J. Neurosci. Methods* 213, 138–146. doi: 10.1016/j.jneumeth.2012.12.014

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Todd Hurst, R., Kendall, C. B., Gotway, M. B., et al. (2016). Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans. Med. Imaging* 35, 1299–1312. doi: 10.1109/tmi.2016.2535302

Thompson, A. J., Banwell, B. L., and Barkhof, F. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 17, 162–173.

Wang, C., Zhao, L., Luo, Y., Liu, J., Miao, P., Shi, L., et al. (2019). Structural covariance in subcortical stroke patients measured by automated MRI-based volumetry. *Neuroimage Clin.* 22:101682. doi: 10.1016/j.nicl.2019.101682

Wang, H., Shen, Y., Wang, S., Xiao, T., Deng, L., Wang, X., et al. (2019). Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease. *Neurocomputing* 333, 145–156. doi: 10.1016/j.neucom.2018.12.018

Wernick, M. N., Yang, Y., and Brankov, J. G. (2010). Machine learning in medical imaging. *IEEE Signal Process. Mag.* 27, 25–38.

Wingerchuk, D. M., Banwell, B., Bennett, J. L., Cabre, P., Carroll, W., Chitnis, T., et al. (2015). International consensus diagnostic criteria for neuromyelitis optica spectrum disorders. *Neurology* 85, 177–189.

Wingerchuk, D. M., Lennon, V. A., Lucchinetti, C. F., Pittock, S. J., and Weinshenker, B. G. (2007). The spectrum of neuromyelitis optica. *Lancet Neurol.* 6, 805–815.

Zacharaki, E. I., Wang, S., and Chawla, S. (2009). Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn. Reson. Med.* 62, 1609–1618. doi: 10.1002/mrm.22147

Conflict of Interest: LS is the director of BrainNow Medical Technology Limited. YL is an employee of BrainNow Medical Technology Limited, which developed AccuBrain® used in this paper.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 07 January 2021
doi: 10.3389/fmolb.2020.613918

A Metabolism-Related Radiomics Signature for Predicting the Prognosis of Colorectal Cancer

Du Cai[1,2†], Xin Duan[1,2†], Wei Wang[3†], Ze-Ping Huang[1,2], Qiqi Zhu[1,2], Min-Er Zhong[1,2], Min-Yi Lv[1,2], Cheng-Hang Li[1,2], Wei-Bin Kou[1,2], Xiao-Jian Wu[1,2] and Feng Gao[1,2*] on behalf of COCC Working Group[‡]*

[1] Department of Colorectal Surgery, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, [2] Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Supported by National Key Clinical Discipline, Guangdong Institute of Gastroenterology, Guangzhou, China, [3] Department of Gynecology, Huzhou Maternity & Child Health Care Hospital, Huzhou, China

OPEN ACCESS

Edited by:
Lihua Li,
Hangzhou Dianzi University, China

Reviewed by:
Xiaoke Hao,
Hebei University of Technology, China
Jun Cheng,
Shenzhen University, China

***Correspondence:**
Xiao-Jian Wu
wuxjian@mail.sysu.edu.cn
Feng Gao
gaof57@mail.sysu.edu.cn

[†] *These authors have contributed equally to this work*

[‡] *The list of members and their affiliations appears at the end of the paper.*

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *04 October 2020*
Accepted: *07 December 2020*
Published: *07 January 2021*

Citation:
Cai D, Duan X, Wang W, Huang Z-P, Zhu Q, Zhong M-E, Lv M-Y, Li C-H, Kou W-B, Wu X-J and Gao F (2021) A Metabolism-Related Radiomics Signature for Predicting the Prognosis of Colorectal Cancer. Front. Mol. Biosci. 7:613918. doi: 10.3389/fmolb.2020.613918

Background: Radiomics refers to the extraction of a large amount of image information from medical images, which can provide decision support for clinicians. In this study, we developed and validated a radiomics-based nomogram to predict the prognosis of colorectal cancer (CRC).

Methods: A total of 381 patients with colorectal cancer (primary cohort: $n = 242$; validation cohort: $n = 139$) were enrolled and radiomic features were extracted from the vein phase of preoperative computed tomography (CT). The radiomics score was generated by using the least absolute shrinkage and selection operator algorithm (LASSO). A nomogram was constructed by combining the radiomics score with clinicopathological risk factors for predicting the prognosis of CRC patients. The performance of the nomogram was evaluated by the calibration curve, receiver operating characteristic (ROC) curve and C-index statistics. Functional analysis and correlation analysis were used to explore the underlying association between radiomic feature and the gene-expression patterns.

Results: Five radiomic features were selected to calculate the radiomics score by using the LASSO regression model. The Kaplan-Meier analysis showed that radiomics score was significantly associated with disease-free survival (DFS) [primary cohort: hazard ratio (HR): 5.65, 95% CI: 2.26–14.13, $P < 0.001$; validation cohort: HR: 8.49, 95% CI: 2.05–35.17, $P < 0.001$]. Multivariable analysis confirmed the independent prognostic value of radiomics score (primary cohort: HR: 5.35, 95% CI: 2.14–13.39, $P < 0.001$; validation cohort: HR: 5.19, 95% CI: 1.22–22.00, $P = 0.026$). We incorporated radiomics signature with the TNM stage to build a nomogram, which performed better than TNM stage alone. The C-index of the nomogram achieved 0.74 (0.69–0.80) in the primary cohort and 0.82 (0.77–0.87) in the validation cohort. Functional analysis and correlation analysis found that the radiomic signatures were mainly associated with metabolism related pathways.

Conclusions: The radiomics score derived from the preoperative CT image was an independent prognostic factor and could be a complement to the current staging strategies of colorectal cancer.

Keywords: radiomics, colorectal cancer, prognosis, nomogram, metabolism

INTRODUCTION

Colorectal cancer (CRC) is one of the most common cancers and ranks as the third cause of cancer-related mortality worldwide (Siegel et al., 2020). Even with the recent progress in cancer treatment, the 5 years overall survival of CRC remains <60% (Moghimi-Dehkordi and Safaee, 2012). Traditionally, the treatment regime of colorectal cancer was mainly determined according to clinicopathological factors, such as the TNM stage, tumor size, differentiated grade, which didn't fully consider the heterogeneity of tumors. The emergence of gene expression-based molecular biomarkers has brought hope for the precision treatment of colorectal cancer in the past decade, but the high cost and long detection time limited its clinical application. In recent years, the medical images, which were routinely detected in clinical practice, have emerged to be promising biomarkers for cancer treatment and management.

Radiomic is a multidisciplinary approach concerning the quantification of medical images, like CT and magnetic resonance imaging (MRI). By transforming medical images into high-dimensional quantitative feature data, radiomics have been successfully used in some medical researches, such as tumor genetic analysis, lesions qualitative, curative effect evaluation and prognosis prediction (Kumar et al., 2012; Lambin et al., 2017; Limkin et al., 2017). Typical radiomic features describe the tissue or lesion characteristics, such as tumor shape, tumor texture, which can provide abundant information for tumor assessment. Compared with traditional clinical diagnosis methods, radiomics has the advantages of cheap, non-invasive, and quantifiable.

Several studies have demonstrated that radiomics analysis combined with clinicopathological information can largely contribute to guiding treatment decisions. Huang Y. Q. et al. (2016) developed a radiomic nomogram incorporating clinical risk factors for preoperative prediction of lymph node metastasis in patients with colorectal cancer. Similarly, CT-based radiomics signature in colorectal cancer shows considerable potential discrimination in preoperative staging (Liang et al., 2016). Kim et al. (2015) reported that some distinct features extracted from CT images can significantly discriminate differentiated grades on colorectal adenocarcinoma. In terms of prognostic evaluation, radiomic features are regarded as independent biomarkers for assessing disease-free survival in patients with early NSCLC. By combining with traditional staging systems and other clinicopathological risk factors, the radiomics signature achieved more effective performance (Huang Y. et al., 2016). Farhidzadeh et al. (2016) found that the radiomic features extracted from MRI images in patients with nasopharyngeal carcinoma (NPC) embody the heterogeneity of the tumor, showing high recurrence prediction in two groups patients. In addition, it has been reported that radiomic features based on CT images may correlate with genomics data underlying clinical outcomes (Segal et al., 2007; Aerts et al., 2014). Therefore, analyzing image radiomic features and excavating the hidden biological information have become a promising direction of image biomarkers research.

In this study, we aimed to develop and validate a radiomics-based nomogram to predict the postoperative outcome of colorectal cancer patients. RNA-seq data from the colorectal cancer subproject (COCC, Clinical Omics Study of Colorectal Cancer in China) of the ICGC-ARGO project (The International Cancer Genome Consortium-Accelerating Research in Genomic Oncology) were used to explore the underlying biological interpretation of the radiomic signature.

MATERIALS AND METHODS

Data Collection

A total of 381 patients with colorectal cancer from The Sixth Affiliated Hospital of Sun Yat-sen University were enrolled in this study. Our study was approved by the Medical Ethics Committee of the Sixth Affiliated Hospital of Sun Yat-sen University. Patients admitted during 2007–2011 were assigned to the primary cohort ($n = 242$), while patients admitted during 2012–2015 were assigned to the validation cohort ($n = 139$). Fifty three patients of 381 patients were enrolled in the COCC project, so they have paired image data and RNA sequencing data. All the CT images are DICOM (Digital Imaging and Communications in Medicine) format from the image archiving and storage system of the Six Affiliated Hospital of Sun Yat-sen University. Baseline clinicopathological information containing age, gender, differentiated grade, lymph node metastasis and carcinoembryonic antigen (CEA, normal < 5 ng/ml, abnormal > 5 ng/ml) were also derived from the hospital archives. Region of interest (ROI) was manually delineated on the tumor outline by skilled doctors using the ITK-snap (Version 3.2). A total of 107 radiomic features were generated using pyradiomics (van Griethuysen et al., 2017) package in python 2.7 platform.

Model Construction

Z-score normalization for radiomic features was used to increase comparability. Only features with high intensity were retained for the following analyses. The least absolute shrinkage and selection operator (LASSO) with cox regression was used to construct the radiomic signature and calculate the radiomics score (Rad-score). A nomogram was constructed by incorporating the radiomics score with clinicopathological risk factors. The performance of the nomogram was evaluated by the calibration curve, receiver operating characteristic (ROC) curve and C-index statistics.

Correlating the Radiomic Features With Gene Expression Data

To explore the association between the radiomic features and the underlying biological mechanism, we conducted a correlation analysis between radiomic features and cancer-related hallmarks. DeepCC (Gao F. et al., 2019) was used to calculate the enrichment score of hallmarks of cancer for each patient. The Pearson's correlation coefficient between each hallmark and radiomic feature was calculated. Hallmarks that have a significant correlation with at least one radiomic feature were displayed in the heatmap.

Statistical Analysis

All statistical analyses were performed by R software (version 3.6.1). Time-dependent ROC curve was used to determine the optimal cut-off value of the radiomics score by "survivalROC" (Heagerty et al., 2000), which can divide patients into different risk groups. The R package "glmnet" was used to perform the LASSO-cox regression analysis (Friedman et al., 2010). Kaplan-Meier curves and log-rank tests were used to perform survival analysis. The primary outcome is disease-free survival (DFS). Univariable and multivariable analyses were performed by the cox proportional hazards regression model. Nomogram incorporating Rad-score with clinicopathologic factors was built by the "rms" packages (Harrell, 2016). The two-sided value of $P < 0.05$ was considered statistical significance in all analyses.

RESULT

Features Selection and Model Construction

In the preprocessing step, radiomic features were first scaled with z-score normalization in the primary and validation cohort. The average signal values of each feature in different patients were calculated and compared. We only retained 85 features (80% of 107 features) with higher signal intensity for subsequent modeling. Further, LASSO-cox regression was applied to select 5 features with non-zero coefficients (**Figure 1A**). Radiomics score was calculated by a linear combination of non-zero coefficients, which was multiplied by the 5 features vectors in the primary and validation cohort, respectively. The radiomics scores of all patients were displayed in **Supplementary Table 1**.

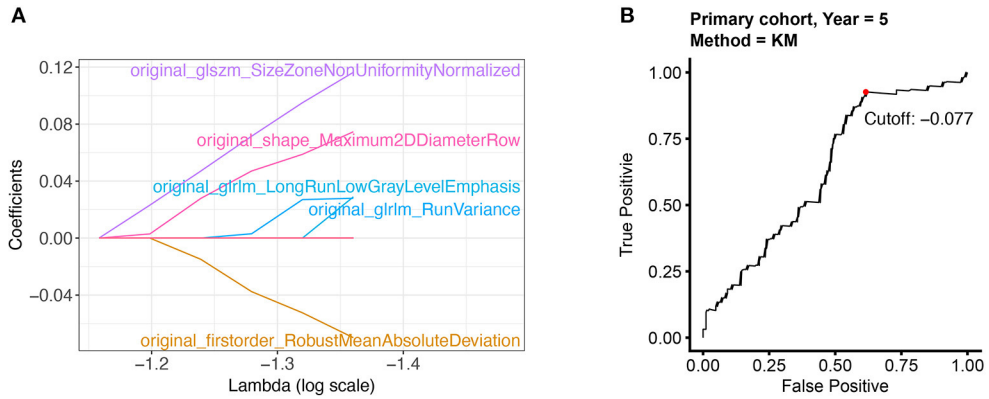


FIGURE 1 | Selecting the radiomic features and the optimal cut-off point. **(A)** 5 radiomic features with non-zero coefficients were selected by using the LASSO algorithm. **(B)** The optimal cut-off point was determined by the time-dependent ROC curve.

TABLE 1 | Baseline characteristic of patients in the primary and validation cohort.

		Primary cohort ($n = 242$)			Validation cohort ($n = 139$)		
	Level	**Low risk**	**High risk**	***P***	**Low risk**	**High risk**	***P***
n		73	169		35	104	
Patients with survival data (*n*)		73	168		35	103	
Sex (%)	F	46 (63.0)	61 (36.1)	<0.001	19 (54.3)	40 (38.5)	0.15
	M	27 (37.0)	108 (63.9)		16 (45.7)	64 (61.5)	
Age [mean (SD)]		71.6 (13.2)	68.8 (14.8)	0.16	65.3 (14.8)	64.0 (12.6)	0.63
Differentiation grade (%)	High	18 (25.0)	45 (31.2)	0.15	10 (28.6)	27 (27.3)	0.54
	Moderate	50 (69.4)	82 (56.9)		19 (54.3)	46 (46.5)	
	Low	4 (5.6)	17 (11.8)		6 (17.1)	26 (26.3)	
TNM stage (%)	I	9 (12.3)	8 (4.7)	0.02	9 (25.7)	7 (6.9)	<0.01
	II	23 (31.5)	68 (40.2)		14 (40.0)	33 (32.4)	
	III	32 (43.8)	54 (32.0)		10 (28.6)	34 (33.3)	
	IV	9 (12.3)	39 (23.1)		2 (5.7)	28 (27.5)	
CEA (%)	Low (<5 ng/ml)	52 (71.2)	110 (65.1)	0.43	27 (79.4)	59 (57.3)	0.04
	High (>5 ng/ml)	21 (28.8)	59 (34.9)		7 (20.6)	44 (42.7)	
Lymph node metastasis (%)	No	34 (46.6)	82 (48.5)	0.89	23 (65.7)	48 (46.2)	0.07
	Yes	39 (53.4)	87 (51.5)		12 (34.3)	56 (53.8)	

The calculation process was presented in the following formula:

$$Rad-score = \text{original\_shape\_Maximum2DDiameterRow} \times 0.075 + \text{original\_firstorder\_RobustMeanAbsoluteDeviation} \times (-0.070) + \text{original\_glrlm\_LongRunLowGrayLevelEmphasis} \times 0.029 + \text{original\_glrlm\_RunVariance} \times 0.028 + \text{original\_glszm\_SizeZoneNonUniformityNormalized} \times 0.116$$

High Radiomics Score Was Associated With Poor Outcome in Colorectal Cancer Patients

The optimal cut-off of Rad-score was determined by the time-dependent ROC curve. Based on the threshold, the patients were divided into the high-risk (>-0.077) and low-risk (<-0.077) groups (**Figure 1B**). Patients' clinical characteristics in the primary and validation cohort were presented in **Table 1**. Survival analysis revealed a significant association between radiomics score and DFS in the primary cohort (HR: 5.65, 95% CI: 2.26–14.13, $P < 0.001$) and validation cohort (HR: 8.49, 95% CI: 2.05–35.17, $P < 0.001$) (**Figure 2**). Patients with a high Rad-score showed a significantly poor outcome. To adjust for the confounding effect of clinicopathological factors, the Rad-score, sex, age, carcinoembryonic antigen (CEA) level, differentiation grade, and TNM stage were added into the multivariable analysis (**Table 2**). Multivariable analysis demonstrated that Rad-score was an independent prognostic predictor of recurrence in the primary cohort (HR: 5.35, 95% CI: 2.14–13.39, $P < 0.001$) and validation cohort (HR: 5.19, 95% CI: 1.22–22.00, $P = 0.026$).

Construction and Performance of the Radiomics Nomogram

Subsequently, based on the results of the multivariable analysis, a nomogram was developed combining the Rad-score and TNM stage (**Figure 3A**). To illustrate the performance of the nomogram prediction, calibration curves were used to evaluate the degree of fitting between the nomogram and the actual

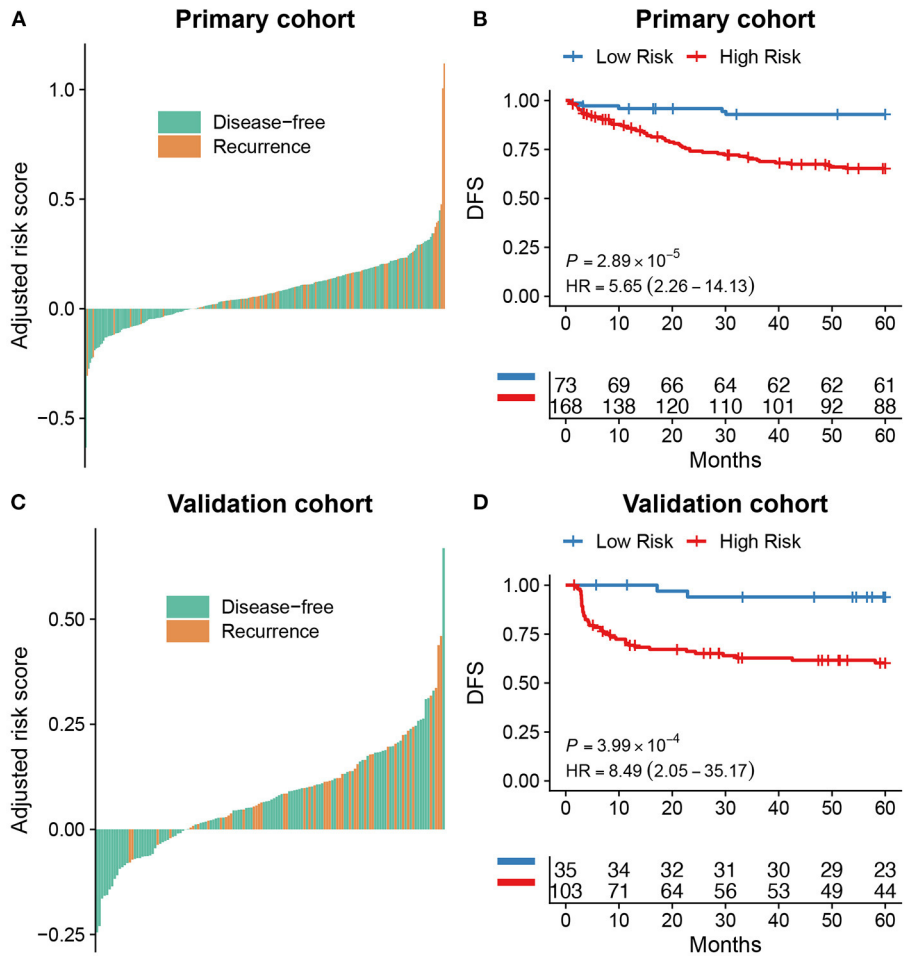


FIGURE 2 | Survival analysis for radiomics score. The distribution of radiomics score in colorectal cancer and its correlation with recurrence status in the primary **(A)** and validation cohort **(C)**. Kaplan-Meier curves showed a significant association between Rad-score groups and disease-free survival (DFS) in the primary cohort **(B)** and validation cohort **(D)**.

TABLE 2 | Univariable and multivariable analysis of clinical factors in the primary and validation cohort.

	Primary cohort				Validation cohort			
	Univariable analysis		Multivariable analysis		Univariable analysis		Multivariable analysis	
	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*
Rad-score	5.65 (2.26–14.13)	<0.001	5.35 (2.14–13.39)	<0.001	8.49 (2.05–35.17)	<0.001	5.19 (1.22–22.00)	0.03
Sex	1.26 (0.74–2.12)	0.39			1.41 (0.75–2.67)	0.28		
Age	1.00 (0.98–1.02)	0.76			0.99 (0.96–1.01)	0.27		
CEA	2.01 (1.20–3.36)	<0.01	1.25 (0.72–2.16)	0.43	2.55 (1.37–4.76)	<0.01	0.96 (0.49–1.86)	0.90
Differentiated grade	1.12 (0.69–1.79)	0.65			1.87 (1.20–2.92)	<0.01	1.80 (1.12–2.89)	0.02
TNM stage	2.51 (1.79–3.52)	< 0.001	2.30 (1.62–3.28)	<0.001	4.82 (3.04–7.63)	<0.001	4.82 (2.86–8.13)	<0.001

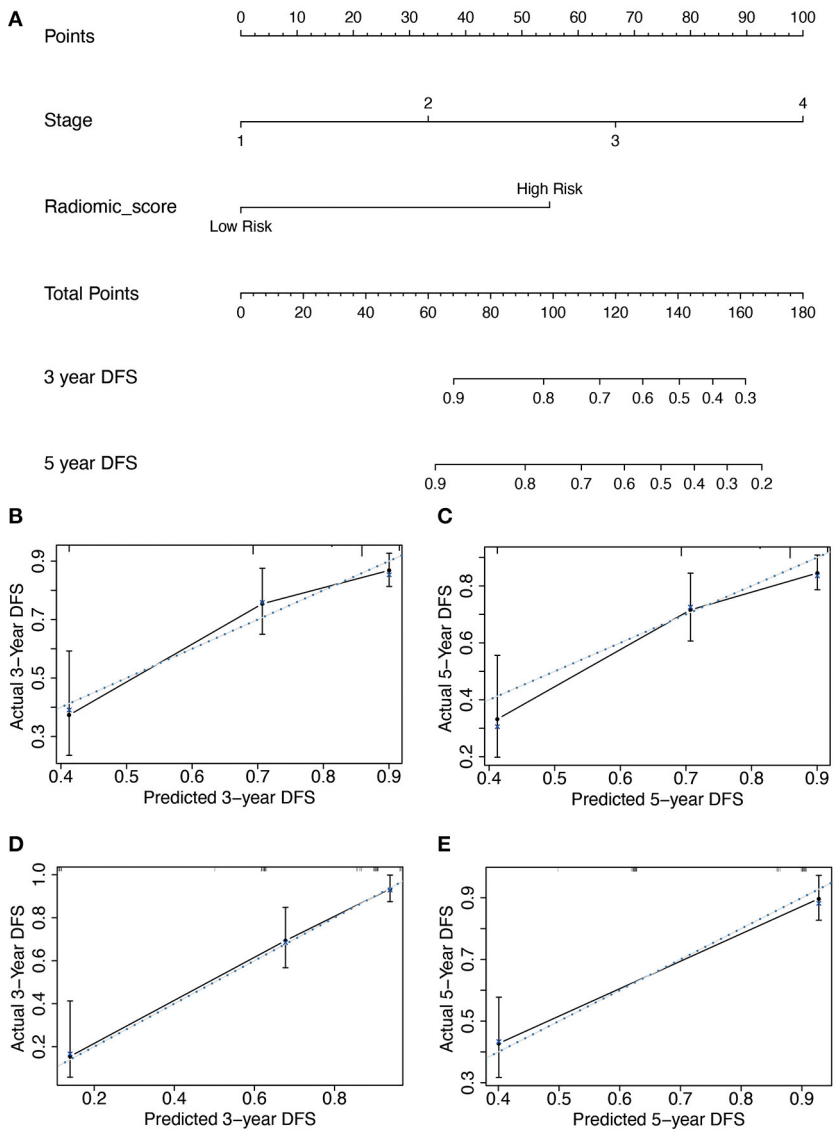


FIGURE 3 | The developed nomogram incorporated Rad-score with TNM stage in the primary cohort. **(A)** The length of the coordinates for the prognostic factor was determined by the coefficient in the regression model. For every patient, the total score was calculated by summing every variable score. The probability of disease-free survival was derived from the mapping relationship between the evaluation results and total score on specified patient survival time. **(B–E)** Calibration curves of Rad-score based nomogram for 3 year DFS and 5 year DFS in the primary **(B,C)** and validation cohort **(D,E)**. The blue dot line is on the diagonal of the figure, indicating a complete fitting between the prognostic model and the actual data. The solid line illustrated the degree of fitting between model prediction and actual survival probability.

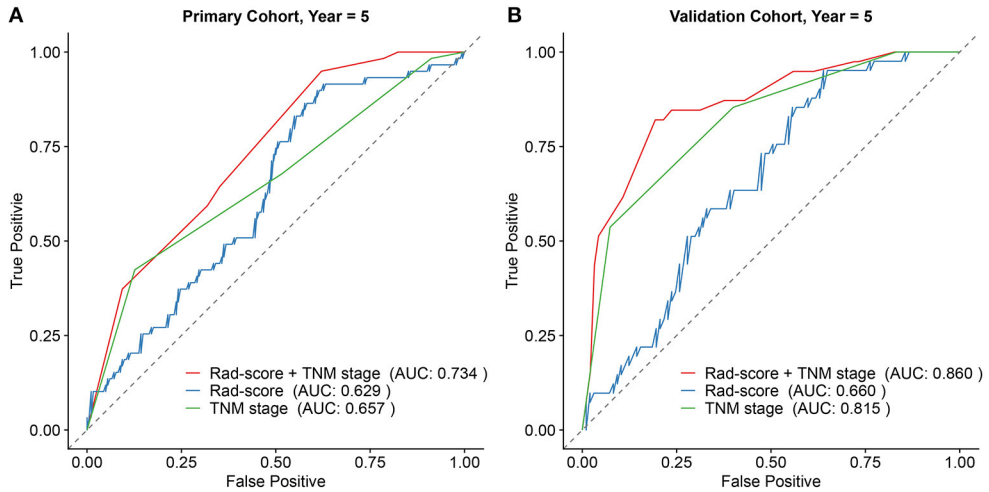


FIGURE 4 | Comparison of the survival discriminative ability between Rad-score based nomogram and clinicopathological factor in the primary cohort **(A)** and validation cohort **(B)**.

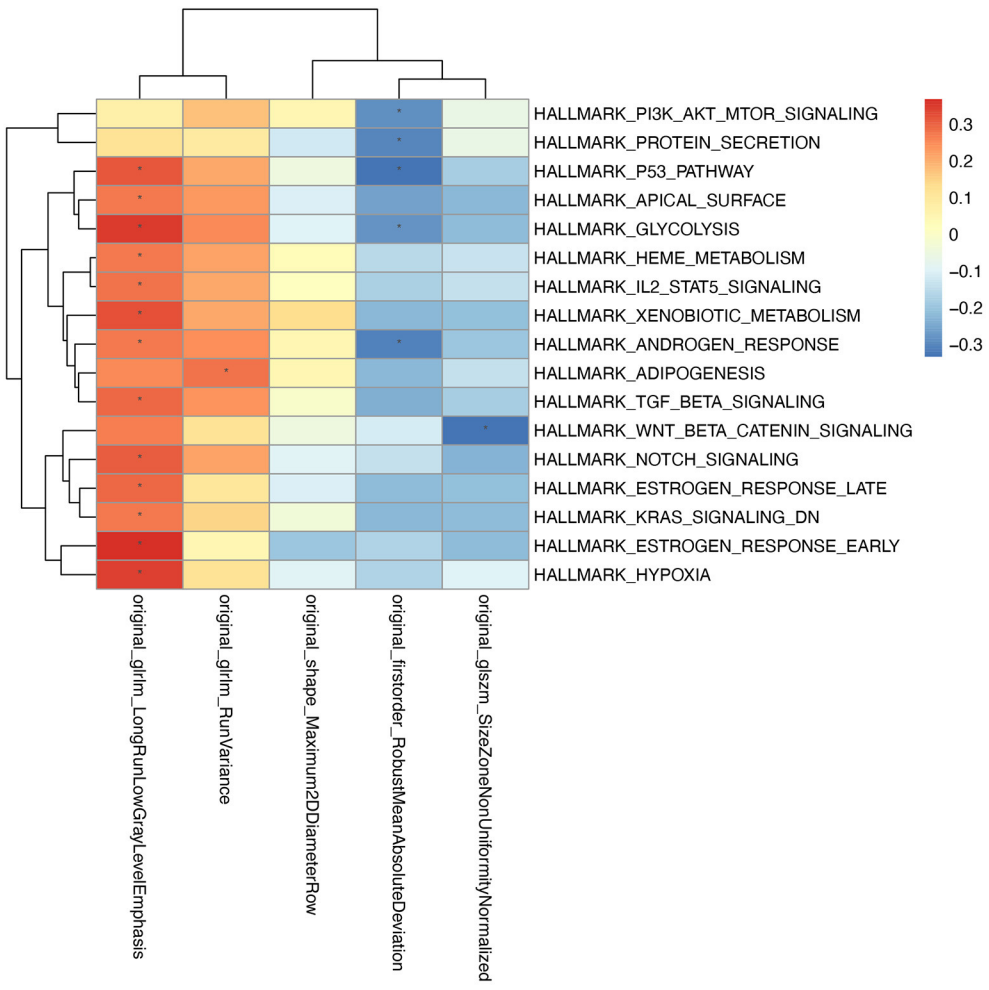


FIGURE 5 | The association between radiomic features and gene expression profiles. For every patient, the enrichment scores of cancer related hallmarks were calculated by DeepCC package. The color block in the heatmap represent the value of Pearson's correlation coefficient between each hallmark and radiomic signature. The hallmarks significantly associated with radiomic features ($P < 0.05$) are marked with "*" in the heatmap.

outcome of patients. The results showed that our nomogram showed good concordance between the predictive and actual survival probability in the primary (**Figures 3B,C**) and the validation cohort (**Figures 3D,E**). The C-index of the nomogram achieved 0.74 (0.69–0.80) in the primary cohort and 0.82 (0.77–0.87) in the validation cohort. To further confirm the effectiveness of the nomogram, we applied the receiver operating characteristic (ROC) to evaluate the discriminative ability of the nomogram for the 5 year DFS. The results showed that the area under curve (AUC) values of Rad-score incorporating the TNM stage reached 0.734 and 0.86 in the primary and validation cohort, respectively, outperforming the result of using the TNM stage alone (**Figure 4**).

Radiomics Features Were Mainly Associated Metabolism-Related Pathway

To explore the association between radiomic features and the underlying biology mechanism, we performed the correlation analysis between the enrichment score of hallmarks and the 5 radiomics features. Gene expression data from 53 patients who have paired image data and RNA sequencing data were used to calculate the enrichment score of hallmarks by DeepCC. The pathways were selected according to the significant association with the radiomics signatures (**Figure 5**). Typically, the radiomics signatures showed significant enrichment in some metabolic pathways, such as protein secretion, glycolysis, heme metabolism, xenobiotic metabolism, adipogenesis.

DISCUSSION

Medical image analysis is a popular issue for precision therapy, which provides non-invasive information for clinical practice and treatment guidance. However, traditional medical image analysis can only find low throughput features or qualitative information manually by radiologists. Recent progress in machine learning enables researchers to extract high dimensional data quickly and quantitatively by radiomics. In this study, we used the radiomic features extracted from the CT image to predict the outcome of CRC patients. Survival analysis showed that high radiomics score was significantly associated with poor outcomes. Univariable and multivariable analyses confirmed the independent prognostic value of radiomic signature. Subsequently, the radiomics based nomogram was developed to predict the DFS, which showed better performance than using the TNM stage alone. Correlation analysis with gene expression profiles revealed that radiomic signature was mainly associated with metabolism-related pathways. Taken together, our results suggested that radiomic signature could be a supplement to the TNM stage for risk stratification of CRC patients.

Although the traditional gene expression-based molecular biomarkers have achieved good performance in many risks predicting tasks of colorectal cancer, there are still some difficulties that limit its clinical application (Walther et al., 2009; Kandimalla et al., 2018, 2019). Genetic test not only requires additional cost and time but also depend on the postoperative detection on pathological samples, which may limit the preoperative treatment intervention. These problems can be avoided by using medical image-based biomarkers. Recent progress in deep learning has generated a series of the image-based model with high accuracy and good performance (Kather et al., 2019; Lu et al., 2020; Skrede et al., 2020). However, a tricky problem of deep learning-based image model is the insufficiency of interpretation, which may raise concerns about its safety and limit its clinical application (Gordon et al., 2019). In contrast, radiomics is more interpretable and less dependent on sample size, which makes it easier to transform into clinical practice. Several studies have successfully use radiomics for individualized risk prediction of colorectal cancer (Liu et al., 2017, 2020). Furthermore, integration analysis of radiomics and gene expression profiles can provide deeper biological interpretation. Our results showed that the radiomics signatures showed significant enrichment in some metabolic pathways, which is an important mechanism for colorectal cancer initiation and progression (Satoh et al., 2017; Gao X. et al., 2019; Tang et al., 2019; La Vecchia and Sebastián, 2020). This indicated that the change of tumor metabolic status may cause morphological change on the image, which could be captured by radiomics features.

Our study not only established a robust radiomics-based nomogram for prognosis prediction of CRC, but also provided biological interpretation by correlation with gene expression profiles. However, there are still some limitations in our study. For example, the image and clinical data are collected from a single center, which may challenge the generalization of our model. Besides, as a retrospective study, the evidence level might be not enough. Prospective multicenter validation would be needed in future studies.

In conclusion, we proposed the Rad-score extracted from CT images as an independent prognostic factor for colorectal cancer. We incorporated Rad-score with the TNM stage to build a nomogram, which outperformed than TNM stage alone, indicating that the Rad-score can be complementary to the current staging strategies of CRC patients. As a non-invasive biomarker, our radiomics-based model can also provide a way of preoperative evaluation, which is helpful for clinical intervention.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are available upon request from the corresponding authors (FG). The image data are not publicly available because they contain information that could compromise patient privacy. Gene expression data of COCC project are not publicly available currently and will be released by ICGC after milestone.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Medical Ethics Committee of the Sixth Affiliated Hospital of Sun Yat-sen University. The ethics

committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

FG and X-JW designed this study. DC, XD, and WW wrote the paper. DC analyzed and interpreted the data and drew the figures. XD extracted the radiomics features. WW, FG, and X-JW revised the paper. Z-PH, QZ, and M-EZ delineated the region of interest. M-YL, C-HL, and W-BK collected and cleaned up the clinical data. COCC Working Group provided assistance for data generation or analysis. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 82002221, FG), the Fundamental Research Funds for the Central Universities (No. 20ykpy05, FG), the Sun Yat-sen University 100 Top Talent Scholars Program—China (No. P20190217202203617, FG), National Natural Science Foundation of China (No. 81972212, X-JW), Natural Science Foundation of Guangdong Province, China (No. 2019A1515010063, X-JW).

ACKNOWLEDGMENTS

The authors appreciated everyone who contributed to this article.

COCC WORKING GROUP

Xin-Juan Fan, Department of Pathology, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China; Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Supported by National Key Clinical Discipline, Guangdong Institute of Gastroenterology, Guangzhou, China

Xuan-Hui Liu, Department of Colorectal Surgery, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China; Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Supported by National Key Clinical Discipline, Guangdong Institute of Gastroenterology, Guangzhou, China

Jia Ke, Department of Colorectal Surgery, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China; Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, Supported by National Key Clinical Discipline, Guangdong Institute of Gastroenterology, Guangzhou, China

Kui Wu, BGI-Shenzhen, Shenzhen, China

Lan-Xin Zhou, BGI-Shenzhen, Shenzhen, China

Shu-Zhen Luo, BGI-Shenzhen, Shenzhen, China

Dong-Bing Liu, BGI-Shenzhen, Shenzhen, China

Xin Wang, Department of Biomedical Sciences, City University of Hong Kong, Hong Kong, China

Tan Wu, Department of Biomedical Sciences, City University of Hong Kong, Hong Kong, China

Zhong-Xu Zhu, Department of Biomedical Sciences, City University of Hong Kong, Hong Kong, China

Rong-Hui He, Hangzhou YITU Healthcare Technology Co., Ltd, Hangzhou, China

Lan-Xuan Liu, Hangzhou YITU Healthcare Technology Co., Ltd, Hangzhou, China

Ming Li, Hangzhou YITU Healthcare Technology Co., Ltd, Hangzhou, China

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.613918/full#supplementary-material

REFERENCES

Aerts, H. J., Velazquez, E. R., Leijenaar, R. T., Parmar, C., Grossmann, P., Carvalho, S., et al. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* 5:4006. doi: 10.1038/ncomms5006

Farhidzadeh, H., Kim, J. Y., Scott, J. G., Goldgof, D. B., Hall, L. O., and Harrison, L. B. (Eds.). (2016). "Classification of progression free survival with nasopharyngeal carcinoma tumors," in *Medical Imaging 2016: Computer-Aided Diagnosis* (San Diego, CA: International Society for Optics and Photonics).

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi: 10.18637/jss.v033.i01

Gao, F., Wang, W., Tan, M., Zhu, L., Zhang, Y., Fessler, E., et al. (2019). DeepCC: a novel deep learning-based framework for cancer molecular subtype classification. *Oncogenesis* 8:44. doi: 10.1038/s41389-019-0157-8

Gao, X., Sanderson, S. M., Dai, Z., Reid, M. A., Cooper, D. E., Lu, M., et al. (2019). Dietary methionine influences therapy in mouse cancer models and alters human metabolism. *Nature* 572, 397–401. doi: 10.1038/s41586-019-1437-3

Gordon, L., Grantcharov, T., and Rudzicz, F. (2019). Explainable artificial intelligence for safe intraoperative decision support. *JAMA Surg.* 154, 1064–1065. doi: 10.1001/jamasurg.2019.2821

Harrell, F. E. Jr. (2016). *rms: Regression Modeling Strategies*. R package version, 5.

Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56, 337–344. doi: 10.1111/j.0006-341X.2000.00337.x

Huang, Y., Liu, Z., He, L., Chen, X., Pan, D., Ma, Z., et al. (2016). Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology* 281, 947–957. doi: 10.1148/radiol.2016152234

Huang, Y. Q., Liang, C. H., He, L., Tian, J., Liang, C. S., Chen, X., et al. (2016). Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J. Clin. Oncol.* 34, 2157–2164. doi: 10.1200/JCO.2015.65.9128

Kandimalla, R., Gao, F., Matsuyama, T., Ishikawa, T., Uetake, H., Takahashi, N., et al. (2018). Genome-wide discovery and identification of a novel miRNA signature for recurrence prediction in stage II and III colorectal cancer. *Clin. Cancer Res.* 24, 3867–3877. doi: 10.1158/1078-0432.CCR-17-3236

Kandimalla, R., Ozawa, T., Gao, F., Wang, X., and Goel, A. (2019). Gene expression signature in surgical tissues and endoscopic biopsies identifies high-risk T1 colorectal cancers. *Gastroenterology* 156, 2338–2341.e3. doi: 10.1053/j.gastro.2019.02.027

Kather, J. N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C. A., et al. (2019). Predicting survival from colorectal cancer histology slides using

deep learning: a retrospective multicenter study. *PLoS Med*. 16:e1002730. doi: 10.1371/journal.pmed.1002730

Kim, J. E., Lee, J. M., Baek, J. H., Moon, S. K., Kim, S. H., Han, J. K., et al. (2015). Differentiation of poorly differentiated colorectal adenocarcinomas from well- or moderately differentiated colorectal adenocarcinomas at contrast-enhanced multidetector CT. *Abdom. Imaging* 40, 1–10. doi: 10.1007/s00261-014-0176-z

Kumar, V., Gu, Y., Basu, S., Berglund, A., Eschrich, S. A., Schabath, M. B., et al. (2012). Radiomics: the process and the challenges. *Magn. Reson. Imaging* 30, 1234–1248. doi: 10.1016/j.mri.2012.06.010

La Vecchia, S., and Sebastián, C. (2020). Metabolic pathways regulating colorectal cancer initiation and progression. *Semin. Cell Dev. Biol*. 98, 63–70. doi: 10.1016/j.semcdb.2019.05.018

Lambin, P., Leijenaar, R. T. H., Deist, T. M., Peerlings, J., de Jong, E. E. C., van Timmeren, J., et al. (2017). Radiomics: the bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol*. 14, 749–762. doi: 10.1038/nrclinonc.2017.141

Liang, C., Huang, Y., He, L., Chen, X., Ma, Z., Dong, D., et al. (2016). The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer. *Oncotarget* 7, 31401–31412. doi: 10.18632/oncotarget.8919

Limkin, E. J., Sun, R., Dercle, L., Zacharaki, E. I., Robert, C., Reuz,é S., et al. (2017). Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann. Oncol*. 28, 1191–1206. doi: 10.1093/annonc/mdx034

Liu, Z., Meng, X., Zhang, H., Li, Z., Liu, J., Sun, K., et al. (2020). Predicting distant metastasis and chemotherapy benefit in locally advanced rectal cancer. *Nat. Commun*. 11:4308. doi: 10.1038/s41467-020-18162-9

Liu, Z., Zhang, X. Y., Shi, Y. J., Wang, L., Zhu, H. T., Tang, Z., et al. (2017). Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin. Cancer Res*. 23, 7253–7262. doi: 10.1158/1078-0432.CCR-17-1038

Lu, M. T., Raghu, V. K., Mayrhofer, T., Aerts, H., and Hoffmann, U. (2020). Deep learning using chest radiographs to identify high-risk smokers for lung cancer screening computed tomography: development and validation of a prediction model. *Ann. Intern. Med*. 173, 704–713. doi: 10.7326/M20-1868

Moghimi-Dehkordi, B., and Safaee, A. (2012). An overview of colorectal cancer survival rates and prognosis in Asia. *World J. Gastrointest. Oncol*. 4, 71–75. doi: 10.4251/wjgo.v4.i4.71

Satoh, K., Yachida, S., Sugimoto, M., Oshima, M., Nakagawa, T., Akamoto, S., et al. (2017). Global metabolic reprogramming of colorectal cancer occurs at adenoma stage and is induced by MYC. *Proc. Natl. Acad. Sci. U.S.A*. 114, E7697–E7706. doi: 10.1073/pnas.1710366114

Segal, E., Sirlin, C. B., Ooi, C., Adler, A. S., Gollub, J., Chen, X., et al. (2007). Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nat. Biotechnol*. 25, 675–680. doi: 10.1038/nbt1306

Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal cancer statistics, 2020. *CA Cancer J. Clin*. 70, 145–164. doi: 10.3322/caac.21601

Skrede, O. J., De Raedt, S., Kleppe, A., Hveem, T. S., Liestøl, K., Maddison, J., et al. (2020). Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet* 395, 350–360. doi: 10.1016/S0140-6736(19)32998-8

Tang, J., Yan, T., Bao, Y., Shen, C., Yu, C., Zhu, X., et al. (2019). LncRNA GLCC1 promotes colorectal carcinogenesis and glucose metabolism by stabilizing c-Myc. *Nat. Commun*. 10:3499. doi: 10.1038/s41467-019-11447-8

van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., et al. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 77, e104–e107. doi: 10.1158/0008-5472.CAN-17-0339

Walther, A., Johnstone, E., Swanton, C., Midgley, R., Tomlinson, I., and Kerr, D. (2009). Genetic prognostic and predictive markers in colorectal cancer. *Nat. Rev. Cancer* 9, 489–499. doi: 10.1038/nrc2645

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fmolb.2020.613347

One-Shot Learning With Attention-Guided Segmentation in Cryo-Electron Tomography

Bo Zhou[1], *Haisu Yu*[2], *Xiangrui Zeng*[2], *Xiaoyan Yang*[2], *Jing Zhang*[3] *and Min Xu*[2]*

[1] Department of Biomedical Engineering, Yale University, New Haven, CT, United States, [2] Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, United States, [3] Computer Science Department, University of California, Irvine, Irvine, CA, United States

Cryo-electron Tomography (cryo-ET) generates 3D visualization of cellular organization that allows biologists to analyze cellular structures in a near-native state with nano resolution. Recently, deep learning methods have demonstrated promising performance in classification and segmentation of macromolecule structures captured by cryo-ET, but training individual deep learning models requires large amounts of manually labeled and segmented data from previously observed classes. To perform classification and segmentation in the wild (i.e., with limited training data and with unseen classes), novel deep learning model needs to be developed to classify and segment unseen macromolecules captured by cryo-ET. In this paper, we develop a one-shot learning framework, called cryo-ET one-shot network (COS-Net), for simultaneous classification of macromolecular structure and generation of the voxel-level 3D segmentation, using only one training sample per class. Our experimental results on 22 macromolecule classes demonstrated that our COS-Net could efficiently classify macromolecular structures with small amounts of samples and produce accurate 3D segmentation at the same time.

Keywords: one shot learning, cryo-ET, macromolecule classification, macromolecular segmentation, attention

OPEN ACCESS

Edited by:
Gaurav Malviya, University of Glasgow, United Kingdom

Reviewed by:
Nikolay Mikhaylovich Borisov, Moscow Institute of Physics and Technology, Russia
Anindya Ghosh, University of Arkansas at Little Rock, United States

***Correspondence:**
Min Xu
mxu1@cs.cmu.edu

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *02 October 2020*
Accepted: *09 December 2020*
Published: *12 January 2021*

Citation:
Zhou B, Yu H, Zeng X, Yang X, Zhang J and Xu M (2021) One-Shot Learning With Attention-Guided Segmentation in Cryo-Electron Tomography. Front. Mol. Biosci. 7:613347. doi: 10.3389/fmolb.2020.613347

1. INTRODUCTION

Cryo-Electron Tomography (cryo-ET) has made possible the observation of cellular organelles and macromolecular structures at nano-meter resolution with native conformations (Lučić et al., 2013). Without disrupting the cell, cryo-ET can visualize both known and unknown cellular structures *in situ*[1] and reveals their spatial and organizational relationships (Oikonomou and Jensen, 2017). Using cryo-ET, it is possible to capture 3D structural information of diverse macromolecular structures inside a given scanned sample.

To analyze the macromolecular structures in cryo-ET, two major subsequent steps need to occur. First, we need to extract the subtomograms[2] and average those that belong to the same macromolecular class, in order to generate a high Signal-to-Noise Ratio (SNR) subtomogram for clear visualization (Zhang, 2019). Second, it is desirable to obtain the macromolecule segmentation in subtomograms to analyze the macromolecular structure parameters such as size distribution and

[1] At their original locations.
[2] Small cubic subvolumes containing one macromolecular structure.

shape. However, the macromolecular structures are highly heterogeneous and contain large quantities of subtomograms. In the past, biologists would spend large amounts of time on a set of tomograms to manually classify and segment subtomograms, but manual annotation is time-consuming and susceptible to the biases of individual biologists. Therefore, it is desirable to automatically classify the extracted subtomograms into subset of macromolecule with similar structure, and automatically generate the macromolecular segmentation.

To automate the process as well as to achieve objective analysis, deep learning methods for classification (Che et al., 2017; Xu et al., 2017; Guo et al., 2018; Zhao et al., 2018; Li et al., 2019, 2020) and segmentation (Chen et al., 2017; Liu et al., 2018; Zhou et al., 2018) have been developed for cryo-ET. Xu et al. (2017) proposed to use Inception3D network and DSRF3D network for cryo-ET subtomogram classification. Then, Chen et al. (2017) further improved the DSRF3D network with residual connection design. Guo et al. (2018) developed a cryo-ET classification model compression technique to reduce the model size while maintaining the classification performance. Zhao et al. (2018) developed a classification model visualization technique for explaining the model's attention on the classified subtomograms. For cryo-ET segmentation, Che et al. (2017) utilized independent 2D CNNs for cryo-ET tomogram components segmentation. Liu et al. (2018) built a SSN3D net for subtomogram segmentation via supervised training with large amounts of segmentation data. While previous deep learning models on cryo-ET improved the accuracy and efficiency on classification and segmentation, there are still two major bottlenecks: (1) as supervised classification methods, previous algorithms still require large amount of manually annotated training data for deep model's training, and (2) previous algorithms need to be trained again to apply to a new dataset of different classes. The open question is: Is it possible to design a generalizable cryo-ET subtomogram classification model that requires only a small reference dataset (such as one manually picked sample in each class) and match the given subtomogram to a reference class, while performing generalizable subtomogram segmentation?

Inspired by one-shot learning models which aim to learn information about object categories from one, or only a few training images (Fe-Fei et al., 2003; Koch et al., 2015), In this work, we develop a Cryo-ET One-Shot Network (COS-Net) that is able to (1) classify macromolecular structure using only a very small amount of samples, (2) simultaneously segment structural regions in a subtomogram based on the classification network, and (3) be readily and directly applied to classify and segment novel structures without needing to be re-trained. Using our COS-Net, biologists can classify and segment thousands of subtomograms by only manually picking a few representative subtomograms as support classes. When there is a need to classify new subtomogram datasets with novel structures, the support classes can be readily changed to accommodate without the need to train the model again. Moreover, unlike previous one-shot learning and few-shot learning algorithms that only address the classification task, our COS-Net can generate both classification and 3D segmentation with application in 3D imaging data of cryo-ET.

Our COS-Net is a Siamese network with pairs of volume encoders, volume decoders, and feature encoders. Given a support set of subtomograms and a target subtomogram, volume encoders first extract the volume's feature presentations. Then, the feature encoders transform the feature presentations for the next stage: one-shot learning. In the meantime, the volume decoders decode the feature presentations to generate the coarse attention/segmentation of the subtomograms. Our COS-Net with additional attention guidance from segmentation information allows better feature embedding for one-shot learning, and thus could provide better one-shot classification performance. During the test stage, we also developed a customized subtomogram processing pipeline to refine the coarse attention/segmentation from COS-Net based on 3D Conditional Random Field (3D-CRF) (Krähenbühl and Koltun, 2011). Our experimental results demonstrated that our method can effectively classify observed or novel macromolecular structures and produce accurate segmentation mask.

2. METHODS

The general structure of our COS-Net is shown in **Figure 1**. The COS-Net is a Siamese network with two encoding-decoding streams. First, each stream consists of one volume encoder, one volume decoder, and one feature encoder. The volume encoders, volume decoders, and feature encoders shared weights between the dual streams. The design of our volume encoders, volume decoders, and feature encoders are illustrated in **Figure 2** and are discussed in detail in our next section. Denoting the input for the upper stream as X_S that is our support set with dimensions of $N \times K$, where N is the number of classes and K is sample per class, support set X_S consists of N classes of macromolecules with K samples per class. In our one-shot learning scheme, $K = 1$. The upper volume encoder takes the support set X_S as input and generates the latent representation of the support set with:

$$F_{S_1} = \mathcal{P}_{VE}(X_S) \tag{1}$$

where F_{S_1} is the latent representation of the support set X_S and $\mathcal{P}_{VE}$ is the volume encoder function. Then, the support set's latent representations F_{S_1} are simultaneously fed into the volume decoder $\mathcal{P}_{VD}$ and feature encoder $\mathcal{P}_{FE}$:

$$M_S = \mathcal{P}_{VD}(F_{S_1}) \tag{2}$$

$$F_{S_2} = \mathcal{P}_{FE}(F_{S_1}) \tag{3}$$

where M_S is the predicted segmentation of the support set, and F_{S_2} is the feature for next stage one-shot learning. Similarly, denoting the input for the lower stream as X_T that is our target set with dimensions of $1 \times K$, target set X_T consists of 1 classes of macromolecules with K samples per class. In our one-shot learning scheme, $K = 1$. Similarly, the same volume encoder

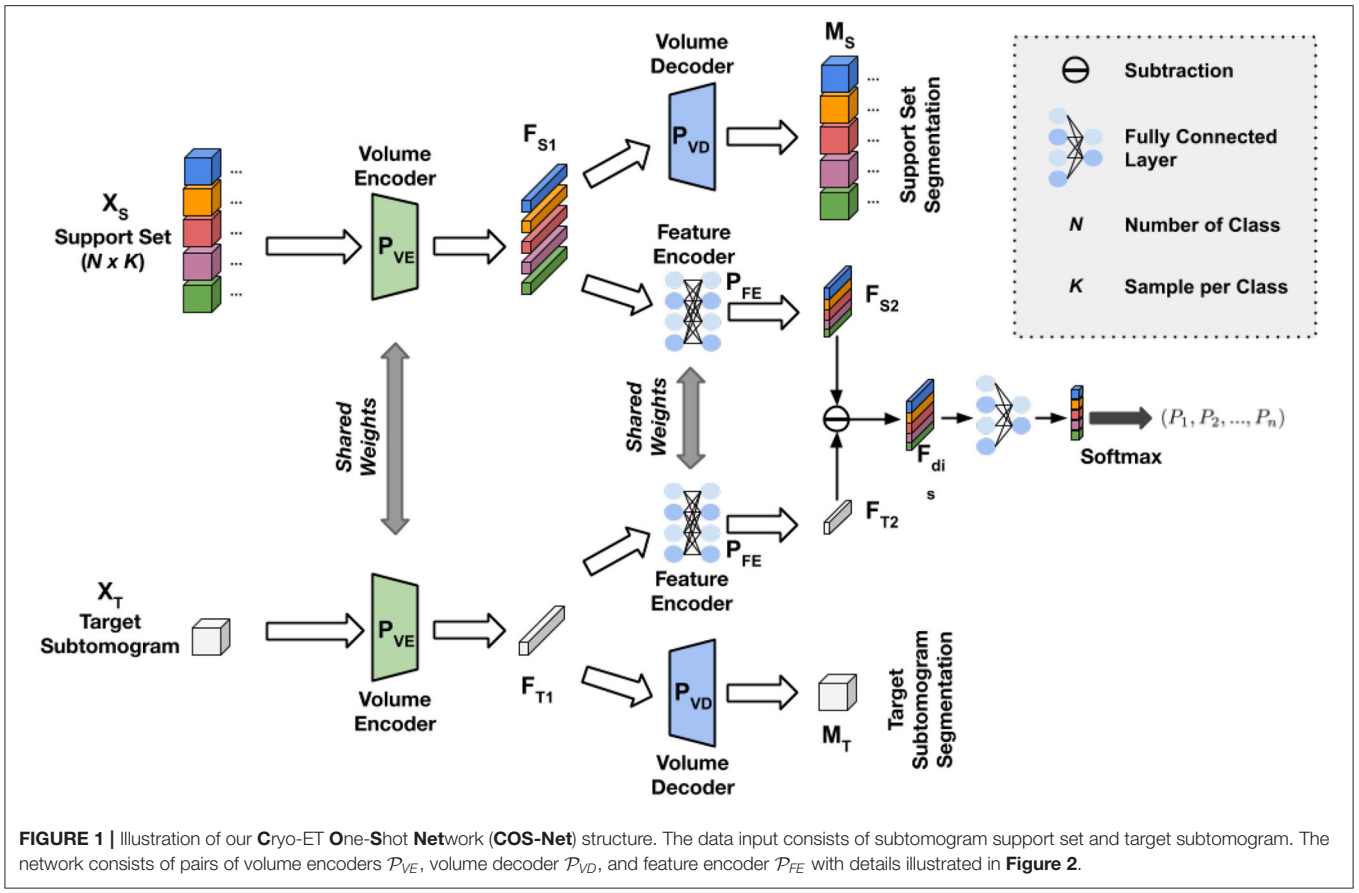


FIGURE 1 | Illustration of our **C**ryo-ET **O**ne-**S**hot **Net**work (**COS-Net**) structure. The data input consists of subtomogram support set and target subtomogram. The network consists of pairs of volume encoders $\mathcal{P}_{VE}$, volume decoder $\mathcal{P}_{VD}$, and feature encoder $\mathcal{P}_{FE}$ with details illustrated in **Figure 2**.

$\mathcal{P}_{VE}$ takes the target set X_T as input and generates the latent representation of the target set with:

$$F_{T_1} = \mathcal{P}_{VE}(X_T) \tag{4}$$

where F_{T_1} is the latent representation of the target set X_T. Then, the target set's latent representations F_{T_1} are simultaneously fed into the shared weights volume decoder $\mathcal{P}_{VD}$ and feature encoder $\mathcal{P}_{FE}$:

$$M_T = \mathcal{P}_{VD}(F_{T_1}) \tag{5}$$

$$F_{T_2} = \mathcal{P}_{FE}(F_{T_1}) \tag{6}$$

where M_T is the predicted segmentation of the target set, and F_{T_2} is the feature for next stage one-shot learning. Given the features F_{S_2} from support set and the features F_{T_2} from target set, we compute the L1 distance between the features to calculate the similarity between the support set features F_{S_2} and the target set features F_{T_2} with:

$$F_{dis} = |F_{S_2} - F_{T_2}| \tag{7}$$

where F_{dis} is the feature distance. F_{dis} is then input into a fully connected layer followed by a softmax function:

$$F_{out} = softmax(\mathcal{P}_{final}(F_{dis})) \tag{8}$$

where F_{out} is the final output with one-shot prediction indicating that the target data matches with which specific class in the support set.

Sub-networks Design: We use a 512 × 512 fully connected layer as our feature encoder. The volume encoder and decoder design are shown in **Figure 2**. Our volume encoder and volume decoder consist of three level of 3D convolution layers. Unlike conventional convolutional encoder and decoder, we concatenate a Dual Squeeze-and-Excitation (DuSE) block at each level's output in order to re-calibrate the features channel-wise and spatial-wise. More specifically, as illustrated in **Figure 2** bottom right, our DuSE block contains two 3D Squeeze-and-Excitation branches for spatial-Squeeze-channel-Excitation (scSE) and channel-Squeeze-spatial-Excitation (csSE), respectively (Hu et al., 2018; Roy et al., 2018).

For scSE, we spatial-wise squeeze the input feature map using global average pooling, where the feature map is formulated as $F = [f_1, f_2, \ldots, f_C]$ here with $f_n \in \mathbb{R}^{H \times W \times D}$ denoting the individual feature channel. We flatten the global average pooling output, generating $v \in \mathbb{R}^C$ with its z-th element:

$$v_z = \frac{1}{H \times W \times D} \sum_{i}^{H} \sum_{j}^{W} \sum_{k}^{D} f_z(i, j, k) \tag{9}$$

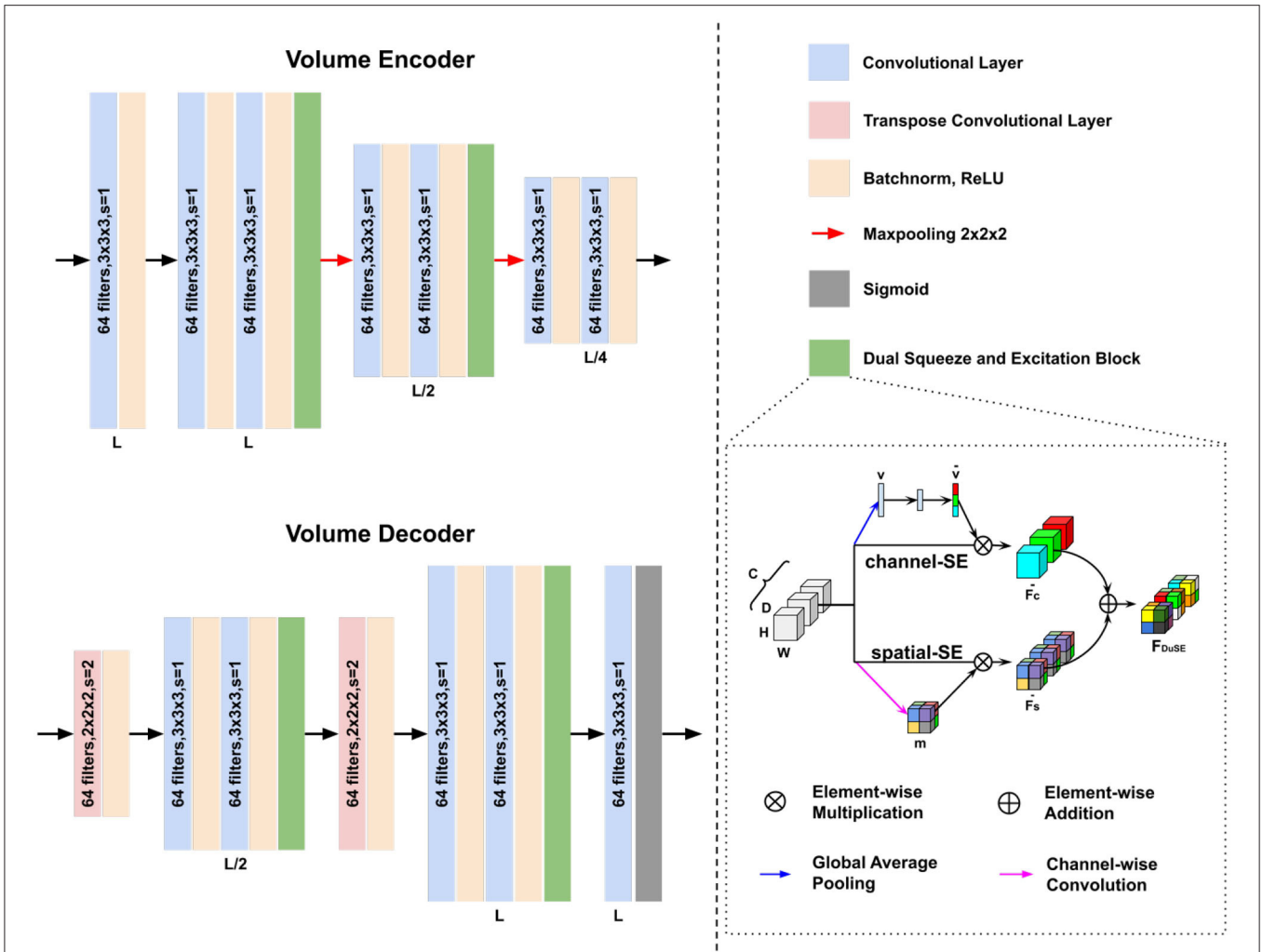


FIGURE 2 | Architectures of our volume encoder and volume decoder in **Figure 1**. The **Du**al **S**queeze-and-**E**xcitation (**DuSE**) block is illustrated on the bottom right.

where vector v embeds the spatial-wise global information. Then, v is feed into two fully connected layers with weights of $w_1 \in \mathbb{R}^{\frac{C}{2} \times C}$ and $w_2 \in \mathbb{R}^{C \times \frac{C}{2}}$, producing the channel-wise calibration vector:

$$\hat{v} = \sigma(w_2 \eta(w_1 v)) \tag{10}$$

where η and σ are the ReLU and Sigmoid activation function, respectively. The calibration vector is applied to the input feature map using channel-wise multiplication, namely channel-Excitation:

$$\hat{F}_{sc} = [f_1 \hat{v}_1, f_2 \hat{v}_2, \ldots, f_C \hat{v}_C] \tag{11}$$

where $\hat{v}_i$ indicates the importance of the i-th feature channel and lies in $[0, 1]$. With scSE embedded into our network, the calibration vector adaptively learns to emphasize the important feature channels while playing down the others.

In csSE, we formulate our feature map as $F = [f^{1,1,1}, \ldots, f^{i,j,k}, \ldots, f^{H,W,D}]$, where $f^{i,j,k} \in \mathbb{R}^C$ indicates the feature at spatial location (i, j, k) with $i \in \{1, \ldots, H\}$, $j \in \{1, \ldots, W\}$, and $k \in \{1, \ldots, D\}$. We channel-wise squeeze the input feature map using a convolutional kernel with weights of $w_3 \in \mathbb{R}^{1 \times 1 \times 1 \times C \times 1}$, generating a volume tensor $m = w_3 \circledast F$ with $m \in \mathbb{R}^{H \times W \times D}$. Each $f^{i,j,k}$ is a linear combination of all feature channel at spatial location (i, j, k). Then, the spatial-wise calibration volume that lies in $[0, 1]$ and can be written as:

$$\hat{m} = \sigma(m) = \sigma(w_3 \circledast F) \tag{12}$$

where σ is the Sigmoid activation function. Applying the calibration volume to the input feature map, we have:

$$\hat{F}_{cs} = [f^{1,1,1}\hat{m}^{1,1,1}, \ldots, f^{i,j,k}\hat{m}^{i,j,k}, \ldots, f^{H,W,D}\hat{m}^{H,W,D}] \tag{13}$$

where calibration parameter of $\hat{m}^{i,j,k}$ provides the relative importance of a spatial information of a given feature map. Similarly, with csSE embedded into our network, the calibration volume learns to stress the most important spatial locations while ignores the irrelevant ones.

Finally, channel-wise calibration and spatial-wise calibration are combined via element-wise addition: $F_{DuSE} = \hat{F}_{sc} + \hat{F}_{cs}$. With the two SE branch fusion, feature at (i, j, k, c) possess high activation only when it receives high activation from both scSE and csSE. Our DuSE encourages the networks to re-calibrate the feature map such that more accurate and relevant feature map can be learned.

Training Strategy and Losses: We design a customized training strategy to train our COS-Net, such that the training procedure matches the inference at test time. Specifically, two support set are randomly generated during the training procedure. Within N classes, the same n classes are randomly sampled for each support set. 1 subtomogram is randomly sampled from these classes to form a *n-way-1-shot* scheme. The ground-truth one-shot classification label is generated by matching the class labels from the two support set, i.e., 1 for matched class label and 0 for unmatched class label.

Our training loss consists of two parts, including a Binary Cross Entropy (BCE) loss for one-shot classification learning and a Dice Similarity Coefficient (DSC) loss for one-shot segmentation. Denoting the ground-truth one-shot classification label as F_{gt}, the BCE loss can be written as:

$$\mathcal{L}_{bce} = -F_{gt} log(F_{out}) - (1 - F_{gt}) log(1 - F_{out}) \tag{14}$$

Denoting the ground-truth subtomogram segmentation for the two support set as M_{gt1} and M_{gt2}, the segmentation loss can be written as:

$$\mathcal{L}_{dsc} = 2 - \frac{2 \times |M_{gt_1} \cap M_{S_1}|}{|M_{gt_1}| + |M_{S_1}|} - \frac{2 \times |M_{gt_2} \cap M_{S_2}|}{|M_{gt_2}| + |M_{S_2}|} \tag{15}$$

where M_{S_1} and M_{S_2} are the predicted segmentation from COS-Net. The total loss thus can be formulated as:

$$\mathcal{L}_{tot} = \mathcal{L}_{dsc} + \mathcal{L}_{bce} \tag{16}$$

In testing, one of the support sets during training can be replaced with the target subtomogram for direct inference.

Attention-guided Segmentation: The segmentation predicted from COS-Net is a probability distribution, which is used for guiding our final segmentation. Specifically, the volume decoder's output is a probability distribution ranging between 0 and 1. We use a 3D Conditional Random Field (CRF) to refine and generate the final 3D subtomogram segmentation. The CRF aims to optimize the following objective function:

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i,j} \psi_p(x_i, x_j) \tag{17}$$

where ψ_u is the unary potential that encourages the CRF output to be loyal to the probability distribution from the COS-Net. ψ_p is the pairwise potential between label on voxel i and j and can be expanded as:

$$\psi_p = \mu(x_i, x_j) \Bigg[w_1 exp\Bigg(- \frac{|p_i - p_j|^2}{2\sigma_\alpha^2} - \frac{|I_i - I_j|^2}{2\sigma_\beta^2} \Bigg) + w_2 exp\Bigg(- \frac{|p_i - p_j|^2}{2\sigma_\gamma^2} \Bigg) \Bigg] \tag{18}$$

where $\mu(x_i, x_j)$ is the compatibility transformation and depends on the labels x_i and x_j such that $\mu(x_i, x_j) = 1$ if $x_i \neq x_j$, and 0 otherwise. I_i and I_j are the intensity value at voxel location i and j. p_i and p_j are the spatial coordinates of voxel i and j. w_1, w_2, σ_α, σ_β, and σ_γ are learnable parameters for CRF. This term penalizes pixels with similar position p and intensity x but with different label.

3. EXPERIMENTS AND RESULTS

3.1. Data Preparation

We prepared a realistically simulated dataset with known macromolecular structures by reconstructing the tomographic image using the projection images (Pei et al., 2016). The limiting factors of cryo-ET, such as noise, missing wedge, and electron optical factors (Modulation Transfer Function, Contrast Transfer Function) were all properly included. The simulation process mimicked the experimental cellular sample imaging condition and tomographic reconstruction process. We took into account the randomness of macromolecule structural poses. The packed volume containing macromolecular structures were projected to a series of 2D projection images with specified tilt angle steps. The resulting projection images were convolved to include optical factors and then back-projected to obtain the reconstructed 3D simulated tomogram. 22 distinct macromolecular structures are chosen from the Proterin Databank (PDB) with their PDB ID information (Berman et al., 2000) of atomic coordinates and connectivity, and secondary structure assignments. We choose very representative macromolecules such as ribosome (4V4Q), proteasome (3DY4), and RNA polymerase (2GHO), which are well-studied due to their abundance and importance in cellular functions. Each simulated tomogram of $600 \times 600 \times 300$ voxels contains 10,000 randomly distributed macromolecules. Given the true position of these macromolecules inside tomograms, we collected 5,835 subtomograms of size $32 \times 32 \times 32$, belonging to 22 structural classes. The dataset with 22 distinct classes was split into a training set with 14 classes and a test set with 8 classes. Three datasets with different levels of signal-to-noise ratio (SNR) were used, including SNR = ∞, SNR = 1, 000, and SNR = 0.5.

3.2. Classification Results

Table 1 summarizes the one-shot classification performance with different sub-network setup. We evaluated the one-shot classification accuracy under different noise level and various one-shot training schemes. First, comparing the COS-Net with and without volume decoder for guiding the one-shot classification, with volume decoder can significantly improve the classification accuracy for sub-networks with or without DuSE block. For example, using the SNR = 1, 000 dataset, the 2way-1shot COS-Net with DuSE improve the accuracy from 0.928 to 0.939 by adding the volume decoder. Second, comparing the COS-Net with and without DuSE block, adding DuSE block to volume encoder/decoder can also improve the classification accuracy. However, the classification accuracy decreases as the SNR decreases, due to the structural details being degraded by noise. Meanwhile, the classification accuracy also decreases as the number of classes (way) increase.

TABLE 1 | The one-shot classification accuracy on three dataset with three different SNR levels.

Data	Networks	2way-1shot	4way-1shot	6way-1shot	8way-1shot
SNR:∞	SCNN w/o Decoder	0.931	0.763	0.613	0.595
	SCNN w Decoder	0.945	0.798	0.663	0.636
	DuSE-SCNN w/o Decoder	0.934	0.772	0.618	0.603
	DuSE-SCNN w Decoder	0.957	0.831	0.672	0.646
SNR:1000	SCNN w/o Decoder	0.923	0.698	0.493	0.473
	SCNN w Decoder	0.935	0.706	0.493	0.473
	DuSE-SCNN w/o Decoder	0.928	0.701	0.504	0.479
	DuSE-SCNN w Decoder	0.939	0.718	0.534	0.513
SNR:0.5	SCNN w/o Decoder	0.812	0.599	0.501	0.387
	SCNN w Decoder	0.824	0.616	0.502	0.399
	DuSE-SCNN w/o Decoder	0.821	0.614	0.510	0.391
	DuSE-SCNN w Decoder	0.829	0.628	0.513	0.403

2way-1shot, 4way-1shot, 6way-1shot, and 8way-1shot learning scenarios are included. The highest accuracy for each learning scenario is marked in blue.

TABLE 2 | The segmentation results for all eight test classes on SNR = 1,000 dataset.

SCNN	1A1S	1BXR	1EQR	1F1B	1FNT	1GYT	1KPB	1LB3
2way-1shot	0.84 ± 0.07	0.85 ± 0.02	0.86 ± 0.02	0.87 ± 0.01	0.89 ± 0.01	0.84 ± 0.01	0.88 ± 0.01	0.83 ± 0.01
4way-1shot	0.84 ± 0.07	0.85 ± 0.02	0.86 ± 0.02	0.87 ± 0.01	0.90 ± 0.01	0.85 ± 0.01	0.88 ± 0.01	0.84 ± 0.02
6way-1shot	0.85 ± 0.08	0.85 ± 0.02	0.85 ± 0.02	0.87 ± 0.01	0.89 ± 0.01	0.84 ± 0.01	0.87 ± 0.01	0.84 ± 0.01
8way-1shot	0.85 ± 0.07	0.84 ± 0.02	0.86 ± 0.02	0.87 ± 0.01	0.90 ± 0.01	0.85 ± 0.01	0.88 ± 0.01	0.83 ± 0.01
DuSE-SCNN	**1A1S**	**1BXR**	**1EQR**	**1F1B**	**1FNT**	**1GYT**	**1KPB**	**1LB3**
2way-1shot	0.85 ± 0.08	0.85 ± 0.02	0.86 ± 0.02	0.87 ± 0.01	0.90 ± 0.01	0.85 ± 0.01	0.88 ± 0.01	0.85 ± 0.01
4way-1shot	0.85 ± 0.07	0.85 ± 0.02	0.85 ± 0.02	0.87 ± 0.01	0.90 ± 0.01	0.85 ± 0.01	0.88 ± 0.01	0.85 ± 0.01
6way-1shot	0.85 ± 0.08	0.85 ± 0.02	0.86 ± 0.02	0.87 ± 0.01	0.90 ± 0.01	0.85 ± 0.01	0.88 ± 0.01	0.85 ± 0.02
8way-1shot	0.84 ± 0.08	0.85 ± 0.01	0.86 ± 0.02	0.87 ± 0.01	0.90 ± 0.01	0.85 ± 0.01	0.88 ± 0.01	0.85 ± 0.02

The mean±standard deviation DSC are reported in the table. 2way-1shot, 4way-1shot, 6way-1shot, and 8way-1shot learning scenarios are reported at different rows. The macromolecular PDB ID is indicated for each classes.

TABLE 3 | The segmentation results for all eight test classes on SNR = ∞ dataset.

SCNN	1A1S	1BXR	1EQR	1F1B	1FNT	1GYT	1KPB	1LB3
2way-1shot	0.92 ± 0.08	0.94 ± 0.03	0.98 ± 0.02	0.97 ± 0.02	0.97 ± 0.03	0.95 ± 0.03	0.96 ± 0.01	0.97 ± 0.02
4way-1shot	0.92 ± 0.08	0.95 ± 0.03	0.98 ± 0.02	0.97 ± 0.02	0.97 ± 0.02	0.95 ± 0.03	0.96 ± 0.03	0.97 ± 0.03
6way-1shot	0.92 ± 0.08	0.94 ± 0.04	0.98 ± 0.01	0.96 ± 0.02	0.97 ± 0.02	0.95 ± 0.03	0.96 ± 0.01	0.96 ± 0.02
8way-1shot	0.92 ± 0.08	0.94 ± 0.03	0.98 ± 0.02	0.96 ± 0.02	0.97 ± 0.02	0.95 ± 0.02	0.96 ± 0.02	0.96 ± 0.02
DuSE-SCNN	**1A1S**	**1BXR**	**1EQR**	**1F1B**	**1FNT**	**1GYT**	**1KPB**	**1LB3**
2way-1shot	0.92 ± 0.08	0.94 ± 0.03	0.98 ± 0.02	0.97 ± 0.02	0.97 ± 0.02	0.95 ± 0.03	0.96 ± 0.02	0.97 ± 0.02
4way-1shot	0.93 ± 0.07	0.96 ± 0.02	0.98 ± 0.01	0.97 ± 0.02	0.97 ± 0.02	0.95 ± 0.03	0.96 ± 0.02	0.97 ± 0.02
6way-1shot	0.92 ± 0.08	0.95 ± 0.03	0.98 ± 0.02	0.97 ± 0.02	0.97 ± 0.02	0.95 ± 0.02	0.96 ± 0.02	0.96 ± 0.02
8way-1shot	0.92 ± 0.07	0.94 ± 0.03	0.98 ± 0.02	0.96 ± 0.02	0.97 ± 0.02	0.95 ± 0.02	0.96 ± 0.02	0.96 ± 0.03

The mean±standard deviation DSC are reported in the table. 2way-1shot, 4way-1shot, 6way-1shot, and 8way-1shot learning scenarios are reported at different rows. The macromolecular PDB ID is indicated for each classes.

3.3. Segmentation Results

The segmentation performance of our attention-guided segmentation is evaluated using the same test set as in the classification section based on DSC:

$$DSC = \frac{2 \times |M_{gt} \cap M_{pred}|}{|M_{gt}| + |M_{pred}|} \tag{19}$$

where M_{pred} is our generated segmentation, and M_{gt} is the ground-truth segmentation. Segmentation results with different training schemes on SNR = 1,000 dataset are visualized in **Figure 3**. As we can see, our method can generate accurate 3D segmentation that does not rely on unseen classes' pixel-level or image-level training data. It is also worth notice that our method can achieve robust and consistent segmentation performance over different way one shot learning schemes.

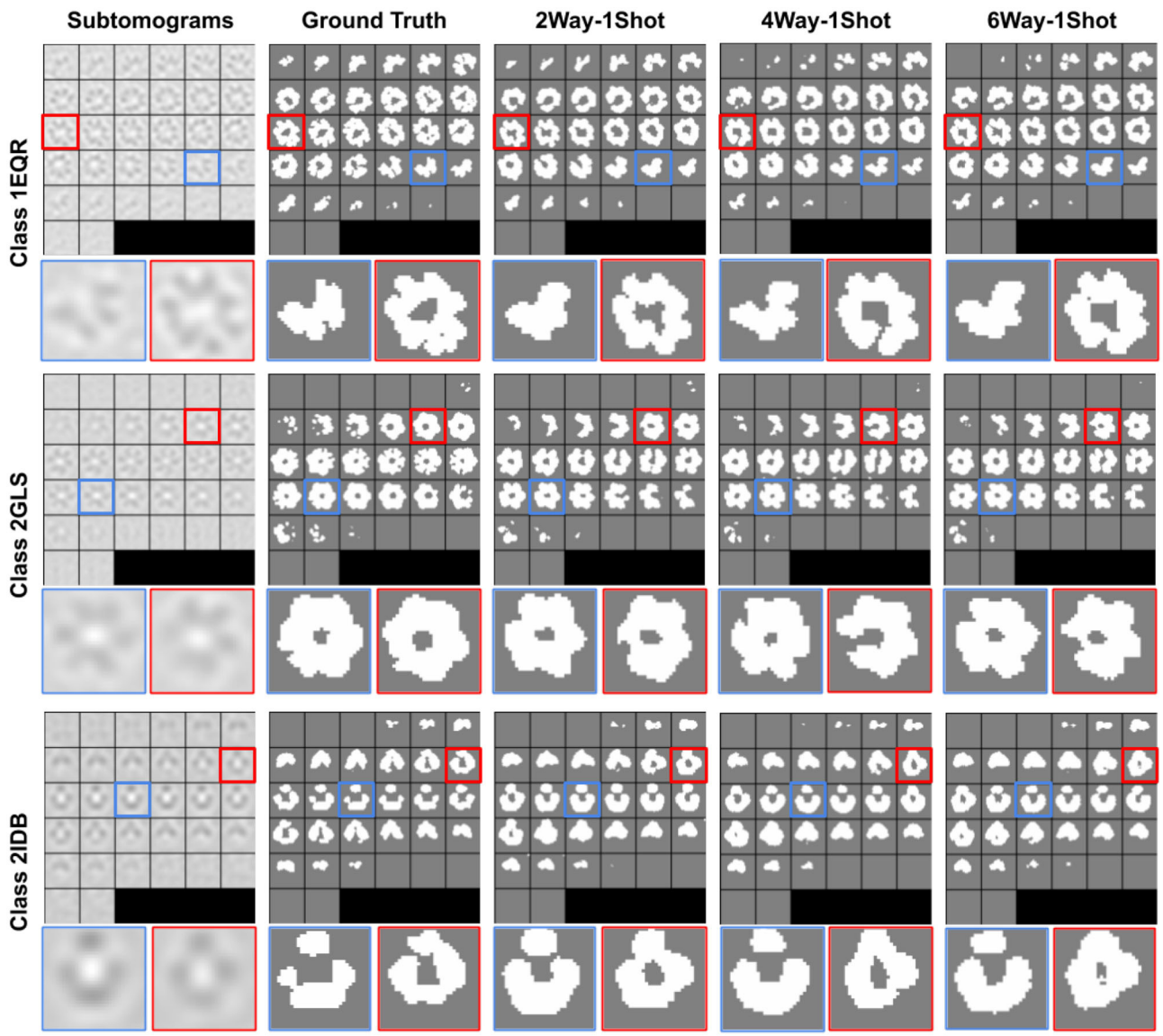


FIGURE 3 | Illustration of segmentation results on all three test classes using COS-Net with DuSESCNN. The macromolecule PDB ID is indicated for each classes on the left. The ground truth segmentation (second column) is compared against COS-Net with 2way-1shot, 4way-1shot, 6way-1shot scenarios from second to fifth column. The enlarged images on selected 2D slices are visualized at the bottom.

Besides, a comparison of segmentation results with and without DuSE block on eight different macromolecule classes is visualized in **Figure 4**. While segmentation with DuSE block does not significantly outperforms segmentation without DuSE block, they both produce reasonable segmentation of macromolecules.

The quantitative results using SNR = 1,000 and SNR = ∞ datasets are summarized in **Tables 2**, **3**, respectively. As we can observe, for all 8 unseen classes, our COS-Net is able to generate reasonable 3D segmentation. For SNR = ∞ data, the DSC of our COS-Net with DuSE are all $>$ 0.92 for all classes, indicating accurate 3D macromolecule segmentation. For SNR = 1,000, the DSC of COS-Net with DuSE are $>$ 0.84. The decrease in segmentation performance is due to the increased noise level that degrades the macromolecule structure details. However, as illustrated in **Figure 3**, our COS-Net can still generate reasonable 3D segmentation for unseen classes.

4. DISCUSSION AND CONCLUSION

In this work, we developed a one-shot learning framework for cryo-ET where simultaneous classification and segmentation

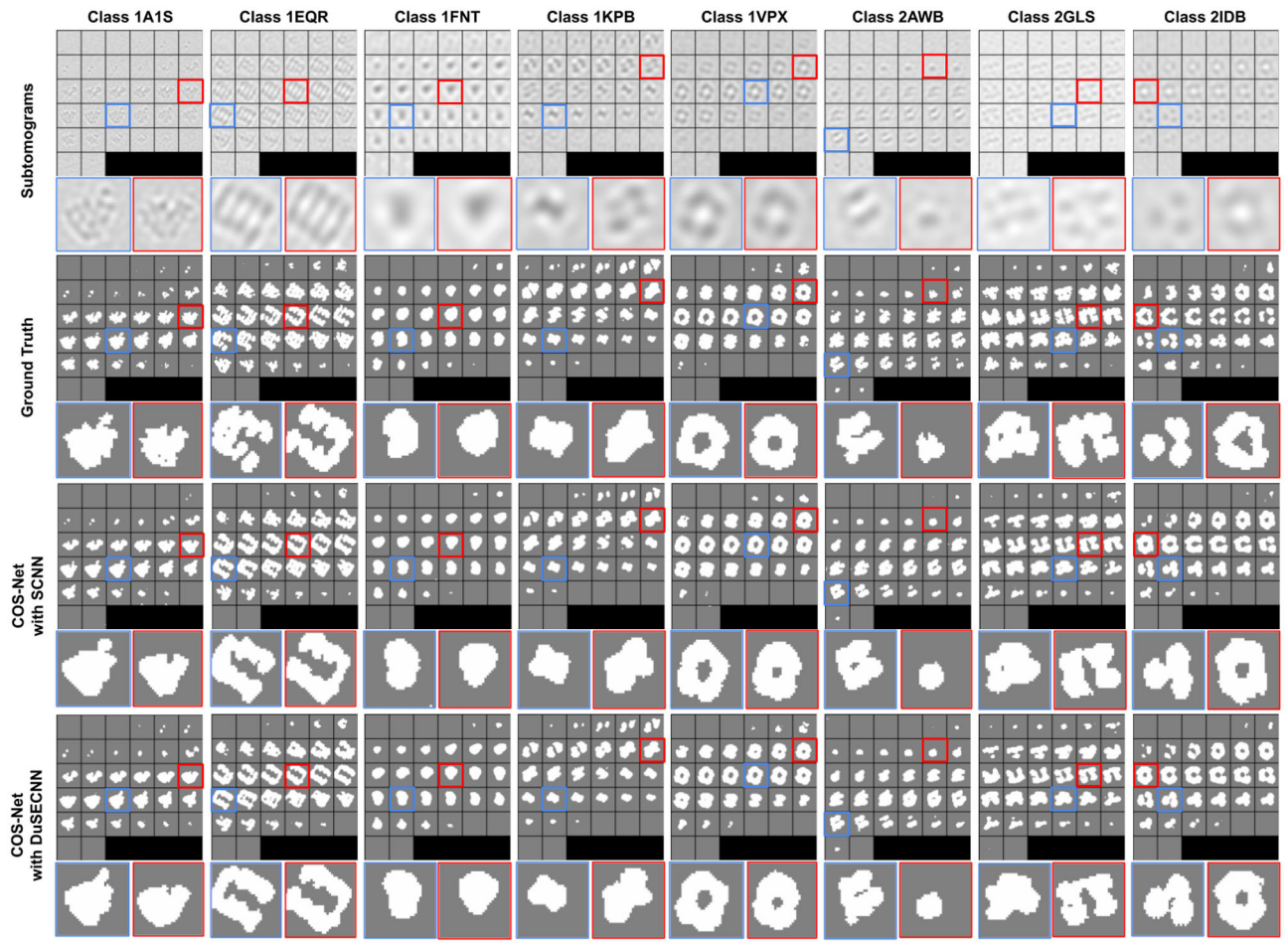


FIGURE 4 | Illustration of segmentation results on all eight test classes using 2way-1shot. The macromolecular PDB ID is indicated for each classes on the top. The ground truth segmentation (second row) is compared against COS-Net with SCNN (third row) and COS-Net with DuSESCNN (fourth row). The enlarged images on selected 2D slices are visualized at the bottom.

can be performed for seen or unseen macromolecule subtomograms. Specifically, we developed a COS-Net to learn the class matching between a support set consisting of multiple classes with only 1 sample per class and a target subtomogram. In COS-Net, the segmentation attention is utilized to better guide the one-shot classification. In the mean time, the volume decoder of COS-Net allows us to generate the coarse segmentation of the macromolecule in the subtomogram. Then, 3D CRF is utilized to refine the 3D macromolecule segmentation from COS-Net.

We demonstrated the successful application of our COS-Net on a cryo-ET dataset consisting of 22 macromolecule classes. First, our method demonstrated accurate one-shot classification performance over dataset with different noise levels. Even with SNR as low as 0.5, the classification accuracy is over 0.8 in a 2way-1shot classification scheme. As compared to previous supervised cryo-ET classification methods with classification accuracy of about 0.9, our method is able to achieve comparable performance without using large-scale high-quality labeled data (Liu et al., 2018; Che et al., 2019). Second, our method can produce high-quality 3D segmentation for unseen macromolecules under different one-shot classification schemes. As we can observe in **Table 3**, our COS-Net can produce 3D segmentation with DSC> 0.84 on all test macromolecules over all one-shot schemes. As compared to previous supervised segmentation methods, our segmentation performance is comparable to these supervised cryo-ET segmentation models with DSC of about 0.88, which require segmentation ground truth on seen macromolecule classes for training (Liu et al., 2018; Che et al., 2019). Therefore, our method provides a solution of both accurate classification and segmentation for unseen macromolecule classes.

The presented work can potentially be further improved from the following perspectives. First of all, the classification accuracy decreases as the number of classes in the support set increases. As more classes are involved in the class matching procedure and only one sample is used for each

classes, the classification difficulty will naturally increase. However, our COS-Net can be extended from one-shot to few-shot if more samples are available for each class, and this strategy could potentially improve the classification accuracy. Moreover, the macromolecule alignment is not considered in the current one-shot classification pipeline. The macromolecule in the support set and target set may not be aligned, i.e., they have different orientations before feeding into our network, which could potentially decrease the classification accuracy. Subtomogram pre-processing by alignment of macromolecule in subtomograms could potentially further improve our classification accuracy and will be a focus in our future work (Lü et al., 2019; Zeng and Xu, 2020). Second, the cryo-ET imaging data is reconstructed from limited angle conditions. The subtomogram image quality could be degraded by the limited angle reconstruction artifacts and potentially impact the downstream COS-Net's performance. Deep learning based limited angle reconstruction algorithms could be incorporated to mitigate these artifacts and potentially further improve our performance (Zhou et al., 2019, 2020). Third, our study is performed based on realistically simulated cryo-ET dataset with sufficient amounts of macromolecule classes for one-shot learning studies. Currently, real cryo-ET data does not provide sufficient amounts of classes for one-shot learning studies, and we will include it in our future studies.

In summary, we developed a COS-Net for one-shot classification and segmentation in cryo-ET, which enables the classification and segmentation for unseen macromolecules in the wild. We believe our algorithm is an important step toward the large-scale and systematic *in situ* analysis of macromolecular structure in single cells captured by cryo-ET.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

BZ: conceptualization, methodology, software, visualization, validation, formal analysis, and writing original draft. HY: methodology, software, visualization, validation, formal analysis, and writing original draft. XZ: conceptualization, methodology, and writing—review and editing. XY: software, visualization, validation, and formal analysis. JZ: writing—review and editing, and supervision. MX: conceptualization, methodology, writing—review and editing, and supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by U.S. National Institutes of Health (NIH) grants P41GM103712, R01GM134020, and K01MH123896, U.S. National Science Foundation (NSF) grants DBI-1949629 and IIS-2007595, AMD COVID-19 HPC Fund, and Mark Foundation 19-044-ASP. BZ was supported by the Biomedical Engineering Ph.D. fellowship from Yale University. XZ was supported by a fellowship from Carnegie Mellon University's Center for Machine Learning and Health.

ACKNOWLEDGMENTS

We would like to thank Erica Chiang at Carnegie Mellon University for improving the manuscript.

REFERENCES

Berman, H. M., Bhat, T. N., Bourne, P. E., Feng, Z., Gilliland, G., Weissig, H., et al. (2000). The protein data bank and the challenge of structural genomics. *Nat. Struct. Mol. Biol.* 7, 957–959. doi: 10.1038/80734

Che, C., Lin, R., Zeng, X., Elmaaroufi, K., Galeotti, J., and Xu, M. (2017). Improved deep learning based macromolecules structure classification from electron cryo tomograms. *arXiv preprint* arXiv:1707.04885.

Che, C., Xian, Z., Zeng, X., Gao, X., and Xu, M. (2019). "Domain randomization for macromolecule structure classification and segmentation in electron cyro-tomograms," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (San Diego, CA: IEEE), 6–11.

Chen, M., Dai, W., Sun, S. Y., Jonasch, D., He, C. Y., Schmid, M. F., et al. (2017). Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nat. Methods* 14:983. doi: 10.1038/nmeth.4405

Fe-Fei, L., Fergus, and Perona. (2003). "A bayesian approach to unsupervised one-shot learning of object categories," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on* (Madison, WI: IEEE), 1134–1141.

Guo, J., Zhou, B., Zeng, X., Freyberg, Z., and Xu, M. (2018). "Model compression for faster structural separation of macromolecules captured by cellular electron cryo-tomography," in *International Conference Image Analysis and Recognition* (Póvoa de Varzim: Springer), 144–152.

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 7132–7141.

Koch, G., Zemel, R., and Salakhutdinov, R. (2015). "Siamese neural networks for one-shot image recognition," in *ICML Deep Learning Workshop* (Lille), Vol. 2.

Krähenbühl, P., and Koltun, V. (2011). "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in Neural Information Processing Systems* (Granada), 109–117.

Li, R., Yu, L., Zhou, B., Zeng, X., Wang, Z., Yang, X., et al. (2020). Few-shot learning for classification of novel macromolecular structures in cryo-electron tomograms. *PLoS Comput. Biol.* 16:e1008227. doi: 10.1371/journal.pcbi.1008227

Li, R., Zeng, X., Sigmund, S. E., Lin, R., Zhou, B., Liu, C., et al. (2019). Automatic localization and identification of mitochondria in cellular electron cryo-tomography using faster-rcnn. *BMC Bioinformatics* 20:132. doi: 10.1186/s12859-019-2650-7

Liu, C., Zeng, X., Lin, R., Liang, X., Freyberg, Z., Xing, E., et al. (2018). "Deep learning based supervised semantic segmentation of electron cryo-subtomograms," in *2018 25th IEEE International Conference on Image Processing (ICIP)* (Athens: IEEE), 1578–1582.

Lü, Y., Zeng, X., Zhao, X., Li, S., Li, H., Gao, X., et al. (2019). Fine-grained alignment of cryo-electron subtomograms based on mpi parallel optimization. *BMC Bioinformatics* 20:443. doi: 10.1186/s12859-019-3003-2

Lučić, V., Rigort, A., and Baumeister, W. (2013). Cryo-electron tomography: the challenge of doing structural biology *in situ*. *J. Cell. Biol.* 202, 407–419. doi: 10.1083/jcb.201304193

Oikonomou, C. M., and Jensen, G. J. (2017). Cellular electron cryotomography: toward structural biology *in situ*. *Annu. Rev. Biochem.* 86, 873–896. doi: 10.1146/annurev-biochem-061516-044741

Pei, L., Xu, M., Frazier, Z., and Alber, F. (2016). Simulating cryo electron tomograms of crowded cell cytoplasm for assessment of automated particle picking. *BMC Bioinformatics* 17:405. doi: 10.1186/s12859-016-1283-3

Roy, A. G., Navab, N., and Wachinger, C. (2018). Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. *IEEE Trans. Med. Imaging* 38, 540–549. doi: 10.1109/TMI.2018.2867261

Xu, M., Chai, X., Muthakana, H., Liang, X., Yang, G., Zeev-Ben-Mordehai, T., et al. (2017). Deep learning-based subdivision approach for large scale macromolecules structure recovery from electron cryo tomograms. *Bioinformatics* 33, i13–i22. doi: 10.1093/bioinformatics/btx230

Zeng, X., and Xu, M. (2020). "Gum-net: unsupervised geometric matching for fast and accurate 3d subtomogram image alignment and averaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Seattle, WA), 4073–4084.

Zhang, P. (2019). Advances in cryo-electron tomography and subtomogram averaging and classification. *Curr. Opin. Struct. Biol.* 58, 249–258. doi: 10.1016/j.sbi.2019.05.021

Zhao, G., Zhou, B., Wang, K., Jiang, R., and Xu, M. (2018). "Respond-cam: analyzing deep models for 3d imaging data by visualizations," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Granada: Springer), 485–492.

Zhou, B., Guo, Q., Wang, K., Zeng, X., Gao, X., and Xu, M. (2018). "Feature decomposition based saliency detection in electron cryo-tomograms," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Madrid: IEEE), 2467–2473.

Zhou, B., Lin, X., and Eck, B. (2019). "Limited angle tomography reconstruction: synthetic reconstruction via unsupervised sinogram adaptation," in *International Conference on Information Processing in Medical Imaging* (Hong Kong: Springer), 141–152.

Zhou, B., Zhou, S. K., Duncan, J. S., and Liu, C. (2020). Limited view tomographic reconstruction using a deep recurrent framework with residual dense spatial-channel attention network and sinogram consistency. *arXiv preprint* arXiv:2009.01782.

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 Molecular Biosciences

REVIEW
published: 28 January 2021
doi: 10.3389/fmolb.2020.614258

Artificial Intelligence for the Future Radiology Diagnostic Service

Seong K. Mun, Kenneth H. Wong, Shih-Chung B. Lo, Yanni Li and Shijir Bayarsaikhan*

Arlington Innovation Center:Health Research, Virginia Tech-Washington DC Area, Arlington, VA, United States

Radiology historically has been a leader of digital transformation in healthcare. The introduction of digital imaging systems, picture archiving and communication systems (PACS), and teleradiology transformed radiology services over the past 30 years. Radiology is again at the crossroad for the next generation of transformation, possibly evolving as a one-stop integrated diagnostic service. Artificial intelligence and machine learning promise to offer radiology new powerful new digital tools to facilitate the next transformation. The radiology community has been developing computer-aided diagnosis (CAD) tools based on machine learning (ML) over the past 20 years. Among various AI techniques, deep-learning convolutional neural networks (CNN) and its variants have been widely used in medical image pattern recognition. Since the 1990s, many CAD tools and products have been developed. However, clinical adoption has been slow due to a lack of substantial clinical advantages, difficulties integrating into existing workflow, and uncertain business models. This paper proposes three pathways for AI's role in radiology beyond current CNN based capabilities 1) improve the performance of CAD, 2) improve the productivity of radiology service by AI-assisted workflow, and 3) develop radiomics that integrate the data from radiology, pathology, and genomics to facilitate the emergence of a new integrated diagnostic service.

Keywords: artificial intelligence, radiology, CNN, productivity, integrated diagnostics, workflow

OPEN ACCESS

Edited by:
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Salva Mena-Mollá,
University of Valencia, Spain
Mohammad Shahid,
Children's National Hospital, United States

****Correspondence:***
Seong K. Mun
munsk@vt.edu

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *05 October 2020*
Accepted: *29 December 2020*
Published: *28 January 2021*

Citation:
Mun SK, Wong KH, Lo S-CB, Li Y and Bayarsaikhan S (2021) Artificial Intelligence for the Future Radiology Diagnostic Service. Front. Mol. Biosci. 7:614258. doi: 10.3389/fmolb.2020.614258

INTRODUCTION

Radiology was one of the first specialty in healthcare to adopt digital technology. Since the 1970s, radiology has adopted many new digital imaging modalities such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Computed Radiography (CR), Single Photon Emission Computed Tomography (SPECT), Digital Ultrasound, Digital Mammography and many others. These digital images were initially printed on films for interpretation, sharing, and archiving. As digital technologies for data capture, data storage, image display, and transmission improved, radiology operations began to convert to a filmless digital environment in the late '90s (Mun et al., 1993; Mun et al., 2007). Today, x-ray films are gone, and the PACS manages all radiological images (Alhajeri et al., 2017). This massive investment in digital technology transformed the radiology service and made radiology images ubiquitous throughout all aspects of healthcare (Hricak, 2018). Digital radiological images enabled the development of many new image-guided surgeries and radiation oncology. Radiology became global as teleradiology was the first successful telemedicine application globally (Mun et al., 1998; Mun and Turner, 1999). Teleradiology, often globally, is a significant portion of radiology operations in the US. Radiology services accumulate massive digital images in their archives, some in the cloud, that laid a technological and

human infrastructure for the next digital transformation based on machine learning (ML) and artificial intelligence (AI).

During the mid-'80s, the radiology community began to explore computer aided diagnosis (CAD) as a possible to aid radiologists (Doi, 2007). Since the mid-2010s, there has an overwhelming interest in machine learning techniques in almost all fields involving data classification or analysis. The number of publications using ML has been exponentially increasing from a few thousand per year in the early 2000s to about 35,000 per year in 2018; and nearly 85% were in neural networks, based on Scopus data (Perrault et al., 2019). Several versions of the neural network technique also were used for drug discovery, computational biology, quantum chemistry, autonomous cars, geology, astronomy, and many others.

Many CAD tools were developed in radiology community, with good performance in terms of sensitivity and specificity. However, most of them remained in research labs, and they did not become an integral part of the radiology service. An earlier success in the use of CAD in digital mammography for breast cancer screening generated much excitement in the community for wider clinical adoption of CAD tools. Some speculated that these intelligent systems would soon replace radiologists.

The National Science and Technology Council of the US published a research and development roadmap for medical imaging (Interagency Working Group on Medical Imaging, Committee on Science, and National Science and Technology Council, 2017). The report envisions changes in medical imaging in 4 general areas; 1) patient referral to imaging service, 2) development and use of high-value imaging capability, 3) use of advanced computation and machine learning, and 4) promoting best practice in medical imaging including reorganizing workflows to improve productivity.

In the article, we discuss AI research in medical imaging from a clinical adoption perspective for patient care and suggest several pathways through which AI will be demanded by radiology as it undergoes the next generation digital transformation toward integrated diagnostic service.

Operational Description of Radiology Service

Radiology service is a very complex operation that includes many inefficiencies. There is a great need to improve overall workflow and productivity. The radiology department provides clinical services to referring physicians and patients by managing a complex workflow involving many layers of people, various technology, many types of time-sensitive information. On any given day, a typical radiology department will conduct more than 50 different types of imaging studies covering all body parts using dozens of imaging modalities, including CT, MRI, ultrasound, nuclear medicine, positron emission tomography, and various conventional radiography systems. The acquired images are managed by a picture archiving and communication system (PACS) and radiology information system (RIS) (Boochever, 2004). Orders for imaging studies from the referring physicians are placed by referring physicians based on the patient's medical history and symptoms. A radiologist determines imaging protocol suitable to address the clinical question. When an imaging study is completed at the imaging system, the PACS will collect all images and generate a worklist for each radiologist based on departmental policies/procedures and the radiologist's specialization (Hricak, 2018).

The radiologist's work has three parts: interpreting and analyzing images, generating reports, and providing further consultation for referring physicians and patients (Halsted and Froehle, 2008). In academic departments, training residents and fellows is also a significant responsibility. The interpretation (reading) time varies greatly depending on the types of study. The radiologists are highly skilled and very fast at detecting abnormalities in the image (Forsberg et al., 2017). The reading times in radiology has been steadily increasing. The imaging devices have improved, and they generate increasingly more images per study. For example, the average number of images for a CT exam increased from 82 images in 1999 to 679 images in 2010. For MRI, the numbers increased from 164 images to 570 images, respectively (McDonald et al., 2015). The image interpretation often requires examining previous studies and comparing them to the current study to determine if the patient has gotten better or worse. These comparison analyses are carried out manually by a radiologist, which can be very time-consuming (Schemmel et al., 2016; Doshi et al., 2018).

Once the radiologist has completed their interpretation of the study, the results are generally recorded on a voice recognition system (a component of the PACS/RIS) to generate written reports that become part of the patient record (Boochever, 2004). In some cases, the report is sufficient by itself. In other cases, the report becomes part of a more complex analysis (Kahn et al., 2009). For example, a patient might have multiple radiology exams, blood chemistry analysis, and tissue sampling, all of which need to be combined to create a comprehensive diagnosis.

Most of the previous efforts to apply AI to radiology in terms of CAD so far have focused exclusively on the interpretation of single images or a single series of images. In comparison, this is an important task, only a fraction of how a typical radiologist spends their time. When the image's reading/analysis can become a part of subsequent decision-making, radiologists would participate in the decision making beyond the reports (Brady, 2017). In essence, AI in radiology has historically tackled challenging yet narrow problems (Schemmel et al., 2016; Hosny et al., 2018). We argue that a refocusing of AI onto different aspects of the radiology workflow and medical error reduction will generated more demand and adoption by radiology community. (Dikici et al., 2020; Montagnon et al., 2020).

Clinical Adoption Case Study: Breast Cancer Screening and Diabetic Retinopathy

One has to distinguish between CAD, CADx, and CADe, a family of AI tools based on convolution neural network (CNN). The concept of computed diagnosis (CAD) research in medical imaging has evolved into two distinct clinical applications; computer aided diagnosis (CADx) and computer aided detection (CADe). CADx means the computer provides a diagnosis for physician review. On the other hand, CADe

means the computer highlights the area of concern (i.e., cancer) for further diagnostic evaluation without providing a diagnosis. The CADe is used to screen cancers such as lung cancer or breast cancer for the asymptomatic but higher risk population. If cancer is suspected from the screening study, the patients will undergo a higher precision diagnostic study. The CADe is a much more challenging problem compared with CADx.

We will review the clinical adoption experience of two different use of AI, breast cancer screening and diabetic retinopathy screening to highlight that clinical adoption is a multifaceted issue beyond technical success in laboratories.

CADe in Screening Mammography for Cancer Detection

Screening mammography is an ideal application for CADe because it has to review many cases by a limited number of radiologists trained in mammography. It is a single type of exam, and the most basic output is a simple yes (disease present) or no (disease absent). FDA approved the first CADe for mammography in 1998, but its adoption rate was initially less than 5% (Lehman et al., 2015). However, as breast cancer screening became more popular as part of a government policy to promote women's health, there was a shortage of skilled radiologists for mammography. To address the shortage, which resulted from low reimbursement rates for reading mammograms, the Center for Medicare and Medicaid Services (CMS) allowed higher reimbursement rates for using CADe in screening mammography (Gold et al., 2012). This financial incentive has dramatically increased the adoption of CADe for mammography. In fact, in the US today, most breast cancer screening mammograms are interpreted by radiologists with CADe assistance. It is highly unlikely that additional reimbursements would be allowed to use AI tools in radiology in the future.

The widespread use of mammography CADe and the large number of exams performed each year allowed assessment of CADe's impact on mammographic interpretations' accuracy. One of the largest, a study involving 271 radiologists and 323,973 women between Jan 2003 and Dec 2009, compared reading mammograms with and without CADe. The study concluded that CADe does not improve mammography's diagnostic accuracy (Lehman et al., 2015).

A more recent study by Schaffter and colleagues conducted a crowd-sourced trial on the use of deep learning in digital mammography involving 300,000 mammograms from the US and Sweden. The project had 126 teams from 44 countries to see if they could meet or beat the radiologist's performances. They concluded that AI tools again did not perform better than radiologists (Schaffter et al., 2020).

CADe Use to Screen Diabetic Retinopathy to Prevent Blindness

Diabetic retinopathy is an eye disease when high blood sugar levels of a diabetic patient can cause damage to blood vessels in the retina. Undetected and untreated the patient can become blind. Screening for diabetic retinopathy is an effective way to prevent the blindness. A special camera takes images of the blood vessels in retina. Interpretation of these images requires special expertise, thus it has been a interests of AI community to develop CAD systems.

Recently Google Corporation deployed its AI tool to detect diabetic retinopathy in Thailand. Initial development and testing involved 3,049 patients. In 2018, they deployed the system at 11 clinics in Thailand, involving 7,600 patients (Beede et al., 2020). This large-scale prospective study was halted mainly because of persistent image quality problems. The system performed poorly, mainly due to the variability of retinal scan images obtained at different nurses' different settings (Beede et al., 2020). The success in the lab did not translate well in real-life situations. Initial deployment of an AI system for lung cancer screening experienced a similar situation in dealing with a considerable variability of image quality of real-life clinical cases (Worrell, 2020). Pre-processing of image normalizing before AI application may be an essential step for scalable clinical deployment.

Currently, 46 AI algorithms have approvals from Food and Drug Administrations and Conformité Européenne (CE) (Tadavarthi et al., 2020). The approval process consists of clinical trial demonstration clinical safety and efficacy, often involving receiver operating curve analysis of sensitivity and specificity. The approval, however, does not guarantee successful clinical adoption. Global adoption of these tools is still a few years away. Since there are no insurance reimbursements, except digital mammography in the US, for the use of AI software, return on investment must be assessed based on significant quality improvement or efficiency improvements. AI tools have not shown significant improvements in the quality of diagnosis and operational efficiency for cost savings. Such improvement may require significant changes in the radiology department functions and possible re-configurations of PACS often owned by vendors other than current AI vendors.

The Technology of CNN and Computer-Aided Diagnosis

Artificial intelligence (AI) is the capability of the machine to imitate intelligent human behavior. In contrast, machine learning (ML) is a subfield of AI that allows the machine to learn from data without being explicitly programmed (Soffer et al., 2019). The concept of neural networks emerged from the biologic neuron system. A neural network in the visual cortex can detect the edges of an object seen by the retina. When the receptors' inner parts are activated simultaneously, the cell neutron integrates the signals and transmits an edge detection signal. An artificial neural network (ANN) is composed of interconnected artificial neurons. Each artificial neuron implements a simple classifier model, which outputs a decision signal based on a weighted sum of evidence, and an activation function integrates signals from the neurons. An ANN system can be built with thousands of these basic computing units. The system can be trained by computing these weights using a learning algorithm where pairs of input signals and desired output decisions are presented, mimicking

brain functions. An individual artificial neuron is a simple neural network; however, multilayer perceptron can model complex nonlinear functions. The deep learning (DL) concept is based on the use of multilayer architecture of multilayer perceptron. In medical imaging, the number of layers tends to be in the range of dozen.

The convolution neural network (CNN) consists of a series of convolution layers equivalent to compositional convolution layers with a set of large kernels. In effect, a CNN acts as a feature learning based on spatial features with multiple channels (Lo et al., 1995; Lo et al., 2018b; Lo et al., 2018c). However, common difficulties in traditional CNN approaches for medical imaging can be grouped into three categories; 1) inability to separate normal from ill-defined abnormal structures, 2) inability to differentiate disease patterns, particularly in subtle cases from a broad spectrum of normal structures, and 3) inability to establish an integrated system between compositional and divide-and-conquer models.

Limitation of Current Generic CNN

Many ML open-source packages such as Tensorflow, Keras, Caffe, and others featuring CNN have been widely used. The core algorithms of CNN in all these packages were designed for general image pattern recognition. They were initially developed for the recognition of alphanumerical handwriting. General image pattern recognition relied on essential graphic pattern features (e.g., edges) and orientation-dependent but size-independent in many situations. On the other hand, medical image pattern recognition should rely more on gray intensity distribution and is orientation independent but size-dependent. Also, some users have experienced inconsistent results from the current CNNs and have tried to use many versions converted from the same input as a part of the augmentation strategy to increase the training samples and stability (Lo et al., 1998).

Improving CNN for Medical Imaging

An ordinary CNN using unconstrained kernel weights entirely based on the backpropagation training (Lo et al., 2018d). The use of rotational and translational versions of each input vector as data augmentation was developed by the authors (Lo et al., 1993). However, many investigators reported that the current method requires a long training time and produces unstable results (Lo et al., 2018b; Lo et al., 2018c).

The current CNN software should be redesigned for medical imaging pattern recognition by (i) the use of an activation function without suppression of the composed signal and (ii) the use of symmetric kernels. This is because current activation functions (Relu, Leaky Relu, sigmoid, and Tanh) used in general CNN tools are signal suppression functions (i.e., $df/dx < 1$). When using them multiple times through multiple convolutional layers, only edge patterns with very few gray value features remained in the feature maps at the end of convolution/activation processing for final classification. These are not acceptable intermediate outcomes for many medical images where subtle gray value differences are used for discerning possible disease characteristics (Lo et al., 1995; Lo et al., 1998).

The symmetric kernels within CNN should be used to stabilize the CNN output consistency. The use of kernels with dihedral symmetry of order 8 (Dih4) is an example with a minimum number of free parameters as element coefficients are symmetric with respect to each corner wedges. In other words, elements on other wedges corresponding to the Dih4 symmetric element positions in the wedge would share the same value. The use of symmetric kernels can be expanded to wavelet decomposition. Though it is different from an ordinary convolution process, each compartment's biorthogonal kernels may be different. However, t the absolute value in each element of the kernel is the same. The multi-dimensional wavelet decomposition is made by a one-dimensional convolution process and down sampling one half at a time. The total number of free parameters is much less (the number of elements in 1D kernel plus 1, divided by 2). In effect, kernels to produce low-low (LL) and high-high (HH) compartments are Dih4 transformation-identical (TI) kernels. Kernels to produce low-high (LH) and high-low (HL) compartments are Dih4 TI with an odd number of elements but are anti-symmetric (i.e., 180^o rotation TI) with an even number of elements. Since each compartment is processed through an independent pipeline in the neural network process, for the latter situation, there is still room to make signals from LH and HL be Dih4 TI, if desired. This can be done by inserting a reflective symmetry kernel in each of these two compartment pipelines (Lo et al., 2018d).

With symmetric kernels, such as the dihedral symmetry of order 8, the intermediate results throughout all convolutional layers would be equivariant for original input and 90^o rotation increment as the flipping version. With this equivariant property on all convolutional layers, the CNN would produce identical output for all eight input image versions. In summary, the CNN can be treated as a whole function of an input vector Vi (i.e., a 2D image or 3D volume), and the output vector can be expressed as $Vo = CNN(T[Vi]) = CNN(Vi)$ as long as $T[K] = K$ within each of the CNN convolutional processes, where Vi and Vo are input and output vectors, respectively. K denotes the convolution kernel and T[.] is a transformation function. This equivariant property at all convolutional layers can be extended for the CNN to produce identical output results for any arbitrarily rotated images by merging the convolution processing before the classification section in the CNN. The use of symmetric kernels in the convolutional layers in the CNN would be a more appropriate tool to systematically produce highly stable results (Lo et al., 2018d).

Open Source CNN

There have been increasing concerns about the ethics, ability to explain and transparency of AI technology (Tang et al., 2018; ESR, 2019), especially in healthcare. These concerns are partly due to difficulties in understanding underlying theories, methods, and assumptions used to generate systematic bias results. In this scenario, the use of open-source software (OSS) strategy could help address some of these concerns because, by definition, OSS offers greater transparency of the technology and opportunities for community-based collaboration.

The OSS concept started in the 1980s as a social movement and a philosophy for software development and distribution (Levine and Prietula, 2014). OSS is defined as a software code made available under a legal license in which the copyright holder provides (depending upon the specific terms) various rights to the licensees to study, change, improve and re-distribute the code without any fees. Today there are many different types of OSS licenses depending on the copyright holders' interests and intentions (Fosfuri et al., 2008; Opensource.org, 2020). These licenses range from permissive licenses such as Apache-2.0 to strongly protective licenses such as general public license (GPL). OSS is typically available as-is; however, it can be made into commercial products with additional services such as warranty, training, documentation, and maintenance under various commercial contracts.

Some of the more popular packages include TensorFlow, Keras, PyTorch, Caffe2, and many others. They all have varying strengths and weaknesses, depending on users' needs. Keras and TensorFlow have a common or similar core, but Keras is much easier to use with limited options than TensorFlow. PyTorch is fast and flexible for experimentation, and it is tightly integrated with the Python language. An extensive table of available software with detail comparison can be found at Wikipedia (Multiple-Editor, 2020).

These OSS packages are developed and sponsored by various organizations and individuals for their use cases and applications other than medical imaging. However, the packages are initial starting platforms for imaging research. These open source packages should be optimized to be suitable for meaningful medical imaging research as discussed in the earlier section. Additionally, these open source codes' users should form or join collaborative communities based on shared medical imaging interests.

Application of CNN in Medical Imaging:

For research in supervised learning such as CNN, the success depends on three technical factors: 1) underlying science and technology of the code, 2) learning supervised by subject matter experts, and 3) the quantity/quality of data.

Another crucially important factor in CNN research is imaging expertise, both clinical and physic of imaging (Giger et al., 2008). Unlike common everyday objects in AI research, the research team in medical imaging AI has to understand the clinical significance of images and imaging physic.

Availability of Data and Realistic Mix of Data

There are two data issues: access to a sufficient volume of data and enough data diversity representing a realistic case mix of the clinical operational environment (Yamashita et al., 2018).

The imaging data requirement in radiology is relatively modest, less than 10,000 cases per disease category. In the case of the recent AI tool development for lung cancer screening with CT images, approximately 2,000 cases consisting of 300,000 CT images were sufficient for training, and approximately 300 cases of 45,000 images with about 20% subtle cases tested by more than 10 radiologists were sufficient for an FDA specified clinical trial (Lo et al., 2018a). For different disease types and imaging modalities, these numbers would be different. If the clinical problem to be addressed has many subtle features, the data volume required would be much higher.

For an AI algorithm to be clinically useful, it must be trained on data that appropriately represent the patient population's variance and diseases' presentation. In a routine data collection effort, the majority of available cases show disease patterns, which are considered relatively easy cases. The cases of subtle disease patterns are relatively rare and thus challenging to collect. It is essential to have a mix of subtle cases in the image archive. If one has a disproportionately large number of similar or easy cases, the system will show bias (Lo et al., 1995; Hosny et al., 2018). In supervised learning, algorithms such as CNN learn from labeled data. When the number of categories and/or patterns to be differentiated increases, the required data volumes would increase. The problem of having more dimensions, yet small data volume, can result in overfitting contributing to low generalizability and scalability (Yamashita et al., 2018; Mutasa et al., 2020).

Quality of Data: Image Quality

The performance of current CNN is fragile, dealing with varying image quality. In fact, in any data science project, one can spend a significant amount of effort to "clean" the data. The same is true in imaging. The data must be of sufficient quality and acquired with uniform parameters to make certain that conclusions can be validated. The image quality can vary depending on the time and day of imaging, image protocol, imaging system set up, patient conditions, and clinical practice standards in different departments (Worrell, 2020). While human vision is good at reading through the images of varying qualities, AI tools are generally not (Tang et al., 2018). One important task to produce a systematic image AI performance is image pre-processing, including optimization of image quality, noise reduction, clutter removal, and enhancement of essential features for differentiation. Various AI tools are used to standardize image quality (Zhu et al., 2017; Mazurowski et al., 2019).

A radiology AI tool for screening or diagnosis of a disease is usually comprised of several components: 1) pre-processing such as image normalization, 2) image segmentation or region of interest (ROI) extraction, and 3) potential disease pattern identification and classification. Various algorithms have been applied to each of these AI sub-components. However, there is a trend to use a fully CNN-based algorithm such as U-net for image segmentation and use a classification CNN- based algorithm for identification and classification of the disease aiming at the ROI. Alternatively, radiomics based classification can be employed on the ROI.

Data Labeling

In radiological imaging, the supervised learning approach is the most popular tool, and it requires labeled data for training and validation. The labeling of images must be done manually by expert radiologists. This process is very labor-intensive and very costly. The truth panel for images is established by having 2 out of 3 radiologists agreeing on the diagnoses and clinical determinations (Lo et al., 2018a).

Research Environment - Access to Research Resources

The role of AI will be different in different parts of the world. The AI tools developed for one region of the world using data from that region may not be useful in other regions with different disease prevalence, limited infrastructure, and different healthcare systems.

Collection and curation of images and related data could face several obstacles such as management of privacy, confidentiality, and the question of ownership. In recent years, the realization of clinical images' possible commercial values makes access more costly and difficult. When images need to be collected from multiple organizations, the data sharing process can become more complicated. International collaboration can be difficult when certain countries do not allow clinical data movement beyond their national borders (Prior et al., 2020). One technical solution for such a situation can be a federated learning system where data remains in place while processing code and processed results can move around (Konečný et al., 2016)

Government agencies and various consortia have established a growing number of open access data repositories to facilitate better access to clinical image data for research. One of the best-known such repositories is the Cancer Imaging Archive (TCIA) (Prior et al., 2013; The Cancer Imaging Archive (TCIA), 2020). It is a publicly available information repository for data about cancer, mostly radiology and pathology data acquired by the lung cancer screening project involving 26,722 participants from 2002 through 2004. It contains 22.3 TB of data. The types and volume of data in the archive are increasing rapidly.

Research and Development Environment in Resource-Limited Regions

The research and development and eventual adoption of AI for medical decision making in global health and low-resource settings are hampered by insufficient infrastructure (Mollura et al., 2020).However, it is essential that local radiology and clinical community, resource-poor or not, have to develop and validate AI tools suitable for their environment. In the resource-poor regions with limited infrastructure, technical and human, such participation could be difficult. However, the research communities in the world's resource-limited regions can access many global imaging AI research resources. The Radiological Society of North America website has a vast amount of information. Many of the AI software is freely available as open-source at no cost to the users. The cancer imaging archive (TCIA) of the National Cancer Institute of the US has many curated radiological images to support imaging research. Most of the CADe products for lung cancer screen started using the openly available images in TCIA, which holds the CT images from the national lung cancer screening trial.

Future of Radiology Service and Radiomics

The digital transformation of radiology services will continue and accelerate. Analog film is gone, and modern imaging systems have evolved far beyond the slow and primitive early MRI and CT systems. PET was only a research tool at a few centers but is now becoming available at small community hospitals. Hybrid scanners that combine multiple modalities and can operate in different healthcare settings will become readily available. The whole-body scanner that could do MRI or CT or PET will be available where the patients receive care (Pichler et al., 2008; Nensa et al., 2018). PACS, teleradiology, and Radiology Information Systems all changed the radiology practice.

Nevertheless, the radiology department structure has not fundamentally changed in 30 years (Kim and Mansfield, 2014). The radiology department continues to operate as a centralized resource to which patients come to complete the study, and radiologists dictate and distribute the reports (Ondategui-Parra et al., 2004). This operational model may soon see some changes.

For the next 30 years of radiology in the future, the pace of change will accelerate. The cost of computing will continue to decrease, and connectivity will be fast and ubiquitous. Radiology and the field of diagnosis will evolve together with pathology. Diagnostic imaging, clinical pathology, and genomics could merge as an integrated diagnostic service that can integrate the various reports from these subspecialty sections and synthesize a coherent diagnosis that is communicated to the appropriate physicians with more actionable specifics. Such an integrated system will allow a rapid on-site point of care diagnosis, rather than the serial process of today involving multiple appointments over many days and weeks.

Some expect that a new profession of clinical diagnosticians, who integrate the work of radiologists and pathologists with other specialists with increasing reliance on AI assistance, will begin to grow (Lundström et al., 2017). Pathologists are already doing biopsies under image guidance. There are movements toward the integration of professional service. Simultaneously, there are parallel significant scientific and technical developments underway in radiomics and pathomics that can facilitate this historical evolution.

Radiomics and pathomics are part of quantitative imaging that attempts to extract additional information from radiology and pathology images that may not be visible via visual inspection (Saltz et al., 2017). Radiomics attempts to extract features from radiological images that quantify its phenotype characteristics in an automated high-throughput manner (Fan et al., 2020). Pathomics attempts to extract similar information from pathology images. These two approaches will meet at a shared space to support personalized medicine. It has been hypothesized that such analysis may help prognosticate, predict treatment outcomes, and assess cancer tissue malignancy.

The value of AI in radiomics is two-fold. First, AI can be used for automated image analysis at scale, enabling rapid evaluation of hypothetical radiomic features. Whereas comparison studies involving human radiologists should take into account for a wide range of ergonomic and perception factors (such as the required number of readers or the need to provide time between different readings of the same image), a comparison of radiomics algorithms is only limited by computational speed and power. Similarly, new features can easily be tested against existing data sets. Secondly, unsupervised learning methods can be used to search for new radiomic features that might be very different from what

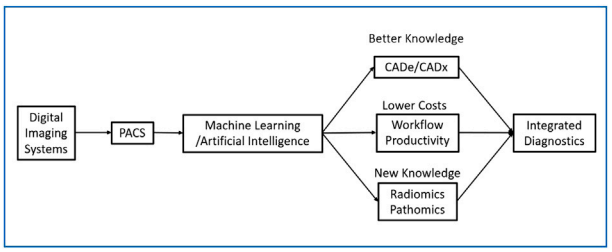


FIGURE 1 | Digital Transformation of Radiology with 3 Categories of Application of ML/AI.

would be noticed by a human observer. The AI can take on the discovery and find new and useful patterns within the existing imaging data (Miles, 2020; Prior et al., 2020).

One fundamental issue in radiomics has to address is the standardization of the factors and processing involved in the quantitative analysis. The Image Biomarker Standardization Initiative (IBSI) is a new organization to address many challenges in 4 different specific areas, 1) standard nomenclature and common radiomic features, 2) radiomics image processing schemes, 3) provide data sets for validation and calibration, and 4) set of reporting guidelines (Zwanenburg et al., 2020). This group defined 174 radiomics features commonly used to quantify the morphologic characteristics and numerous others needed to define the quantitative information. The group tries to standardize the image processing steps of data conversion, post-acquisition processing, segmentation, interpolation, masking, and others (Zwanenburg et al., 2020)(Zwanenburg et al., 2020). Such standardization is expected to make radiomics and pathomics clinically useful and scalable for the integrated diagnosis service (Kuhl and Truhn, 2020).

CONCLUSION

The clinical adoption of AI can be driven by either technology push or market pull (Chidamber and Kon, 1994; Di Stefano et al., 2012). Technology push arises when a new idea or new tool creates a capability that did not previously exist. The market pull is defined by the need to address pain points, inefficiencies, and problems with the current way of doing business. Ideally, these two forces synergistically combine to accelerate technology development and deployment. Over the last 30 years, radiology has benefited from this combination to develop teleradiology, PACS/RIS, and advanced imaging modalities. However, much of CAD's early development has been a technology push; it was usually not well aligned with clinical needs. We have proposed that a better alignment could arise from focusing on the radiology workflow – which includes many tasks beyond image interpretation – and the need to create an integrated diagnostics service combining radiological images, pathological data, and genomics (Santos et al., 2019).

We envision three AI trajectories in radiology, as shown in **Figure 1**. First, AI will undergo advances in CADe and CADx to make image interpretation better and faster. Despite the significant progress in developing CNN algorithms, there are still many areas for improvement as proposed in this paper. Second, a variety of AI tools, supervised learning and unsupervised learning, will be needed to improve workflow and increase productivity, and, at the same time, reduce the cost of operation. This operationally focused research will require a holistic understanding of radiology operations. Third, quantitative imaging, including radiomics, pathomics, and genomics, will emerge and become a standardized approach for integrated diagnostics. In summary, we predict that AIs will facilitate the merging of disparate medical and scientific domains into an integrated diagnostic service for personalized precision medicine.

AUTHOR CONTRIBUTIONS

SM designed the document and integrated different ideas into a single document. KW drafted the section dealing with the clinical adoption of AI and radiomics and edited the final document. SL drafted the section on CNN and data needs. YL did a literature review and developed the list of relevant references. SB was responsible for developing radiology workflow issues.

REFERENCES

Alhajeri, M., Aldosari, H., and Aldosari, B. (2017). Evaluating latest developments in PACS and their impact on radiology practices: A systematic literature review. *Informatics in Medicine Unlocked* 9, 181–190. doi:10.1016/j.imu.2017.08.005

Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P., et al. (2020). "A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy," in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Honolulu, HI: Hawaii Convention Center, April, 2020, (Accessed August 11, 2020). doi:10.1145/3313831.3376718

Boochever, S. S. (2004). HIS/RIS/PACS integration: getting to the gold standard. *Radiol. Manage.* 26 (3), 16–7.

Brady, A. P. (2017). Error and discrepancy in radiology: inevitable or avoidable? *Insights Imaging* 8 (1), 171–182. doi:10.1007/s13244-016-0534-1

Chidamber, S. R., and Kon, H. B. (1994). A research retrospective of innovation inception and success: the technology–push, demand–pull question. *Internat. J. Tech. Manag.* 9 (1), 94–112. doi:10.1504/IJTM.1994.025565

Di Stefano, G., Gambardella, A., and Verona, G. (2012). Technology push and demand pull perspectives in innovation studies: Current findings and future research directions. *Research policy* 41 (8), 1283–1295. doi:10.1016/j.respol.2012.03.021

Dikici, E., Bigelow, M., Prevedello, L. M., White, R. D., and Erdal, B. S. (2020). Integrating AI into radiology workflow: levels of research, production, and feedback maturity. *J. Med. Imaging (Bellingham)* 7 (1), 016502. doi:10.1117/1.JMI.7.1.016502

Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput. Med. Imaging Graph* 31 (4–5), 198–211. doi:10.1016/j.compmedimag.2007.02.002

Doshi, A. M., Moore, W. H., Kim, D. C., Rosenkrantz, A. B., Fefferman, N. R., Ostrow, D. L., et al. (2018). Informatics solutions for driving an effective and efficient radiology practice. *RadioGraphics* 38 (6), 1810–1822. doi:10.1148/rg.2018180037

ESR (2019). What the radiologist should know about artificial intelligence - an ESR white paper. *Insights Imaging* 10 (1), 44. doi:10.1186/s13244-019-0738-2

Fan, M, Xia, P., Clarke, R., Wang, Y., and Li, L. (2020). Radiogenomic signatures reveal multiscale intratumour heterogeneity associated with biological functions and survival in breast cancer. *Nat. Commun.* 11 (1), 4861.doi:10.1038/s41467-020-18703-2

Forsberg, D., Rosipko, B., and Sunshine, J. L. (2017). Radiologists' variation of time to read across different procedure types. *J. Digit. Imaging* 30 (1), 86–94. doi:10.1007/s10278-016-9911-z

Fosfuri, A., Giarratana, M. S., and Luzzi, A. (2008). The penguin has entered the building: the commercialization of open source software products. *Organization science* 19 (2), 292–305. doi:10.1287/orsc.1070.0321

Giger, M. L., Chan, H. P., and Boone, J. (2008). Anniversary paper: history and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. *Med. Phys.* 35 (12), 5799–5820. doi:10.1118/1.3013555

Gold, L. S., Klein, G., Carr, L., Kessler, L., and Sullivan, S. D. (2012). The emergence of diagnostic imaging technologies in breast cancer: discovery, regulatory approval, reimbursement, and adoption in clinical guidelines. *Cancer Imaging* 12 (1), 13–24. doi:10.1102/1470-7330.2012.0003

Halsted, M. J., and Froehle, C. M. (2008). Design, implementation, and assessment of a radiology workflow management system. *AJR Am. J Roentgenol.* 191 (2), 321–327. doi:10.2214/AJR.08.1566

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., and Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nat. Rev. Cancer* 18 (8), 500–510. doi:10.1038/s41568-018-0016-5

Hricak, H. (2018). 2016 new horizons lecture: beyond imaging-radiology of tomorrow. *Radiology* 286 (3), 764–775. doi:10.1148/radiol.2017171503

Interagency Working Group on Medical Imaging, Committee on Science, and National Science and Technology Council (2017). The interagency working group on medical imaging released a final report in December 2017. Available at: https://www.whitehouse.gov/wp-content/uploads/2017/12/Roadmapfor-Medical-Imaging-Research-and-Development-2017.pdf

Kahn, C. E., Jr, Langlotz, C. P., Burnside, E. S., Carrino, J. A., Channin, D. S., Hovsepian, D. M., et al. (2009). Toward best practices in radiology reporting. *Radiology* 252 (3), 852–856. doi:10.1148/radiol.2523081992

Kim, Y. W., and Mansfield, L. T. (2014). Fool me twice: delayed diagnoses in radiology with emphasis on perpetuated errors. *AJR Am. J. Roentgenol.* 202 (3), 465–70. doi:10.2214/AJR.13.11493

Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., and Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. *arXiv*.

Kuhl, C. K., and Truhn, D. (2020). The long route to standardized radiomics: unraveling the knot from the end. *Radiology* 295 (2), 339–341. doi:10.1148/radiol.2020200059

Lehman, C. D., Wellman, R. D., Buist, D. S., Kerlikowske, K., Tosteson, A. N., Miglioretti, D. L., et al. (2015). Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern. Med.* 175 (11), 1828–1837.doi:10.1001/jamainternmed.2015.5231

Levine, S. S., and Prietula, M. J. (2014). Open collaboration for innovation: principles and performance. *Organization Science* 25 (5), 1414–1433. doi:10.1287/orsc.2013.0872

Lo, S.-C. B., Chan, H.-P., Lin, J.-S., Li, H., Freedman, M. T., and Mun, S. K. (1995). Artificial convolution neural network for medical image pattern recognition. *Neural Networks* 8 (7–8), 1201–1214. doi:10.1016/0893-6080(95)00061-5

Lo, S.-C. B., Lin, J.-S., Freedman, M. T., and Mun, S. K. (1993). Computer-assisted diagnosis of lung nodule detection using artificial convoultion neural network. *Medical Imaging: Image Processing* 1898, 859–869. doi:10.1117/12.154572

Lo, S.-C. B., Lin, J.-S.J., Freedman, M. T., and Mun, S. K. (1998). Application of artificial neural networks to medical image pattern recognition: detection of clustered microcalcifications on mammograms and lung cancer on chest radiographs. *J. VLSI signal processing systems for signal, image video technology* 18 (3), 263–274.

Lo, S. B., Freedman, M. T., Gillis, L. B., White, C. S., and Mun, S. K. (2018a). JOURNAL CLUB: computer-aided detection of lung nodules on CT with a computerized pulmonary vessel suppressed function. *AJR Am. J. Roentgenol.* 210 (3), 480–488. doi:10.2214/AJR.17.18718

Lo, S. B., Freedman, M. T., and Mun, S. K. (2018b). Transformationally Identical and Invariant Convolutional Neural Networks by Combining Symmetric Operations or Input Vectors. arXiv preprint arXiv:1807.11156.

Lo, S. B., Freedman, M. T., Mun, S.K., and Chan, H.-P. (2018c). Geared Rotationally Identical and Invariant Convolutional Neural Network Systems. arXiv preprint arXiv:1808.01280.

Lo, S. C. B., Freedman, M. T., Mun, S. K., and Gu, S. (2018d). Transformationally identical and invariant convolutional neural networks through symmetric element operators. arXiv preprint arXiv:1806.03636.

Lundström, C. F., Gilmore, H. L., and Ros, P. R. (2017). Integrated diagnostics: the computational revolution catalyzing cross-disciplinary practices in radiology, pathology, and genomics. *Radiological* 285 (1), 12–15. 10.1148/radiol.2017170062

Mazurowski, M. A., Buda, M., Saha, A., and Bashir, M. R. (2019). Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *J. Magn. Reson. Imaging* 49 (4), 939–954. doi:10.1002/jmri.26534

McDonald, R. J., Schwartz, K. M., Eckel, L. J., Diehn, F. E., Hunt, C. H., Bartholmai, B. J., et al. (2015). The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. *Acad. Radiol.* 22 (9), 1191–1198. doi:10.1016/j.acra.2015.05.007

Miles, K (2020). Radiomics for personalised medicine: the long road ahead. *Br J. Cancer* 122 (7), 929–930. doi:10.1038/s41416-019-0699-8

Mollura, D. J., Culp, M. P., Pollack, E., Battino, G., Scheel, J. R., Mango, V. L., et al. (2020). Artificial intelligence in low-and middle-income countries: innovating global health radiology. *Radiology* 297 (3), 513–520. 10.1148/radiol.2020201434

Montagnon, E., Cerny, M., Cadrin-Chênevert, A., Hamilton, V., Derennes, T., Ilinca, A., et al. (2020). Deep learning workflow in radiology: a primer. *Insights Imaging* 11 (1), 22. doi:10.1186/s13244-019-0832-5

Multiple-Editor (2020). Comparison of deep-learning software [Online]. Available from: https://en.wikipedia.org/wiki/Comparison_of_deep-learning_software.

Mun, S. K., Freedman, M., and Kapur, R. (1993). Image management and communications for radiology. *IEEE Engineering in Medicine and Biology Magazine* 12 (1), 70–80. doi:10.1109/51.195943

Mun, S. K., Levine, B., Cleary, K., and Dai, H. (1998). Deployable teleradiology and telemedicine for the US military. *Comput. Methods Programs Biomed.* 57 (1–2), 21–27. doi:10.1016/s0169-2607(98)00040-6

Mun, S. K., Prior, F., Caramella, D., and Ratib, O. (2007). Introduction to the special section on image management in the healthcare enterprise. *IEEE Trans. Inf. Technol. Biomed.* 11 (1), 1–6. doi:10.1109/titb.2006.888236

Mun, S. K., and Turner, J. W. (1999). Telemedicine: Emerging e-medicine. *Annu. Rev. Biomed. Eng.* 1 (1), 589–610. doi:10.1146/annurev.bioeng.1.1.589

Mutasa, S., Sun, S., and Ha, R. (2020). Understanding artificial intelligence based radiology studies: What is overfitting? *Clin. Imaging* 65, 96–99. 10.1016/j.clinimag.2020.04.025

Nensa, F., Bamberg, F., Rischpler, C., Menezes, L., Poeppel, T.D., La Fougère, C., et al. (2018). Hybrid cardiac imaging using PET/MRI: a joint position statement by the European Society of Cardiovascular Radiology (ESCR) and the European Association of Nuclear Medicine (EANM). *European Journal of Hybrid Imaging* 28 (1), 4086–4101. doi:10.1007/s00330-017-5008-4

Ondategui-Parra, S., Bhagwat, J. G., Zou, K. H., Gogate, A., Intriere, L. A., Kelly, P., et al. (2004). Practice management performance indicators in academic radiology departments. *Radiology* 233 (3), 716–722. doi:10.1148/radiol.2333031147

Opensource.org (2020). Open Source Licenses by Category [Online]. Open Source Initiative. Available at: https://opensource.org/licenses/category.

Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., et al. (2019). *The AI Index 2019 Annual Report. AI Index Steering Committee*. Stanford, CA: Human-Centered AI Institute Stanford University.

Pichler, B. J., Judenhofer, M. S., and Wehrl, H. F. (2008). PET/MRI hybrid imaging: devices and initial results. *Eur. Radiol.* 18 (6), 1077–1086. doi:10.1007/s00330-008-0857-5

Prior, F., Almeida, J., Kathiravelu, P., Kurc, T., Smith, K., Fitzgerald, T.J., et al. (2020). Open access image repositories: high-quality data to enable machine learning research. *Clin. Radiol.* 75 (1), 7–12. doi:10.1016/j.crad.2019.04.002

Prior, F. W., Clark, K., Commean, P., Freymann, J., Jaffe, C., Kirby, J., et al. (2013). TCIA: an information resource to enable open science, *Annu. Int. Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2013, 1282–1285. doi:10.1109/EMBC.2013.6609742

Saltz, J., Almeida, J., Gao, Y., Sharma, A., Bremer, E., DiPrima, T., et al. (2017). Towards generation, management, and exploration of combined radiomics and pathomics datasets for cancer research. *AMIA Jt. Summits Transl. Sci. Proc.* 2017, 85–94.

Santos, M. K., Ferreira Júnior, J. R., Wada, D. T., Tenório, A. P. M., Barbosa, M. H. N., and Marques, P. M. A. (2019). Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine. *Radiol. Bras.* 52 (6), 387–396. doi:10.1590/0100-3984.2019.0049

Schaffter, T., Buist, D. S. M., Lee, C.I., Nikulin, Y., Ribli, D., Guan, Y., et al. (2020). Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. *JAMA Netw. Open* 3 (3), e200265. doi:10.1001/jamanetworkopen.2020.0265

Schemmel, A., Lee, M., Hanley, T., Pooler, B.D., Kennedy, T., Field, A., et al. (2016). Radiology workflow disruptors: a detailed analysis. *J. Am. Coll. Radiol.* 13 (10), 1210–1214.doi:10.1016/j.jacr.2016.04.009

Soffer, S., Ben-Cohen, A., Shimon, O., Amitai, M. M., Greenspan, H., and Klang, E. (2019). Convolutional neural networks for radiologic images: a radiologist's guide. *Radiology* 290 (3), 590–606. doi:10.1148/radiol.2018180547

Tadavarthi, Y., Vey, B., Krupinski, E., Prater, A., Gichoya, J., Safdar, N., et al. (2020). The state of radiology AI: considerations for purchase decisions and current market offerings. *Radiology: Artificial Intelligence* 2 (6), e200004. doi:10.1148/ryai.2020200004

Tang, A., Tam, R., Cadrin-Chênevert, A., Guest, W., Chong, J., Barfett, J., et al. (2018). Canadian association of radiologists white paper on artificial intelligence in radiology. *Can. Assoc. Radiol. J.* 69 (2), 120–135. doi:10.1016/j.carj.2018.02.002

The Cancer Imaging Archive (TCIA) (2020). The Cancer Imaging Archive [Online]. Available at: https://www.cancerimagingarchive.net/.

Worrell, S. (2020). Making AI Tools Clinically Relevant- Sustainability and Scalability. (*Central VA Health Care System Artificial Intelligence for Medical Imaging Virtual Workshop: Virginia Tech-AIC)*. Available at: https://aic.ncr.vt.edu/aic-workshops-and-meetings/aimiworkshophome/aimirecording.html

Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9 (4), 611–629. doi:10.1007/s13244-018-0639-9

Zhu, Q., Du, B., Turkbey, B., Choyke, P. L., and Yan, P. (2017). Deeply-supervised CNN for prostate segmentation", in 2017 International Joint Conference on Neural Networks (Ijcnn), Anchorage, AK, May 14–19, 2017 (IEEE) doi:10.1109/IJCNN.2017.7965852

Zwanenburg, A., Vallières, M., Abdalah, M. A., Aerts, H. J. W. L., Andrearczyk, V., Apte, A., et al. (2020). The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 295 (2), 328–338. doi:10.1148/radiol.2020191145

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 17 February 2021
doi: 10.3389/fmolb.2021.614174

A Hybrid-Attention Nested UNet for Nuclear Segmentation in Histopathological Images

*Hongliang He[1,2], Chi Zhang[1,2], Jie Chen[1,2], Ruizhe Geng[1], Luyang Chen[3], Yongsheng Liang[4], Yanchang Lu[5], Jihua Wu[6] and Yongjie Xu[6]**

[1]School of Electronic and Computer Engineering, Peking University, Shenzhen, China, [2]Peng Cheng Laboratory, Shenzhen, China, [3]College of Engineering, Pennsylvania State University, State College, PA, United States, [4]Harbin Institute of Technology, Shenzhen, China, [5]Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China, [6]PLA Strategic Support Force Characteristic Medical Center, Beijing, China*

Nuclear segmentation of histopathological images is a crucial step in computer-aided image analysis. There are complex, diverse, dense, and even overlapping nuclei in these histopathological images, leading to a challenging task of nuclear segmentation. To overcome this challenge, this paper proposes a hybrid-attention nested UNet (Han-Net), which consists of two modules: a hybrid nested U-shaped network (H-part) and a hybrid attention block (A-part). H-part combines a nested multi-depth U-shaped network and a dense network with full resolution to capture more effective features. A-part is used to explore attention information and build correlations between different pixels. With these two modules, Han-Net extracts discriminative features, which effectively segment the boundaries of not only complex and diverse nuclei but also small and dense nuclei. The comparison in a publicly available multi-organ dataset shows that the proposed model achieves the state-of-the-art performance compared to other models.

Keywords: Histopathological image, Nuclear segmentation, Nested UNet, Hybrid attention, Dilated convolution

OPEN ACCESS

Edited by:
Xin Gao,
King Abdullah University of Science and Technology, Saudi Arabia

Reviewed by:
Qing Liu,
Central South University, China
Xipeng Chen,
Sun Yat-Sen University, China

****Correspondence:***
Yongjie Xu
armxyj007@aliyun.com

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *05 October 2020*
Accepted: *06 January 2021*
Published: *17 February 2021*

Citation:
He H, Zhang C, Chen J, Geng R, Chen L, Liang Y, Lu Y, Wu J and Xu Y (2021) A Hybrid-Attention Nested UNet for Nuclear Segmentation in Histopathological Images. Front. Mol. Biosci. 8:614174. doi: 10.3389/fmolb.2021.614174

1 INTRODUCTION

Histopathological imaging diagnosis is an important significance of cancer diagnosis, known as the "gold standard" of clinical tumors. Nuclear segmentation of histopathological images is a crucial step in computer-aided image analysis. Accurately segmenting the nucleus in pathological tissue sections provides powerful support for disease diagnosis, cancer staging, and postoperative treatment. However, the task of nuclear segmentation in histopathological images is still challenging, for which 1) the types of histopathological structures are complex and diverse, and there are many types and complex appearances of nuclei; 2) the nuclei are usually small and dense, leading to an overlapping challenge for nuclear segmentation.

Traditional nuclear segmentation methods have contributed to some extent, such as Otsu (Otsu, 1979), the watershed method (Yang et al., 2006), K-mean clustering (Filipczuk et al., 2011), and Grab Cut (Rother et al., 2004). However, some specific parameters or thresholds are required to set while using these methods for nuclear segmentation. Besides, the lack of generalization ability makes these methods only effective for a few types of histopathological images. With the application and development of deep learning technology in image segmentation, these traditional nuclear segmentation methods are only used as pre/post-processing steps.

In recent years, some models based on convolutional neural networks have been proposed for histopathological image analysis (Ronneberger et al., 2015), (Zhou et al., 2018), (Wollmann et al.,

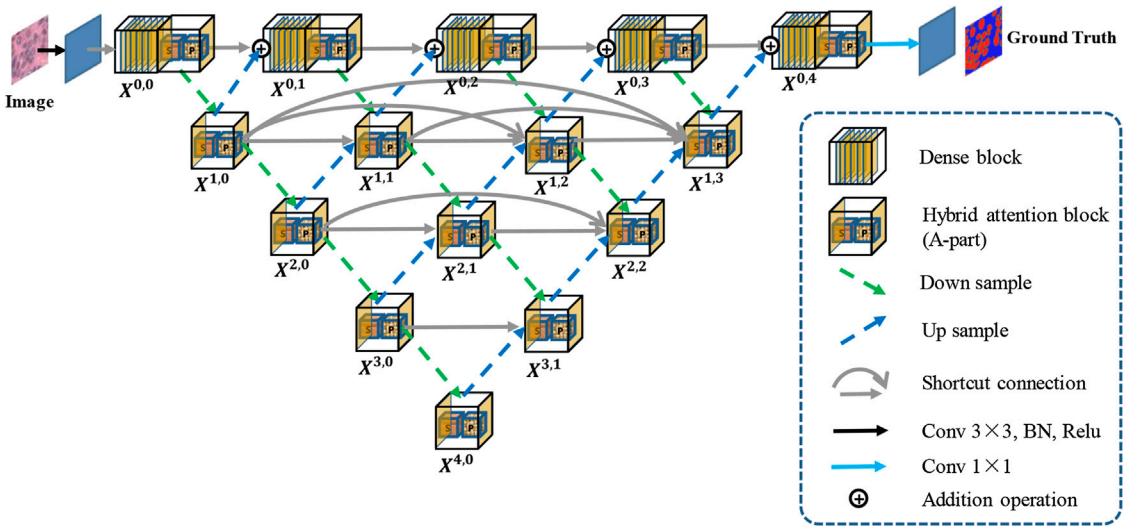


FIGURE 1 | Overview of the proposed Han-Net.

2019), (Qu et al., 2019). UNet (Ronneberger et al., 2015) is successfully applied to segmentation tasks in medical image analysis. The network adopts encode-decode design and has skip connections to combine low-level and high-level feature information. This framework effectively captures the contextual information of image features and has become a basic framework of the current mainstream model for segmentation tasks. However, only using the UNet framework cannot efficiently separate dense or overlapping instances. Therefore, some models have been proposed to improve the performance of UNet. Zhou et al. (2018) proposed UNet++, which reduces the semantic gap between feature maps of encoder and decoder subnets through a series of nested, dense skip pathways. Wollmann et al. (2019) proposed GRUU-Net, which integrates convolutional neural networks and gated recurrent neural networks on multiple image scales to combine the advantages of both types of networks. In addition to encode-decode architecture, dilated convolution has also been proposed and applied to segmentation tasks. Yu and Koltun first proposed a new convolutional network module specifically for dense prediction, which uses dilated convolution to systematically aggregate multi-scale context information without loss of resolution (Yu and Koltun, 2016). Qu et al. introduced the idea of dilated convolution into nuclear segmentation and proposed FullNet (Qu et al., 2019). FullNet uses several densely connected layers with different dilation factors to replace the encoding-decoding operation, thereby avoiding the loss of feature information. Experiments show that the performance of FullNet in cell nuclear segmentation is better than other comparison models. In general, although some latest models have turned their attention to the gaps of the UNet model and achieved competitive performance to some extent, they still have limitations. For example, 1) they fail to effectively identify small or dense objects; 2) they treat each pixel as a separate classification task, failing to fully consider global feature information and the relevance between different pixels.

To address the issues mentioned above, the main contributions of this paper are 1) We propose a hybrid-attention nested UNet (Han-Net), which consists of two modules: a hybrid nested U-shaped network (H-part) and hybrid attention block (A-part). 2) In H-part, we integrate a nested multi-depth U-shaped network and a dense network with full resolution to capture more effective feature information. The excellent feature extraction capability of H-part can effectively segment the boundaries of complex and diverse nuclei. 3) We propose a novel hybrid attention block (A-part) to boost attention information and explore the correlation between different pixels, thereby to effectively capture some small or dense nuclei. 4) Han-Net is compared with some recently proposed models in a multi-organ segmentation dataset. The comparison results show that the proposed model has the state-of-the-art performance.

2 METHOD

Figure 1 shows an overview of the proposed Han-Net, which consists of the backbone module (H-part) and hybrid attention module (A-part). H-part is proposed to obtain multi-scale feature information and improve the capability for exploring effective features. A-part is proposed to boost attention information and capture the correlation between features. In **Figure 1**, each different block and process is represented by different icons and arrows. The detailed information of H-part and A-part is described in the following sub-sections.

2.1 Hybrid Nested U-Shaped Network (H-part)

The framework of H-part is similar to a multilayer regular triangle structure, which is composed of multiple encoders and decoders. The framework of the proposed H-part is shown in

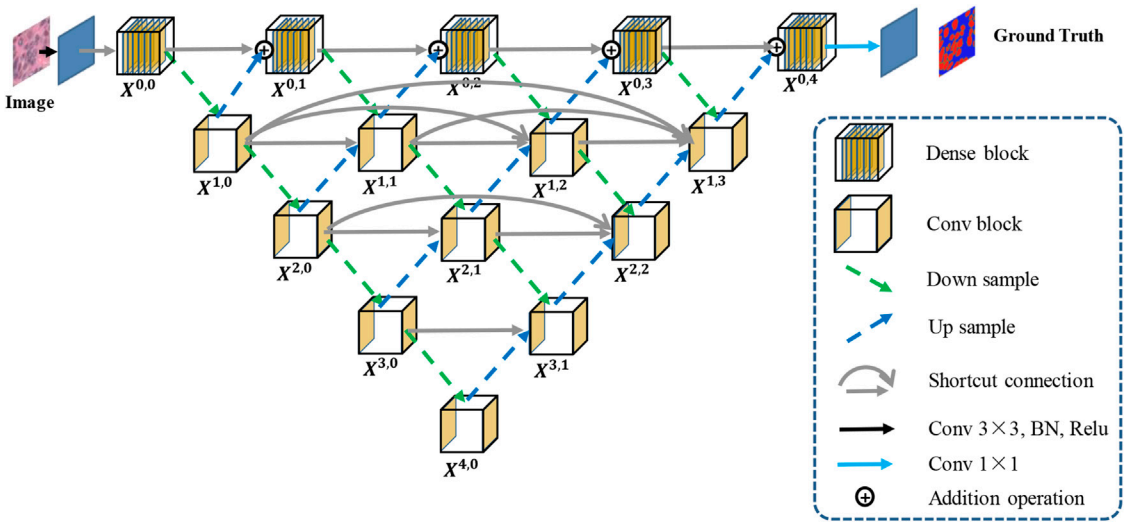


FIGURE 2 | The framework of the proposed H-part.

Figure 2. Inspired by UNet++ (Zhou et al., 2018), we have nested multiple conv blocks in UNet to bridge the possible semantic gap between the corresponding levels of encoder-decoder in classic UNet.

In the top layer, we added a dense network using dilated convolution to make it obtain more global feature information at full resolution. Further, we unify the channel number of the output feature map obtained from up-sampling and previous dense block, so that feature map fusion can be carried out by adding and acting as the input of next dense block. Except the top layer, we set up shortcut connections to prevent losing feature information, and use concatenate operation for the fusion of feature maps. A dense block contains L_n convolutional layers, and a conv block contains two convolutional layers.

Moreover, we explain the calculation relationship between each block: let $x^{i,j}$ denote the output of block $X^{i,j}$, where i represents the index of the encoder, and j represents the index of dense block at the top layer. The formula for calculating the output $x^{i,j}$ of each block is shown in **Eq. 1**.

$$x^{i,j} = \begin{cases} \mathcal{M}\left(A\left(x^{i,j-1} + x^{i+1,j-1}\right)\right), & i = 0, j > 0 \\ \mathcal{C}\left(\left[\mathcal{D}\left(x^{i-1,j}\right)\right]\right), & i > 0, j = 0 \\ \mathcal{C}\left(\left[\left[x^{i,k}\right]_{k=0}^{j-1}, \mathcal{D}\left(x^{i-1,j}\right), \left(x^{i+1,j-1}\right)\right]\right) & i > 0, j > 0 \end{cases} \quad (1)$$

where $\mathcal{M}(\cdot)$ represents the convolution operation in dense block, $\mathcal{C}(\cdot)$ represents the convolution operation in conv block, $(\cdot)$ represents the addition layer, and $[\cdot]$ represents the concatenate layer. $\mathcal{D}(\cdot)$ and $(\cdot)$ refer to the down-sampling layer and the up-sampling layer, respectively. $x^{0,0}$ has only one input and serves as the starting position for other blocks. When $j > 0$, each dense block has two inputs and each conv block has three or more inputs.

2.2 Hybrid Attention Block (A-Part)

To highlight effective features and explore the correlation between different pixels, we propose a hybrid attention block (A-part). Its structure is shown in **Figure 3**. In A-part, channel attention and spatial attention constitute a series of attention module and parallel attention module through series and parallel operations, respectively. Assume that the input feature map are $X \in \mathbb{R}^{H \times W \times C}$. Here H and W are spatial height and width, respectively, and C represents the input channel. Series attention module and parallel attention module are used in convolutions of different kernel sizes, and the results obtained from these attention modules are added to obtain the output $X' \in \mathbb{R}^{H \times W \times C'}$, where C' represents the output channel.

Channel attention captures the importance of different channels in feature maps, thereby enhancing or suppressing different channels. The operation is as follows: a branch is separated after a normal convolution operation. Squeeze operation is first performed on the branch (i.e., $F_{sq}(\cdot)$ in **Figure 3**). It uses a global pooling operation to compress the spatial dimensions; that is, each two-dimensional feature map becomes a real number, and the number of feature channels does not change. Then, Excitation operation is performed to generate a corresponding weight for each channel through parameter w (i.e., $F_{ex}(\cdot)$ in **Figure 3**). In other words, w is learned to represent the correlation between feature channels.

Spatial attention aims at exploring the relative importance of each pixel on feature map. It emphasizes related spatial locations and ignores unrelated locations. The operation flow is as follows: First, a 1-channel kernel size of 1 × 1 convolution is performed on the input feature map, and then a sigmoid operation is performed. Then, the attention rate corresponding to each pixel is multiplied with the original feature map to obtain new feature map.

In this paper, A-part is added after each dense block in H-part to strengthen the effective information while controlling the number of channels. In the nested U-shape network of H-part,

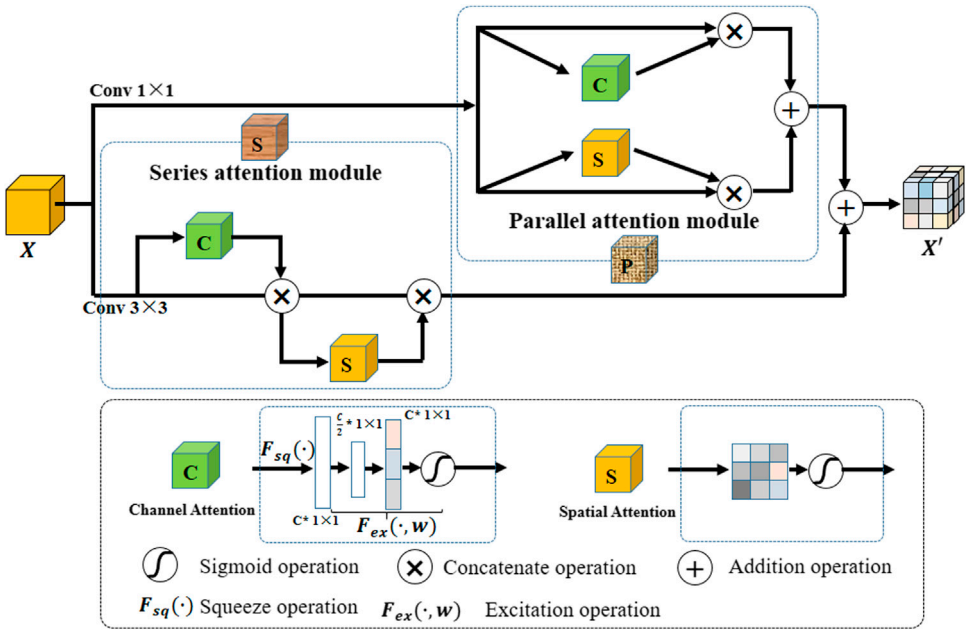


FIGURE 3 | The structure of the hybrid attention block (A-part).

TABLE 1 | Ablation study for the Han-Net.

Method	Test 1				Test 2			
	F1	Dice	AJI	H	F1	Dice	AJI	H
UNet	0.8651	0.7987	0.5923	6.2683	0.8286	0.7701	0.5442	7.6555
NUNet	0.8765	0.7962	0.5993	6.2063	0.8422	0.7959	0.6004	6.9937
NUNet + A-part	0.8847	0.7992	0.6043	6.1121	0.8632	0.7990	0.6169	6.9149
NUNet + Dense-top	0.8777	0.7977	0.5999	6.1287	0.8688	0.8123	0.6321	6.3849
Ours (Han-Net)	**0.8875**	**0.8021**	**0.6091**	**6.0157**	**0.8802**	**0.8185**	**0.6419**	**5.9524**

The bold numbers show the best performance.

we replace the original conv block with A-part to improve extraction capability for effective feature information.

3 EXPERIMENTAL AND RESULTS

3.1 Dataset and Evaluation Metrics

We validate the performance of our proposed model in the Multi-Organ nuclear segmentation (MoNuSeg) dataset (Kumar et al., 2017). The dataset consists of 30 hematoxylin and eosin (H & E) stained histopathology images of size 1000 × 1000 from seven different organs. Following (Kumar et al., 2017), the data division mode is as follows. We extract the first three images from four organs (six images per organ) as the training set (12 images in total), the 4th image as the validation set (four images in total), and the remaining two as the Test 1 (eight images in total). six images of the remaining three organs are used as the Test 2.

In this paper, four evaluation metrics are used to comprehensively evaluate the performance of the proposed model, namely F1-score (F1), Dice coefficient (Dice), average Hausdorff distance (H), and Aggregated Jaccard Index (AJI).

TABLE 2 | The performance comparison of different methods on the MoNuSeg dataset.

Method	Test 1		Test 2	
	F1	AJI	F1	AJI
DCAN Chen et al. (2017)	0.8265	0.6082	0.8214	0.5449
BES-Net Oda et al. (2018)	0.8118	0.5906	0.7952	0.5823
CIA-Net Zhou et al. (2019)	0.8244	**0.6129**	0.8458	0.6306
Spa-Net Koohbanani et al. (2019)	0.8281	0.6239	0.8451	0.6340
FullNet Qu et al. (2019)	0.8552	0.5946	0.8639	0.6164
Ours (Han-Net)	**0.8875**	0.6091	**0.8802**	**0.6419**

The bold numbers show the best performance.

Among them, F1 and Dice are evaluated from pixel-level, and Hausdorff and AJI are evaluated from object-level.

3.2 Implementation Details

We implemented the proposed model with PyTorch 1.0. One NVIDIA Tesla V100 with CUDA 10.1 is used for computation. We treat the nuclear segmentation task as a three-class classification problem, including nucleoli, nuclear boundary, and background. During training, we use Adam as the optimizer. The initial learning

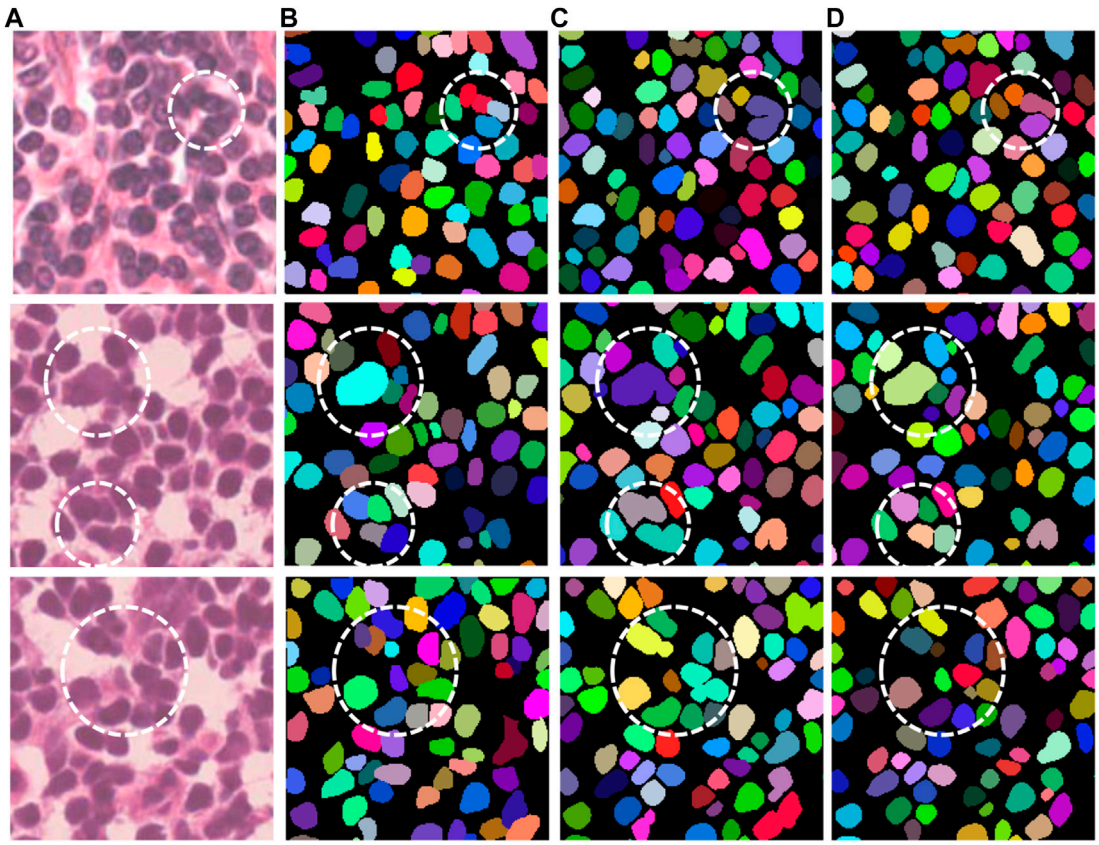


FIGURE 4 | Segmentation results of UNet and Han-Net in MoNuSeg dataset.

rate is set as 0.001 and drop rate is set as 0.1. The dilation factors are $d = (2, 4, 8, 4, 2)$ for the dense net in the top layer of Han-Net. The training epoch for all models in the experiment is set to be 300. In the proposed model, the number of dense layers in each dense block is 6 (i.e. $L_n = 6$), and the number of output features maps for each layer is 16. Due to the limited size of data in the MoNuSeg dataset, we use data augmentation strategies to expand the training set to improve robustness and reduce overfitting. Data augmentations include random color transformation, random crop, horizontal flip, random elastic transformation, and random rotation transformation.

Although the model predicts three-class results during testing, we use inside class area to retrieve the boundary class area and distinguish different nuclei instances according to the boundary class area in the post-processing process. The final boundary area is obtained according to the following steps: 1) Perform dilation and erosion operations in the inside class area; 2) Subtract the area obtained by erosion from the area obtained by dilation to obtain the boundary class area.

3.3 Evaluation and Comparison

3.3.1 Ablation Study

In the experiments, we perform several ablation studies to prove the effectiveness of the proposed method and block. We use UNet as a benchmark and compare the following strategies respectively: 1) Our nested UNet (NUNet): nested U-shaped network, that is, UNet with multiple depths, and skip connections are added to the remaining layers except for the top layer; 2) NUNet + A-part: on the basis of (1), A-part is added without dense network; 3) NUNet +Dense-top: on the basis of (1), dense network is added in the top layer; 4) Ours (Han-Net): both H-part and A-part are adopted. The performance comparison results are shown in **Table 1**.

Comparing the performance in **Table 1**, some conclusions can be drawn: (1) In terms of UNet, our nested UNet performs better in the following aspects: In Test 1, F1 improves 1.14%, and other metrics also improve slightly. In Test 2, the performance of each metric gains a significant improvement. Among them, AJI shows an improvement of 5.62%. These comparison results prove that nested multi-depth U-shaped network can effectively improve segmentation performance. (2) Comparing the performance before and after using A-part, F1 achieves an improvement of 0.82%, Dice achieves 0.30%, and AJI achieves 0.50% in Test 1. In Test 2, F1 improves 2.10%, Dice improves 0.31%, and AJI improves 1.65%, respectively. The Hausdorff distance is reduced in both test sets. This shows that adding the A-part can indeed improve the segmentation performance. (3) Similarly, replacing the original conv block with dense network in the top layer effectively improves the model segmentation performance. Among them, the improvement of AJI is 3.17% in Test 2. These

results prove the effectiveness of the dense network in H-part. (4) In Han-Net, which combines H-part and A-part, the segmentation performance gains more significant improvement basis of our nested UNet. Achieving F1 of 0.8875 and AJI of 0.6091 in Test 1, F1 of 0.8802, and AJI of 0.6419 in Test 2 shows an improvement of 1.10% for F1 and 0.98% for AJI in Test 1, 3.80% for F1 and 4.15% for AJI in Test 2, respectively. Dice and Hausdorff distance also have the same trend. The above results show that the several modules we proposed are effective. Further, compared with the performance improvement in Test 1, the performance improvement in Test 2 is more obvious, which reflects that after adopting the above modules, the Han-Net shows better robustness for the nuclear segmentation of different organs.

Moreover, we compare Han-Net with the novel methods proposed in previous studies. These methods include: DCAN (Chen et al., 2017), BES-Net (Oda et al., 2018), CIA-Net (Zhou et al., 2019), Spa-Net (Koohbanani et al., 2019), FullNet (Qu et al., 2019). They achieved competitive segmentation performance in the MoNuSeg dataset, respectively. Under the two evaluation metrics of F1 and AJI, the performance comparison between different methods is shown in **Table 2**.

The number of parameters and inference time are two important aspects to assess the utility of a new method. In order to further illustrate the effectiveness of the proposed Han-Net, we compared the number of parameters and inference time between Han-Net and vanilla UNet. In our experiment, Han-Net and vanilla UNet have same number of channels in each feature layer Xi (i.e. Xi, i = {0, 1, 2, 3, 4}). The channel numbers in each feature layer Xi are 64, 128, 256, 512, and 1024 respectively. The experimental results show that the parameters of Han-Net and vanilla UNet are 31.04 MB and 30.58 MB, respectively. In the experiment, we replaced the original 'TransposeConv' upsampling approach in UNet with 'bilinear interpolation' to reduce the parameters. In other words, the increase in parameters brought by the proposed modules is close to the decrease in parameters brought by the above operation. On the other hand, the experimental results show that the inference time required by Han-Net and vanilla UNet is basically the same, which means that Han-Net does not require additional inference time. These two comparative experiments also reflect the effectiveness of our proposed Han-Net.

From the performance comparison in **Table 2**, our proposed Han-Net achieves the state-of-the-art performance under F1 in Test 1, Test 2, and AJI in Test 2. It achieves 0.8875 of F1 in Test 1, 0.8802 of F1 and 0.6419 of AJI in Test 2 respectively, which shows an improvement of 3.23% for F1 in Test 1, 1.63% of F1 and 0.79% of AJI in Test 2. The performance of Han-Net's AJI in Test 1 is almost the same as the optimal CIA-Net. Therefore, Han-Net can be considered to reach the state-of-the-art performance in MoNuSeg dataset.

3.3.2 Qualitative analysis

Figure 4 shows several representative examples with challenging cases from the MoNuSeg dataset, which includes nuclei with irregular and densely distributed nuclei. That is, the relevant cases in **Figure 4** are shown by a white dotted circle. It can be obtained from these images that compared with UNet, our proposed Han-Net achieves better segmentation results in some challenging regions.

4 DISCUSSION

In this paper, we propose a hybrid nested attention UNet (Han-Net) for nuclear segmentation in histopathological images, which consists of H-part and A-part. Among them, H-part combines nested U-shaped network and dense network with full-resolution to obtain more effective multi-scale feature information. The A-part is proposed to enhance the effective features and suppress the invalid features, so that the proposed model can learn the morphological information of the nuclei. The experiment results prove that the Han-Net achieves state-of-the-art segmentation performance in the MoNuSeg dataset. In future work, we will consider pruning Han-Net to make it a lightweight network, and try to integrate other methods to decouple the boundary and inside.

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

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

HH proposed the method and wrote most of the sections of this paper, CZ and LC wrote the most of code, RG carried out data preprocessing, JC provided support on computing resources and guidance on deep learning techniques. YoL and YaL provided suggestions on model improvement and guidance on paper writing, JW and YX provided guidance on pathological diagnosis and reviewed experimental results. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by the Nature Science Foundation of China (No.62081360152, No.61972217), Natural Science Foundation of Guangdong Provice in China (No.2019B1515120049, 2020B1111340056).

REFERENCES

Chen, H, Qi, X, Yu, L, Dou, Q, Qin, J, and Heng, PA (2017). Dcan: deep contour-aware networks for object instance segmentation from histology images. *Med Image Anal.* 36, 135–146. doi:10.1016/j.media.2016.11.004

Filipczuk, P., Kowal, M., and Obuchowicz, A. (2011). "Automatic breast cancer diagnosis based on k-means clustering and adaptive thresholding hybrid segmentation". in *Image processing and communications challenges 3 Advances in Intelligent and Soft Computing*. Editor R. S. Choraś (Berlin, Heidelberg: Springer). 295–302

Koohbanani, N. A., Jahanifar, M., Gooya, A., and Rajpoot, N. (2019). "Nuclear instance segmentation using a proposal-free spatially aware deep learning framework". in International Conference on Medical Image Computing and Computer-Assisted Intervention. Shenzhen, China, 13–17 October, (Cham: Springer), 622–630

Kumar, N, Verma, R, Sharma, S, Bhargava, S, Vahadane, A, and Sethi, A (2017). A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans Med Imaging.* 36, 1550–1560. doi:10.1109/TMI.2017.2677499

Oda, H., Roth, H. R., Chiba, K., Sokolić, J., Kitasaka, T., Oda, M., et al. (2018). "Besnet: boundary-enhanced segmentation of cells in histopathological images". in International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer), 228–236

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst., Man, Cybern.*. 9, 62–66. doi:10.1109/tsmc.1979.4310076

Qu, H., Yan, Z., Riedlinger, G. M., De, S., and Metaxas, D. N. (2019). "Improving nuclei/gland instance segmentation in histopathology images by full resolution neural network and spatial constrained loss". in International Conference on Medical Image Computing and Computer-Assisted Intervention Shenzhen, China, 13–17 October, (Cham: Springer), 378–386

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, Munich, Germany, 5–9 October (Cham: Springer), 234–241

Rother, C., Kolmogorov, V., and Blake, A. (2004). "GrabCut" interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*. 23, 309–314. doi:10.1145/1015706.1015720

Wollmann, T, Gunkel, M, Chung, I, Erfle, H, Rippe, K, and Rohr, K (2019). Gruu-net: Integrated convolutional and gated recurrent neural network for cell segmentation. *Med Image Anal.* 56, 68–79. doi:10.1016/j.media.2019.04.011

Yang, X., Li, H., and Zhou, X. (2006). Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy. *IEEE Trans. Circuits Syst. I.* 53, 2405–2414. doi:10.1109/tcsi.2006.884469

Yu, F., and Koltun, V. (2016). "Multi-scale context aggregation by dilated convolutions". in International Conference on Learning Representations (ICLR)

Zhou, Y., Onder, O. F., Dou, Q., Tsougenis, E., Chen, H., and Heng, P.-A. (2019). "Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation". in International Conference on Information Processing in Medical Imaging (Cham: Springer), 682–693

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2018). "Unet++: A nested u-net architecture for medical image segmentation". in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Editors D. Stoyanov, et al. (Cham: Springer). 3–11

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 22 March 2021
doi: 10.3389/fmolb.2021.622219

Radiomics of Tumor Heterogeneity in Longitudinal Dynamic Contrast-Enhanced Magnetic Resonance Imaging for Predicting Response to Neoadjuvant Chemotherapy in Breast Cancer

*Ming Fan[1], Hang Chen[1], Chao You[2], Li Liu[2], Yajia Gu[2], Weijun Peng[2], Xin Gao[3] and Lihua Li[1]**

[1]Institute of Biomedical Engineering and Instrumentation, Hangzhou Dianzi University, Hangzhou, China, [2]Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai, China, [3]Computational Bioscience Research Center (CBRC), Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

OPEN ACCESS

Edited by:
Nikolay Mikhaylovich Borisov, Moscow Institute of Physics and Technology, Russia

Reviewed by:
Yuchen Qiu, University of Oklahoma, United States
Zhi-Cheng Li, Shenzhen Institutes of Advanced Technolog, China

***Correspondence:**
Lihua Li
lilh@hdu.edu.cn

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 28 October 2020
Accepted: 15 February 2021
Published: 22 March 2021

Citation:
Fan M, Chen H, You C, Liu L, Gu Y, Peng W, Gao X and Li L (2021) Radiomics of Tumor Heterogeneity in Longitudinal Dynamic Contrast-Enhanced Magnetic Resonance Imaging for Predicting Response to Neoadjuvant Chemotherapy in Breast Cancer. Front. Mol. Biosci. 8:622219. doi: 10.3389/fmolb.2021.622219

Breast tumor morphological and vascular characteristics can be changed during neoadjuvant chemotherapy (NACT). The early changes in tumor heterogeneity can be quantitatively modeled by longitudinal dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), which is useful in predicting responses to NACT in breast cancer. In this retrospective analysis, 114 female patients with unilateral unifocal primary breast cancer who received NACT were included in a development (n = 61) dataset and a testing dataset (n = 53). DCE-MRI was performed for each patient before and after treatment (two cycles of NACT) to generate baseline and early follow-up images, respectively. Feature-level changes (delta) of the entire tumor were evaluated by calculating the relative net feature change (deltaRAD) between baseline and follow-up images. The voxel-level change inside the tumor was evaluated, which yielded a Jacobian map by registering the follow-up image to the baseline image. Clinical information and the radiomic features were fused to enhance the predictive performance. The area under the curve (AUC) values were assessed to evaluate the prediction performance. Predictive models using radiomics based on pre- and post-treatment images, Jacobian maps and deltaRAD showed AUC values of 0.568, 0.767, 0.630 and 0.726, respectively. When features from these images were fused, the predictive model generated an AUC value of 0.771. After adding the molecular subtype information in the fused model, the performance was increased to an AUC of 0.809 (sensitivity of 0.826 and specificity of 0.800), which is significantly higher than that of the baseline imaging- and Jacobian map-based predictive models (p = 0.028 and 0.019, respectively). The level of tumor heterogeneity reduction (evaluated by texture feature) is higher in the NACT responders than in the nonresponders. The results suggested that changes in DCE-MRI features that reflect a reduction in tumor heterogeneity following NACT could provide early prediction of breast tumor response.

The prediction was improved when the molecular subtype information was combined into the model.

Keywords: dynamic contrast-enhanced magnetic resonance imaging, breast cancer, neoadjuvant chemotherapy, volumetric change, feature change

INTRODUCTION

Neoadjuvant chemotherapy (NACT) is commonly used in treatment of locally advanced or large operable breast cancers with the aim of downstaging before surgery (Taghian et al., 2004; Kaufmann et al., 2007). The achievement of a pathologic complete response (pCR) is associated with improved survival in patients with breast cancer (Cortazar et al., 2014). Despite the benefit, a subset of patients may experience a failure of treatment and suffer from the side effects of NACT. Therefore, accurate determination of the outcome of NACT is of vital importance for tailored treatment of patients with breast cancer.

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), which is routinely used in clinical practice, provides morphological tumor characteristics and functional information about the blood flow, vascular status and angiogenesis (Pinker et al., 2017; Mann et al., 2019). A systematic review demonstrated that MRI-based radiomics achieved overall promising performance in NACT response prediction (Granzier et al., 2019) and residual tumor size evaluation (Kim et al., 2018a), while a DCE-MRI-based predictive model achieved better accuracy than a model using other parametric images (Fowler et al., 2017). Radiomics features derived from the pretreatment MRI have been used for predicting response to NACT for breast cancer (Uematsu et al., 2010; Braman et al., 2017; Santamaría et al., 2017; Reig et al., 2020). Our previous study used DCE-MRI to identify and validate predictive imaging biomarkers for NACT using two datasets with different imaging protocols for training and testing (Fan et al., 2017). These studies were performed using radiomics of preoperative breast MRI without considering the imaging features of longitudinal changes in MRI features that could be promising in predicting tumor responses to NACT.

The NACT regimen usually takes six to eight cycles to finish the whole treatment procedure. Longitudinal imaging is usually performed during the procedure to monitor and evaluate treatment response. The changes of tumor heterogeneity in DCE-MRI between the preoperative and early NACT (e.g., two cycles of treatment) may provide information for early prediction of the eventual treatment outcome. Previous studies have demonstrated evidence of longitudinal changes in pharmacokinetic parameters (Dogan et al., 2019), tumor sizes (Tudorica et al., 2016), and tumor MRI texture parameters (Parikh et al., 2014; Henderson et al., 2017; Eun et al., 2020; Nadrljanski and Milosevic, 2020) being correlated with responses to NACT in breast cancer patients. These studies mainly analyzed the feature-level changes of the heterogeneity by evaluating longitudinal images within a tumor. Despite the advances of these methods, the voxelwise changes inside a tumor between baseline and post-NACT MRI scans may not be captured by feature analysis of the entire tumor.

To this end, attempts have been conducted by aligning intermediate MRI to baseline images to evaluate changes in tumor heterogeneity in a voxel-by-voxel manner. A previous study implemented an accurate image registration technique using a parametric response map (PRM), which can provide quantitative voxel-based information regarding heterogeneous changes within the tumor during treatment (Galban et al., 2011; Galban et al., 2012; Cho et al., 2014). The nonrigid nature of the human breast requires methods using deformable registration of longitudinal tumor changes during NACT (Li et al., 2009; Ou et al., 2015). A recent study uses deformable methods to capture tumor heterogeneity for early prediction of response to NACT (Jahani et al., 2019). However, whether the quantitative evaluation of longitudinal tumor changes by radiomic analysis is associated with tumor responses is still unclear.

To predict NACT responses in breast cancer, changes in tumor heterogeneity were evaluated both in voxel-by-voxel and feature-level manners using longitudinal DCE-MR images. Radiomic features were extracted at baseline and post-treatment images and the voxel-level map of volumetric change before and after early NACT. Additionally, feature-level changes in tumor heterogeneity were evaluated by calculating the relative net radiomic feature change between baseline and follow-up images (deltaRAD). Predictive models were then established using the radiomic features derived from these images. Our comprehensive analyses demonstrated how the heterogeneity changes in DCE-MRI before and after early NACT could affect the accuracy of prediction of the response to NACT.

MATERIALS AND METHODS

Framework Overview

The framework of this study is illustrated in **Figure 1**. The voxelwise volumetric changes during treatment were evaluated to generate a Jacobian map by aligning the post-treatment MRI scans to the baseline ones. Radiomic analysis was performed on the pre-and post-NACT images and the Jacobian map. Feature-level changes in tumor heterogeneity were obtained by calculating the relative net change (the percent change) in features between baseline and post-NACT scans. Predictive models were established using radiomics based on the evaluation of these longitudinal images to discriminate tumors that responded to NACT from those that did not.

Patient Selection

This study was approved by the Institutional Review Board of Fudan University Shanghai Cancer Center. Due to the retrospective nature of this study, use of a consent form was waived. The data collection and selection procedure in these two cohorts are illustrated in **Figure 2**. The original dataset

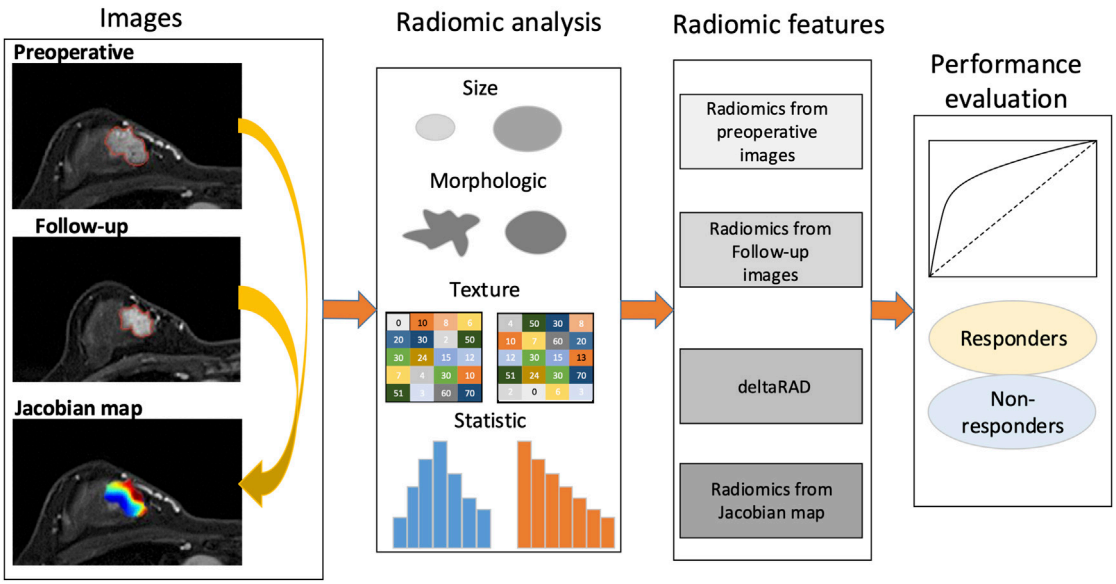


FIGURE 1 | Study framework. A Jacobian map for each tumor was derived based on aligning the post-treatment images to the preoperative ones. Radiomics were calculated using the pre- and the follow-up images, the Jacobian map and the feature changes using longitudinal images (deltaRAD).

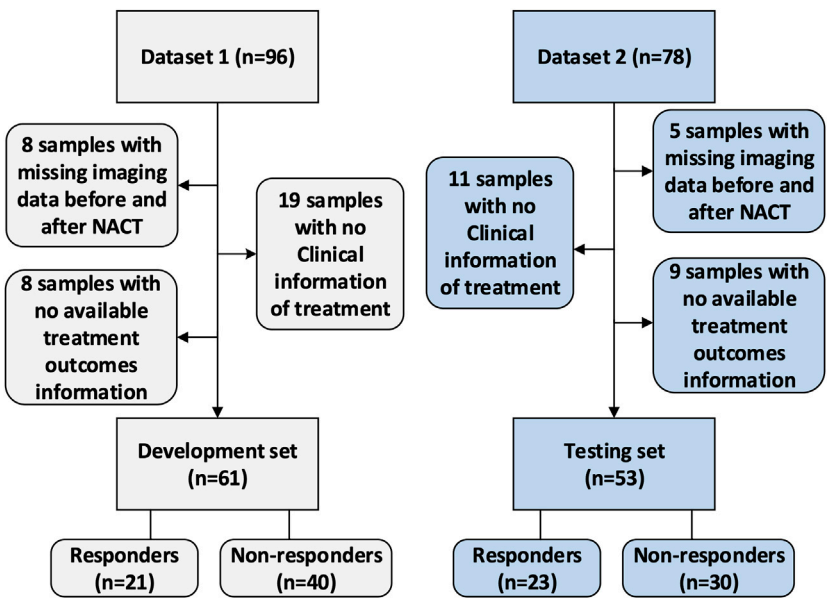


FIGURE 2 | Data selection procedure.

collected from the hospital included 174 samples with paired images acquired at the baseline and post-treatment (after two cycles of NACT). Dataset 1 (the development set) initially included 96 samples. After excluding eight samples with missing imaging sequences at baseline or after early NACT, eight samples with no available treatment outcome data evaluated by the Miler-Payne (MP) score, and 19 samples with no available clinical information, 61 samples were retained in this study. Dataset 2 (the testing set) initially included 78 samples, of which 25 were excluded: 11 with no clinical information, five with incomplete imaging sequences, and nine with no available MP data. The remaining data included 53 samples for testing.

Data Analysis

Pathological response was assessed after the whole cycle of NACT according to the surgical specimen-determined MP grading

TABLE 1 | Imaging parameters in the development and testing datasets.

Parameter	Development dataset	Testing dataset
Repetition time (TR) [ms]	4.5	29
Echo time (TE) [ms]	1.56	4.8
Flip angle (FA) [°]	10	90
Field of view (FOV) [mm]	360 × 360	360 × 360
Matrix	384 × 384	512 × 512
Slice thickness (mm)	2.2	1.48
In-plane resolution (mm)	0.9375 × 0.9375	0.7031 × 0.7031

system by comparing with the preoperative core biopsy (Ogston et al., 2003). This grading system includes five grades. According to a previous study, tumors with MP scores four and five (total cell loss of more than 90%), also termed almost pCR and pCR, respectively, were grouped as responders, while the others (grades 1, 2 or 3 with a total cell loss of up to 90%) were grouped as nonresponders (Zhu et al., 2014). Estrogen receptor (ER), progesterone receptor (PR) and Ki-67 status were defined according to immunohistochemistry (IHC) with streptavidin-peroxidase (SP) detection (Hammond et al., 2010; Wolff et al., 2013). Hormone receptor (HR) positivity was defined as HR and/or ER positive. HER2 positivity was defined as IHC score of 3+ or 2+ with confirmation of gene amplification by fluorescence *in situ* hybridization (FISH) (Wolff et al., 2013). Tumor subtypes were categorized as follows: luminal A (HR-positive and HER2-negative), luminal B (HR-positive and HER2-positive), HER2-enriched (HR-negative and HER2-positive) and triple-negative (HR-negative and HER2-negative) subtypes. The HR-positive and HER2-negative tumors with a Ki-67 expression level higher than 14% were specifically determined to be luminal B subtype tumors.

Imaging Protocols

Imaging was performed following the specific requirements of the hospital. For the development dataset (n = 61), the images were acquired using a 3.0-T scanner (Siemens Healthcare, Erlangen, Germany). DCE-MRI was acquired with a fat-suppressed T1-weighted imaging sequence, which generated one precontrast (S0) followed by five or eight sequential postcontrast image series after injection of a gadobutrol-based contrast agent. The time interval between the first postcontrast image and S0 was 90 s, while the time intervals between the subsequent image series were 43 or 44 s.

For the testing dataset (n = 53), DCE-MRI was acquired using a dedicated 1.5-T breast magnetic resonance imaging system (Aurora Dedicated Breast MRI Systems, United States). The imaging system generated one precontrast image and three postcontrast images at 120, 245, and 371 s after beginning the intravenous administration of gadobutrol injection. The detailed imaging parameters for these two datasets are shown as in **Table 1**.

Image Preprocessing

Nonuniform intensity normalization (N4) bias correction was implemented to reduce the effect of MR imaging artifacts. Images from the patients were resampled to the same spatial resolution for feature extraction. The tumor region-of-interest (ROI) was identified by using a spatial Fuzzy C-means method on the third postcontrast image series where the highest enhancement valuate were usually achieved (Yang et al., 2014; Fan et al., 2020).

Analysis of Volumetric Change in Longitudinal MRI Scans

Voxelwise volumetric changes were evaluated by aligning the follow-up images to preoperative scans by finding an optimal and deformable transformation for image registration (Ou et al., 2015). Based on this approach, the aligned image along with a Jacobian map was generated, in which each pixel of the Jacobian map represented a volumetric shrink/expansion pattern. Specifically, Jacobian values for each voxel greater than one indicate volume expansion, while those less than one indicate volume shrinkage, and those equal to one indicate volume preservation. The Jacobian value is calculated using the following equation (**Eq. 1**):

$$Jacobian\ value\ = \frac{v_{follow-up}}{v_{baseline}} \tag{1}$$

where v_2 denotes the resisted voxel volume in follow-up image, and v_1 denotes the voxel volume in the baseline image.

Radiomic Features

Features were extracted from the tumor ROI using a publicly available radiomics analysis software, Pyradiomics (van Griethuysen et al., 2017). For each ROI, 102 features were calculated, including the shape (n = 14), first-order statistics (n = 18), texture features using gray level cooccurrence matrix (GLCM) (n = 24), gray-level run-length matrix (GLRLM) (n = 16), gray-level size-zone matrix (GLSZM) (n = 16) and gray-level dependence matrix (GLDM) (n = 14). The imaging heterogeneity of the entire tumor was evaluated based on the subtraction images between the intermediate image series that unusually exhibited the maximum enhancement signal and the precontrast image. Radiomics features were calculated on the pre- and post-treatment images and the Jacobian map.

Feature-Level Changes in Tumor Heterogeneity

Feature-level changes were calculated by the relative net change between the features derived from the baseline and the follow-up image. For the i^{th} radiomics feature (f^i) calculated from the tumor ROI, the feature change f^i_Δ is illustrated as shown in the following equation (**Eq. 2**):

$$f^i_\Delta\ = \frac{f^i_{baseline} - f^i_{follow-up}}{f^i_{baseline}} \tag{2}$$

where $f^i_{baseline}$ stands for the feature i obtained from the baseline image, and $f^i_{follow-up}$ indicates the feature i from the follow-up image.

Statistical Analysis and Machine Learning Methods

The distributions of the histopathological information of the molecular subtypes, menopausal status, family history between the development and testing groups were compared by using the χ^2

TABLE 2 | Patient characteristics.

	All	Development set	Testing set	*p*-value
Number	114	61 (54%)	53 (46%)	
Age	48 (27–79)	49 (27–66)	47 (29–79)	0.407[a]
Menopausal status				0.670[b]
Pre	46 (40%)	23 (38%)	23 (43%)	
Post	68 (60%)	38 (62%)	30 (57%)	
Family history				0.642[b]
No	87 (76%)	45 (74%)	42 (79%)	
Yes	27 (24%)	16 (26%)	11 (21%)	
Miller Payne				0.706[c]
1	9 (8%)	6 (10%)	3 (6%)	
2	21 (18%)	10 (16%)	11 (21%)	
3	40 (35%)	24 (40%)	16 (30%)	
4	10 (9%)	5 (8%)	5 (9%)	
5	34 (30%)	16 (26%)	18 (34%)	
Molecular subtypes				0.409[b]
Luminal A	12 (10%)	9 (15%)	3 (5%)	
Luminal B	58 (51%)	30 (49%)	28 (53%)	
Basal-like	20 (18%)	9 (15%)	11 (21%)	
HER-2	24 (21%)	13 (21%)	11 (21%)	

[a]*Analysis of variance.*
[b]χ^2 *test with Yates' continuity correction.*
[c]*Fisher's exact test.*

test or Fisher's exact test when the expected frequency in any tablet was less than five. Analysis of variance (ANOVA) was performed to compare continuous variables between the development and the testing groups. The area under the receiver operating characteristic (ROC) curve (AUC) was calculated to assess the performance of the predictive model. The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated. The sensitivity and specificity were determined at the operation point at ROC curve by using the Youden index by the maximum sum of the specificity and the sensitivity. Statistical tests with p values less than 0.05 were considered significant.

A support vector machine (SVM) with a Gaussian kernel was used as a base classifier for prediction. Predictive model establishment and model tuning were performed on the development set and were tested on the testing set. SVM-recursive feature elimination (RFE) was used to rank the features that were most relevant to the target, and these were then sequentially added into the predictive model. The feature sets were fed into the predictive model, in which the SVM parameters α and γ were tuned using a grid search method in each iteration with a 10-fold cross-validation framework. An optimized model with the selected feature subset and the tuned model parameters was established using all the samples in the development set and was applied to the testing set to evaluate the model performance. Statistical analysis and machine learning methods were performed using R (version 4.0) and Matlab (MathWorks, Natick, Massachusetts, version 2018 b).

RESULTS

Patient

Patient characteristics including age, menopausal status, family history, molecular subtypes and MP grade are illustrated in **Table 2**. The development dataset included 61 samples (mean age 49, ranges from 27 to 66°years), while the testing dataset included 53 samples (mean age 47, range from 29 to 79°years). There were 44 (38.6%) patients who had an MP grade larger than three (i.e., 4, 5), and they were categorized as the responders, while the others (n = 70, 61.4%) who had an MP grade of no more than three (i.e., 1, 2 or 3) were defined as the nonresponders. No significant differences in histological information were observed between the development and testing datasets ($p > 0.05$, **Table 2**).

Voxelwise Changes in Tumor Heterogeneity Associated With the Response to NACT

After registering the follow-up images to the baseline ones, a Jacobian map was obtained for each tumor that reflects the level of voxelwise volumetric shrink/expansion. An example of a statistical feature (e.g., mean value) calculated on the Jacobian map of tumors is illustrated in **Figure 3**. Tumor volume was reduced in both the nonresponse (**Figures 3A–C**) and the response (**Figures 3D–F**) groups after NACT. The mean Jacobian value inside the tumor was significantly higher in the nonresponders than in the responders, with a p value of $4.9e^{-5}$ (**Figure 3G**). This result indicated that a high Jacobian value that represents a lower level of voxelwise shrink inside a tumor is associated with a failure of treatment.

In addition to statistical features, examples of texture features derived from tumor Jacobian maps are illustrated in **Figure 4**. A low MP grade (nonresponder) patient showed a lower level of volume shrinkage (**Figures 4A,B**) after early NACT than a patient with a higher MP grade (responder) (**Figures 4D,E**); this pattern is illustrated in the Jacobian map (**Figures 4C,F**). The texture feature (large dependence high gray-level emphasis) obtained from the Jacobian map were significantly higher in the nonresponders than in the responders (**Figure 4G**, $p = 2.45e^{-4}$). This result suggested that a higher level of this texture feature, which reflects a higher voxelwise spatial rearrangement heterogeneity of the shrinkage/expansion pattern inside a tumor during NACT, is more likely to be associated with a worse response to NACT.

Feature-Level Changes in Tumor Heterogeneity Associated With Response to NACT

To assess how the tumor heterogeneity changed during treatment, radiomics analysis was conducted on the baseline and the follow-up images. It should be noted that the relative net change in volume size for the entire tumor between pre- and post-treatment images showed no significant (p = 0.09) differences between the responders and nonresponders (**Supplementary Figure S1**). This result suggests that volumetric changes in the entire tumor after early NACT may not be significantly related to the eventual treatment outcomes.

A more significant (p = 0.001) decrease in a statistical feature (energy) after early NACT was observed in the

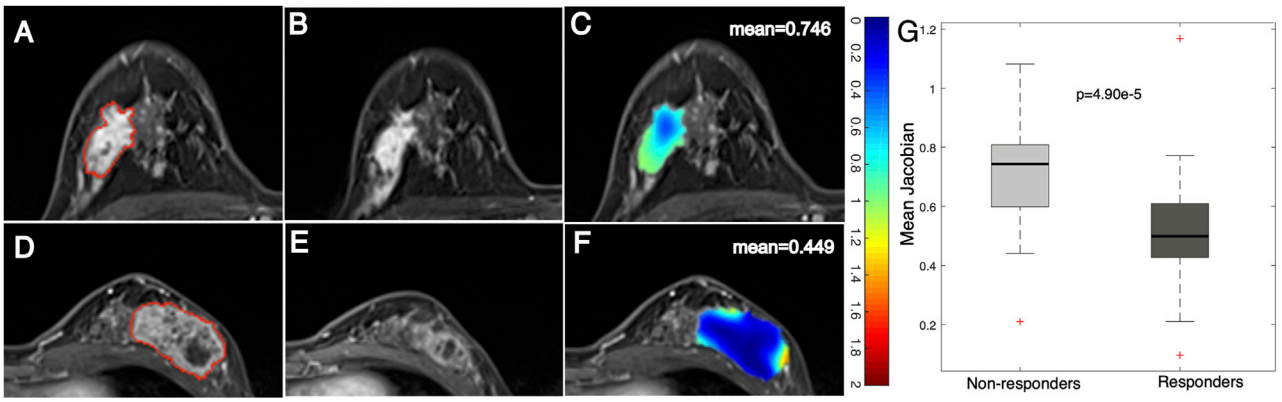


FIGURE 3 | Example images and distribution of mean Jacobian values in nonresponders and responders. Images from a breast cancer patient (aged 45 years old) with a low MP (nonresponder) **(A)** pre- and **(B)** post-treatment and **(C)** a Jacobian map of the ROI (mean Jacobian value = 0.746). Images from a breast cancer patient (aged 41 years old) with a high MP (responder) **(D)** pre- and **(E)** post-treatment and **(F)** a Jacobian map of the ROI (mean Jacobian value = 0.449). **(G)** Boxplot representing the feature distribution between nonresponders and responders.

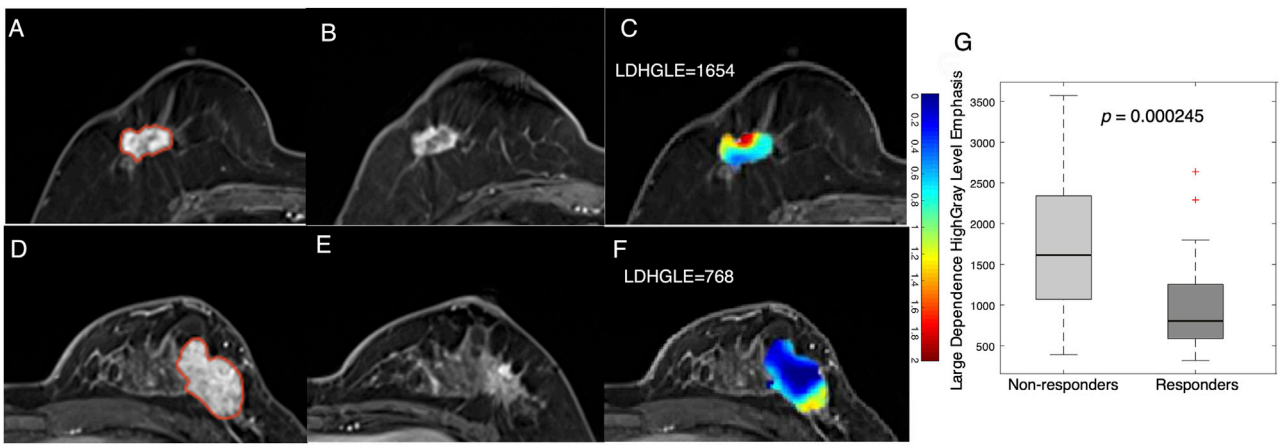


FIGURE 4 | Examples feature of large dependence high gray-level emphasis (LDHGLE) in nonresponders and responders. Images from a nonresponder breast cancer patient (aged 59 years old) **(A)** pre- and **(B)** post-treatment and **(C)** a Jacobian map of the tumor ROI (LDHGLE = 1654). Images from a responder breast cancer patient [aged 43 years old) **(D)**] pre- and **(E)** post-treatment and **(F)** a Jacobian map of the tumor ROI (LDHGLE = 768). **(G)** Boxplot representing the feature distributions in nonresponders and responders.

responders than in the nonresponders (**Figure 5**). This feature measures the magnitude of voxel values, and a higher value suggests a greater sum of the squares of these values. The result suggests that a decrease in the enhancement level of tumors is associated with NACT response in breast cancer.

An example of a texture feature (i.e., autocorrelation) is also illustrated in **Figure 6**. This feature value was significantly reduced after early NACT in the responders (p = 0.006), while the difference was not significant in the nonresponders (p = 0.241). This feature measures the level of the fineness and coarseness of the texture of an object, in which a high value is correlated with high gray-level heterogeneity within the tumor. The results suggest that the level of tumor heterogeneity reduction is higher in the NACT responders than in the nonresponders.

Fusion of Longitudinal MRI Features for Predicting Response to NACT

To evaluate the collective effect of longitudinal radiomics, the features from different images were combined and evaluated. The individual features from the images at baseline, post-treatment, Jacobian map and deltaRAD features were evaluated, and the results showed that features from the follow-up image have the highest performance (in terms of AUC values), while the deltaRAD features and Jacobian map-based features showed intermediate performance (**Supplementary Figure S2**).

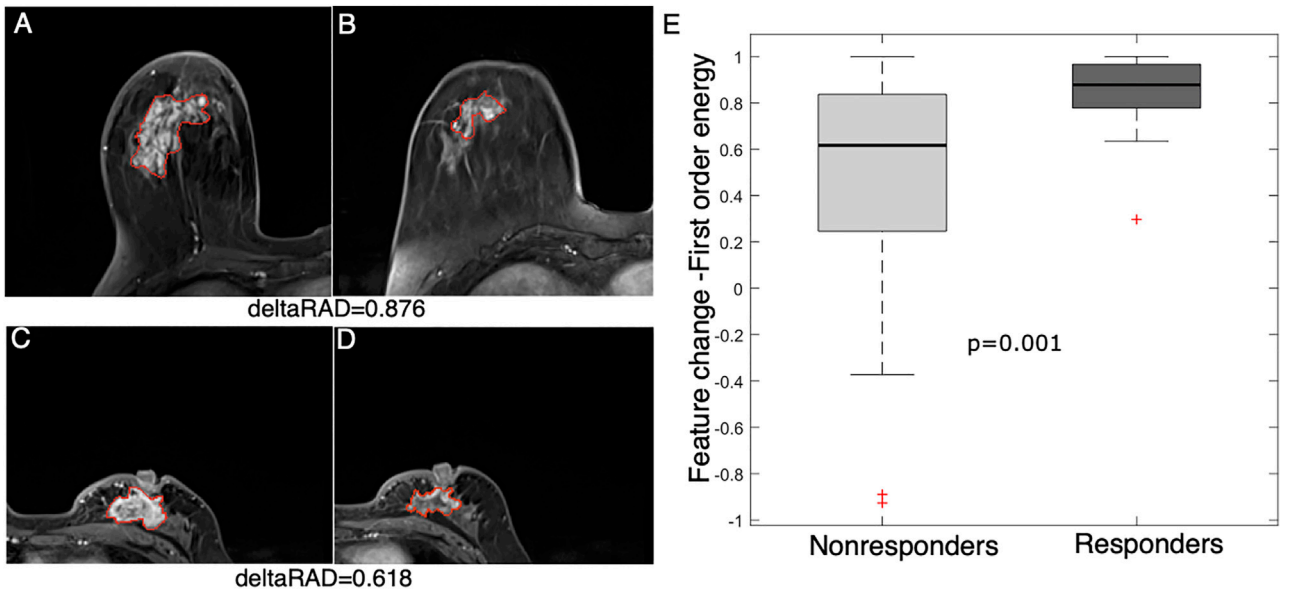


FIGURE 5 | Feature (energy) change between baseline and early NACT images. Images from a 58 year-old woman with a deltaRAD value of 0.876 in the responders at **(A)** baseline and **(B)** early NACT. Images from a 49 year-old woman with a deltaRAD value of 0.618 in the nonresponders at **(C)** baseline and **(D)** early NACT. **(E)** The distribution of the change in the energy value is shown in the boxplot, in which the feature value is significantly higher in responders than in nonresponders.

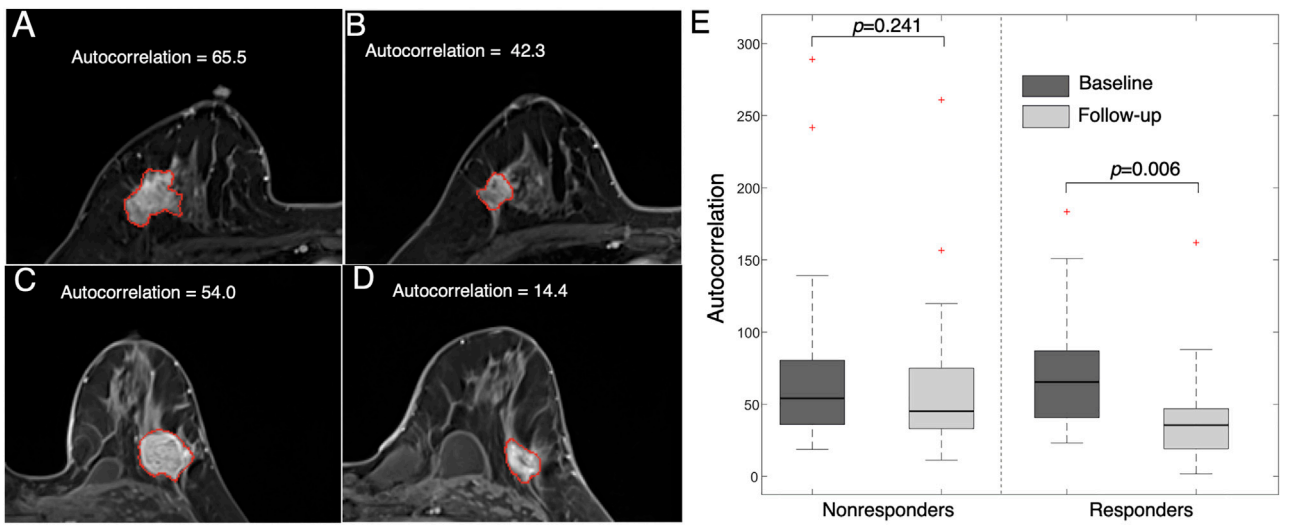


FIGURE 6 | Images representing feature (autocorrelation) changes between pre- and post-treatment images. Images from a 47 year-old woman who responded to NACT (high MP grade) at **(A)** baseline (autocorrelation = 65.5) and **(B)** follow-up (autocorrelation = 42.3). Images from a 36 year-old woman who did not respond to NACT (low MP grade) at **(C)** baseline (autocorrelation = 54.0) and **(D)** follow-up (autocorrelation = 14.4). **(E)** Boxplot showing that the feature value is significantly reduced in responders (p = 0.006) but is not significantly changed in nonresponders (p = 0.241).

Radiomic features from these images were used separately to establish predictive models in the development set and was tested on the testing set (**Table 3**; **Figure 7**). Among these, the predictive model based on the baseline image generated lowest performance with an AUC of 0.568 (sensitivity of 0.913 at a specificity of 0.367). Radiomic features based on Jacobian map, follow-up image and deltaRAD showed a higher prediction performance with AUC of 0.628, 0.757 and 0.718, respectively. When the features from these images were fused, the classifier generated an AUC of 0.771 with sensitivity of 0.522 and specificity of 0.967. Finally, imaging features were combined with the clinical and histologic information for prediction to facilitate a more accurate prediction. The results showed an improved performance with an AUC value of 0.809 (sensitivity of 0.826 at a specificity of 0.800),

TABLE 3 | Performance of predictive model based on images at longitudinal times.

Images	AUC (±SE)	SD	*p* value	Sensitivity	Specificity	PPV	NPV
Baseline image	0.568 ± 0.155	0.079	0.028	0.913	0.367	0.525	0.846
Follow-up image	0.767 ± 0.128	0.065	0.508	0.565	0.900	0.813	0.730
DeltaRAD	0.726 ± 0.137	0.070	0.301	0.913	0.533	0.600	0.889
Jacobian map	0.630 ± 0.154	0.079	0.019	0.609	0.700	0.609	0.700
Feature fusion	0.771 ± 0.136	0.069	0.356	0.522	0.967	0.923	0.725
Feature + MS	0.809 ± 0.131	0.067	—	0.826	0.800	0.760	0.857

SE, standard error; SD, standard derivation; deltaRAD, relative net feature change between baseline and follow-up images; MS, molecular subtype. p *value indicates significance of the comparison between baseline imaging- and the other image-based predictive models.*

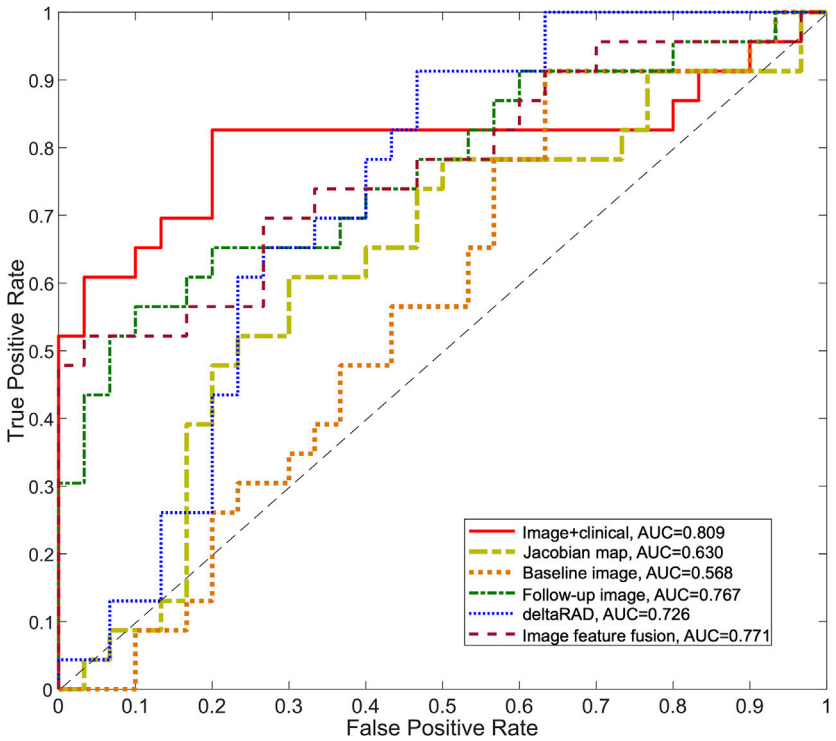


FIGURE 7 | ROC curves for the predictive models using longitudinal images. The ROC curves for the predictive model using deltaRAD and radiomics derived from pre- and post-treatment images, the Jacobian map and the fused imaging features are shown. The ROC curve of the predictive model combining imaging features and molecular subtype information is also shown.

which is significantly better than the baseline image (p = 0.028) and the Jacobian map (p = 0.019) based predictive model.

DISCUSSION

During NACT, the breast tumor size and morphological and functional changes are associated with the eventual treatment outcomes. In this study, the pattern of the changes in tumor heterogeneity during NACT was evaluated using baseline and post-treatment images to predict responses to NACT in breast cancer. The voxelwise shrinkage/expansion inside the tumor and the feature-level changes of the entire tumor were both obtained. Radiomics features from longitudinal images and the changes in tumor heterogeneity were fused for the prediction. The molecular subtype information was combined with radiomics features, which generated an increased prediction performance.

Previous studies have conducted radiomic analysis using features derived from tumors for NACT response prediction. Jahani et al. analyzed voxelwise changes in DCE-MRI features to characterize heterogeneous changes within the tumor and to predict pCR and recurrence free survivals (Jahani et al., 2019). An earlier study evaluated image feature-level changes in tumor heterogeneity to assess for pCR to NACT (Parikh et al., 2014). In our study, radiomic features based on changes in tumor heterogeneity were evaluated in both feature- and voxel-level to facilitate a quantitative analysis of longitudinal heterogeneity during treatment in breast cancer. A recent study extracted texture and statistical features and identified that tumor kurtosis in T_2-weighted MR images was independently

associated with pCR in non-triple-negative breast cancer (Chamming's et al., 2018). Additionally, the molecular subtypes were associated with the prediction accuracy of NACT response (Drisis et al., 2016; Liu et al., 2019). In this study, we have incorporated molecular subtype information in the predictive model and observed the highest performance, which is partly consistent with previous study.

In this study, radiomics analysis of Jacobian maps showed that statistical features (e.g., mean) and texture features (e.g., large dependence high gray level emphasis) decreased after early NACT, while the level of feature reduction was lower in the responders than in the nonresponders. On the other hand, the voxelwise volumetric reduction inside tumors was significantly associated with the responders. Additionally, texture features (e.g., large dependence high gray-level emphasis) were reduced after early NACT, and the level of the reduction was higher in patients who responded to NACT than in those who did not. In our results, tumor heterogeneity was decreased after early NACT, and more importantly, the high level of reduction in heterogeneity was associated with good response to NACT. This indicated that decreased heterogeneity within a tumor may likely be exhibit in the patients who benefitted from the NACT.

In addition to the evaluation of voxel-vize volumetric changes by image registration, longitudinal feature-level changes between the baseline and follow-up images were also evaluated for their associations with tumor response to NACT. In our study, the performance of the model based on vascular characteristics measured by DCE-MRI was higher than that of the model based on morphologic features, which is partly consistent with the findings of a previous study that dynamic features have better accuracy in response prediction than tumor size (Marinovich et al., 2012). Our results indicated that tumor heterogeneity-related features are decreased after treatment, and the extent is higher in responders than in nonresponders. Therefore, longitudinal feature changes in tumor heterogeneity, rather than size changes of the entire tumor, might be more correlated with tumor response to NACT.

We observed a relatively lower performance in terms of AUC for features from the baseline images. A related study identified significant change in the tumor maximum diameter between the responders and nonresponders (Minarikova et al., 2017). In our study, changes in tumor heterogeneity at the feature level and voxel level were both evaluated, and predictive performance was improved after fusing the features from different images at varied times. The results suggested that multiple levels of features and different stages of features at treatment may be complementary, and altogether, these contributed to enhanced model performance.

Despite the potential significance of tumor radiomics using longitudinal images in this study, several limitations should also be addressed. First, only the tumor region was analyzed for image feature extraction. It would also be interesting to analyze the peritumoral tissues that surrounds the tumor (Kim et al., 2018b) to conduct a comprehensively analysis of the pattern of heterogeneity on baseline and post-treatment images. Second, this was a retrospective study, and the sample size was relatively small to conduct a fair statistical analysis. Further studies with more samples and refined analyses should be conducted to confirm the findings of this study. Third, features were derived from two datasets with different magnetic field strengths (3.0 and 1.5 T for the development and testing datasets, respectively), which may have affected the feature calculations and induced bias. Despite this limitation, the features were calculated based on the relative differences in the feature/voxel values between baseline and follow-up images, which may have partly reduced the bias between different protocols. In our study, radiomics features were calculated using publicly available Pyradiomics software, with the aim of ensuring the repeatability of this study (van Griethuysen et al., 2017).

In conclusion, longitudinal changes in tumor heterogeneity at the voxel and feature levels were examined to determine their contribution to the prediction of tumor response. It was found that molecular subtypes add more predictive power in assessing the response to NACT.

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

This study was approved by the Institutional Review Board of Fudan University Shanghai Cancer Center. Due to the retrospective nature of this study, use of a consent form was waived.

AUTHOR CONTRIBUTIONS

MF and LL designed the study. HC, CY, and LL performed the image processing, machine learning and statistical analyses. MF and LL wrote the manuscript. YG, WP, and XG analyzed data. All authors reviewed the manuscript.

FUNDING

This work was supported in part by grants from the National Natural Science Foundation of China (Nos. 61731008, 61871428, U1809205, 81901703, and 82071878), the Natural Science Foundation of Zhejiang Province of China (No. LJ19H180001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.622219/full#supplementary-material.

REFERENCES

Braman, N. M., Etesami, M., Prasanna, P., Dubchuk, C., Gilmore, H., Tiwari, P., et al. (2017). Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res.* 19 (1), 57. doi:10.1186/s13058-017-0846-1

Chamming's, F., Ueno, Y., Ferré, R., Kao, E., Jannot, A. S., Chong, J., et al. (2018). Features from computerized texture analysis of breast cancers at pretreatment MR imaging are associated with response to neoadjuvant chemotherapy. *Radiology* 286 (2), 412–420. doi:10.1148/radiol.2017170143

Cho, N., Im, S. A., Park, I. A., Lee, K. H., Li, M., Han, W., et al. (2014). Breast cancer: early prediction of response to neoadjuvant chemotherapy using parametric response maps for MR imaging. *Radiology* 272 (2), 385–396. doi:10.1148/radiol.14131332

Cortazar, P., Zhang, L., Untch, M., Mehta, K., Costantino, J. P., Wolmark, N., et al. (2014). Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* 384 (9938), 164–172. doi:10.1016/S0140-6736(13)62422-8

Dogan, B. E., Yuan, Q., Bassett, R., Guvenc, I., Jackson, E. F., Cristofanilli, M., et al. (2019). Comparing the performances of magnetic resonance imaging size vs pharmacokinetic parameters to predict response to neoadjuvant chemotherapy and survival in patients with breast cancer. *Curr. Probl. Diagn. Radiol.* 48 (3), 235–240. doi:10.1067/j.cpradiol.2018.03.003

Drisis, S., Metens, T., Ignatiadis, M., Stathopoulos, K., Chao, S. L., and Lemort, M. (2016). Quantitative DCE-MRI for prediction of pathological complete response following neoadjuvant treatment for locally advanced breast cancer: the impact of breast cancer subtypes on the diagnostic accuracy. *Eur. Radiol.* 26 (5), 1474–1484. doi:10.1007/s00330-015-3948-0

Eun, N. L., Kang, D., Son, E. J., Park, J. S., Youk, J. H., Kim, J. A., et al. (2020). Texture analysis with 3.0-T MRI for association of response to neoadjuvant chemotherapy in breast cancer. *Radiology* 294 (1), 31–41. doi:10.1148/radiol.2019182718

Fan, M., Wu, G., Cheng, H., Zhang, J., Shao, G., and Li, L. (2017). Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients. *Eur. J. Radiol.* 94, 140–147. doi:10.1016/j.ejrad.2017.06.019

Fan, M., Yuan, W., Zhao, W., Xu, M., Wang, S., Gao, X. X., et al. (2020). Joint prediction of breast cancer histological grade and ki-67 expression level based on DCE-MRI and DWI radiomics. *IEEE J. Biomed. Health Inform.* 24 (6), 1632–1642. doi:10.1109/JBHI.2019.2956351

Fowler, A. M., Mankoff, D. A., and Joe, B. N. (2017). Imaging neoadjuvant therapy response in breast cancer. *Radiology* 285 (2), 358–375. doi:10.1148/radiol.2017170180

Galbán, C. J., Chenevert, T. L., Meyer, C. R., Tsien, C., Lawrence, T. S., Hamstra, D. A., et al. (2011). Prospective analysis of parametric response map-derived MRI biomarkers: identification of early and distinct glioma response patterns not predicted by standard radiographic assessment. *Clin. Cancer Res.* 17 (14), 4751–4760. doi:10.1158/1078-0432.CCR-10-2098

Galbán, C. J., Han, M. K., Boes, J. L., Chughtai, K. A., Meyer, C. R., Johnson, T. D., et al. (2012). Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat. Med.* 18 (11), 1711–1715. doi:10.1038/nm.2971

Granzier, R. W. Y., van Nijnatten, T. J. A., Woodruff, H. C., Smidt, M. L., and Lobbes, M. B. I. (2019). Exploring breast cancer response prediction to neoadjuvant systemic therapy using MRI-based radiomics: a systematic review. *Eur. J. Radiol.* 121, 108736. doi:10.1016/j.ejrad.2019.108736

Hammond, M. E., Hayes, D. F., Wolff, A. C., Mangu, P. B., and Temin, S. (2010). American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J. Oncol. Pract.* 6 (4), 195–197. doi:10.1200/JOP.777003

Henderson, S., Purdie, C., Michie, C., Evans, A., Lerski, R., Johnston, M., et al. (2017). Interim heterogeneity changes measured using entropy texture features on T_2-weighted MRI at 3.0 T are associated with pathological response to neoadjuvant chemotherapy in primary breast cancer. *Eur. Radiol.* 27 (11), 4602–4611. doi:10.1007/s00330-017-4850-8

Jahani, N., Cohen, E., Hsieh, M. K., Weinstein, S. P., Pantalone, L., Hylton, N., et al. (2019). Prediction of treatment response to neoadjuvant chemotherapy for breast cancer via early changes in tumor heterogeneity captured by DCE-MRI registration. *Sci. Rep.* 9 (1), 12114. doi:10.1038/s41598-019-48465-x

Kaufmann, M., von Minckwitz, G., Bear, H. D., Buzdar, A., McGale, P., Bonnefoi, H., et al. (2007). Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: new perspectives 2006. *Ann. Oncol.* 18 (12), 1927–1934. doi:10.1093/annonc/mdm201

Kim, S. Y., Cho, N., Park, I. A., Kwon, B. R., Shin, S. U., Kim, S. Y., et al. (2018a). Dynamic contrast-enhanced breast MRI for evaluating residual tumor size after neoadjuvant chemotherapy. *Radiology* 289 (2), 327–334. doi:10.1148/radiol.2018172868

Kim, S. Y., Cho, N., Shin, S. U., Lee, H. B., Han, W., Park, I. A., et al. (2018b). Contrast-enhanced MRI after neoadjuvant chemotherapy of breast cancer: lesion-to-background parenchymal signal enhancement ratio for discriminating pathological complete response from minimal residual tumour. *Eur. Radiol.* 28 (7), 2986–2995. doi:10.1007/s00330-017-5251-8

Li, X., Dawant, B. M., Welch, E. B., Chakravarthy, A. B., Freehardt, D., Mayer, I., et al. (2009). A nonrigid registration algorithm for longitudinal breast MR images and the analysis of breast tumor response. *Magn. Reson. Imaging* 27 (9), 1258–1270. doi:10.1016/j.mri.2009.05.007

Liu, Z., Li, Z., Qu, J., Zhang, R., Zhou, X., Li, L. L. F., et al. (2019). Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: a multicenter study. *Clin. Cancer Res.* 25 (12), 3538–3547. doi:10.1158/1078-0432.CCR-18-3190

Mann, R. M., Cho, N., and Moy, L. (2019). Breast MRI: state of the art. *Radiology* 292 (3), 520–536. doi:10.1148/radiol.2019182947

Marinovich, M. L., Sardanelli, F., Ciatto, S., Mamounas, E., Brennan, M., Macaskill, P., et al. (2012). Early prediction of pathologic response to neoadjuvant therapy in breast cancer: systematic review of the accuracy of MRI. *Breast* 21 (5), 669–677. doi:10.1016/j.breast.2012.07.006

Minarikova, L., Bogner, W., Pinker, K., Valkovič, L., Zaric, O., Bago-Horvath, Z., et al. (2017). Investigating the prediction value of multiparametric magnetic resonance imaging at 3 T in response to neoadjuvant chemotherapy in breast cancer. *Eur. Radiol.* 27 (5), 1901–1911. doi:10.1007/s00330-016-4565-2

Nadrljanski, M. M., and Milosevic, Z. C. (2020). Tumor texture parameters of invasive ductal breast carcinoma in neoadjuvant chemotherapy: early identification of non-responders on breast MRI. *Clin. Imaging* 65, 119–123. doi:10.1016/j.clinimag.2020.04.016

Ogston, K. N., Miller, I. D., Payne, S., Hutcheon, A. W., Sarkar, T. K., Smith, I., et al. (2003). A new histological grading system to assess response of breast cancers to primary chemotherapy: prognostic significance and survival. *Breast* 12 (5), 320–327. doi:10.1016/s0960-9776(03)00106-1

Ou, Y., Weinstein, S. P., Conant, E. F., Englander, S., Da, X., Gaonkar, B., et al. (2015). Deformable registration for quantifying longitudinal tumor changes during neoadjuvant chemotherapy. *Magn. Reson. Med.* 73 (6), 2343–2356. doi:10.1002/mrm.25368

Parikh, J., Selmi, M., Charles-Edwards, G., Glendenning, J., Ganeshan, B., Verma, H., et al. (2014). Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy. *Radiology* 272 (1), 100–112. doi:10.1148/radiol.14130569

Pinker, K., Helbich, T. H., and Morris, E. A. (2017). The potential of multiparametric MRI of the breast. *Br. J. Radiol.* 90 (1069), 20160715. doi:10.1259/bjr.20160715

Reig, B., Heacock, L., Lewin, A., Cho, N., and Moy, L. (2020). Role of MRI to assess response to neoadjuvant therapy for breast cancer. *J. Magn. Reson. Imaging* 52, 1587–1606. doi:10.1002/jmri.27145

Santamaría, G., Bargalló, X., Fernández, P. L., Farrús, B., Caparrós, X., and Velasco, M. (2017). Neoadjuvant systemic therapy in breast cancer: association of contrast-enhanced MR imaging findings, diffusion-weighted imaging findings, and tumor subtype with tumor response. *Radiology* 283 (3), 663–672. doi:10.1148/radiol.2016160176

Taghian, A., Jeong, J. H., Mamounas, E., Anderson, S., Bryant, J., Deutsch, M., et al. (2004). Patterns of locoregional failure in patients with operable breast cancer treated by mastectomy and adjuvant chemotherapy with or without tamoxifen and without radiotherapy: results from five national surgical adjuvant breast

and bowel project randomized clinical trials. *J. Clin. Oncol.* 22 (21), 4247–4254. doi:10.1200/JCO.2004.01.042

Tudorica, A., Oh, K. Y., Chui, S. Y., Roy, N., Troxell, M. L., Naik, A., et al. (2016). Early prediction and evaluation of breast cancer response to neoadjuvant chemotherapy using quantitative DCE-MRI. *Transl. Oncol.* 9 (1), 8–17. doi:10.1016/j.tranon.2015.11.016

Uematsu, T., Kasami, M., and Yuen, S. (2010). Neoadjuvant chemotherapy for breast cancer: correlation between the baseline MR imaging findings and responses to therapy. *Eur. Radiol.* 20 (10), 2315–2322. doi:10.1007/s00330-010-1813-8

van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., et al. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 77 (21), e104–e107. doi:10.1158/0008-5472.CAN-17-0339

Wolff, A. C., Hammond, M. E., Hicks, D. G., Dowsett, M., McShane, L. M., Allison, K. H., et al. (2013). Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: american society of clinical oncology/college of american pathologists clinical practice guideline update. *J. Clin. Oncol.* 31 (31), 3997–4013. doi:10.1200/JCO.2013.50.9984

Yang, Q., Li, L., Zhang, J., Shao, G., Zhang, C., and Zheng, B. (2014). Computer-aided diagnosis of breast DCE-MRI images using bilateral asymmetry of contrast enhancement between two breasts. *J. Digit Imaging* 27 (1), 152–160. doi:10.1007/s10278-013-9617-4

Zhu, Q., Wang, L., Tannenbaum, S., Ricci, A., DeFusco, P., and Hegde, P. (2014). Pathologic response prediction to neoadjuvant chemotherapy utilizing pretreatment near-infrared imaging parameters and tumor pathologic criteria. *Breast Cancer Res.* 16 (5), 456. doi:10.1186/s13058-014-0456-0

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 30 March 2021
doi: 10.3389/fmolb.2021.614207

Chest CT Images for COVID-19: Radiologists and Computer-Based Detection

Qingli Dou[1†], Jiangping Liu[1†], Wenwu Zhang[1], Yanan Gu[1], Wan-Ting Hsu[2], Kuan-Ching Ho[3], Hoi Sin Tong[4], Wing Yan Yu[4] and Chien-Chang Lee[5]*

[1] Department of Emergency Medicine, The Second Affiliated Hospital of Shenzhen University, Shenzhen, China, [2] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, United States, [3] Radiology Department, St George Hospital Sydney, Kogarah, NSW, Australia, [4] Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pok Fu Lam, Hong Kong, [5] Department of Emergency Medicine, National Taiwan University Hospital, Taipei, Taiwan

OPEN ACCESS

Edited by:
Lihua Li,
Hangzhou Dianzi University, China

Reviewed by:
Ana Cláudia Coelho,
University of Trás-os-Montes and Alto Douro, Portugal
Prabhat Kumar Sharma,
Children's Hospital of Philadelphia, United States
Yi Jiang,
Sichuan University, China

***Correspondence:**
Chien-Chang Lee
cclee100@gmail.com;
hit3transparency@gmail.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: 05 October 2020
Accepted: 02 March 2021
Published: 30 March 2021

Citation:
Dou Q, Liu J, Zhang W, Gu Y, Hsu W-T, Ho K-C, Tong HS, Yu WY and Lee C-C (2021) Chest CT Images for COVID-19: Radiologists and Computer-Based Detection. Front. Mol. Biosci. 8:614207. doi: 10.3389/fmolb.2021.614207

Background: Characteristic chest computed tomography (CT) manifestation of 2019 novel coronavirus (COVID-19) was added as a diagnostic criterion in the Chinese National COVID-19 management guideline. Whether the characteristic findings of Chest CT could differentiate confirmed COVID-19 cases from other positive nucleic acid test (NAT)-negative patients has not been rigorously evaluated.

Purpose: We aim to test whether chest CT manifestation of 2019 novel coronavirus (COVID-19) can be differentiated by a radiologist or a computer-based CT image analysis system.

Methods: We conducted a retrospective case-control study that included 52 laboratory-confirmed COVID-19 patients and 80 non-COVID-19 viral pneumonia patients between 20 December, 2019 and 10 February, 2020. The chest CT images were evaluated by radiologists in a double blind fashion. A computer-based image analysis system (uAI System, Lianying Inc., Shanghai, China) detected the lesions in 18 lung segments defined by Boyden classification system and calculated the infected volume in each segment. The number and volume of lesions detected by radiologist and computer system was compared with Chi-square test or Mann-Whitney *U* test as appropriate.

Results: The main CT manifestations of COVID-19 were multi-lobar/segmental peripheral ground-glass opacities and patchy air space infiltrates. The case and control groups were similar in demographics, comorbidity, and clinical manifestations. There was no significant difference in eight radiologist identified CT image features between the two groups of patients. There was also no difference in the absolute and relative volume of infected regions in each lung segment.

Conclusion: We documented the non-differentiating nature of initial chest CT image between COVID-19 and other viral pneumonia with

suspected symptoms. Our results do not support CT findings replacing microbiological diagnosis as a critical criterion for COVID-19 diagnosis. Our findings may prompt re-evaluation of isolated patients without laboratory confirmation.

Keywords: COVID-19, 2019-nCoV, chest computed tomography, computer-aided detection, computer-based detection

INTRODUCTION

Due to high transmissibility and so far lack of proven treatment, the 2019 novel coronavirus disease, 2019-nCoV, has quickly disseminated worldwide (Bogoch et al., 2020; Dong et al., 2020). As symptoms of COVID-19 are similar to other acute respiratory infections, diagnosis relies on positive nucleic acid test (NAT). Given the long turnaround time and suboptimal sensitivity of NAT, chest computed tomography (CT) was proposed as a first line diagnostic tool by the Chinese national guideline (trial version 5) (Ai et al., 2020). This however, created problems, particularly resulting isolating patients with COVID-19 with patients with similar respiratory symptoms due to other infections, as well as delaying appropriate treatment for other treatable infections. We believe findings on CT might not be specific enough to differentiate infection from COVID-19 from non-COVID-19 etiology. It is not a coincidence that after the new definition implementation, 14,840 new cases with 242 deaths were reported on February 13th, which was the record by far reported in a single day since the outbreak (Coronavirus COVID-19 Global Cases by John Hopkins, 2020).

Several case series have reported characteristic CT findings of COVID-19, including ground glass opacities in bilateral peripheral lung, crazy-paving changes, reticular thickening, or consolidations (Fang et al., 2020b; Lei et al., 2020). Given that clinically it is difficult to distinguish COVID-19 from other infective causes for patients with respiratory symptoms, we aim to evaluate whether radiologists or a computer-based image analysis system can reliably differentiate COVID-19 cases from non-COVID-19 but suspected patients.

MATERIALS AND METHODS

Study Population

From 20 December, 2019 to 10 February, 2020, 52 laboratory-confirmed COVID-19 patients in Shenzhen were identified that fulfilled the diagnostic criteria of Chinese national guideline (Lin and Li, 2020). The criteria for a confirmed case include documented laboratory evidence, compatible clinical symptoms, and exposure history. Documented laboratory evidence is defined by positive NAT result either from respiratory tract, bronchoalveolar lavage fluid, or blood sample. Compatible clinical symptoms refer to fever, cough, imaging characteristics of pneumonia, and/or normal or decreased white blood cells count or decreased lymphocyte count. Exposure history includes travel/residence history in Wuhan city, contact history with laboratory-confirmed patients, or contact history with patients with fever or respiratory symptoms from Wuhan and its surrounding areas or endemic communities, within 14 days before the onset of illness. We randomly selected 80 laboratory-confirmed non-COVID-19 viral pneumonia patients as controls. These presented with suspected symptoms and exposure history and underwent NAT and chest CT exams in the same period. To be eligible for inclusion as a control, patients must have at least two negative NAT results from two respiratory specimens and laboratory evidence of other respiratory viruses including positive antigen or NAT test results for influenza A, influenza B, parainfluenza, respiratory syntactical virus, or adenovirus.

Chest CT Evaluation by Radiologists

All CT images were independently retrospectively analyzed using a structured form by two experienced radiologists in a double blinded fashion without knowing the clinical diagnosis. A third senior radiologist was consulted to solve any discrepancy by consensus. Evaluation was focused on the presence of ground glass opacities, patchy infiltration, patchy consolidation, pleural effusion, mediastinal lymphadenopathy, air bronchogram, pleural thickening, and interstitial change.

Machine-Learning Based CT Lesion Detection and Quantification System

We used the uAI image analysis system (Lianying Intelligent Medical Technology Co., Ltd., Shanghai, China) for detection and quantification of chest CT lesions (Computer-based detection in **Supplementary Appendix**). The system classifies the lung fields into five lung lobes and 18 lung segments based on Boyden classification, detects infected regions in each anatomical region, and quantifies the cumulative and relative infected volume. The infected volumes between the two groups were also compared in four different CT windows. Different CT windows refer to different status of brightness and contrast of a CT image manipulated via the CT numbers in order to highlight particular structures. This study was approved by the Institutional Review Board of the Second Affiliated Hospital of Shenzhen University.

Statistical Analysis

Categorical variables were expressed as number and proportion and compared with a Chi-square test. Continuous variables were presented with mean ± standard deviation for data with a normal distribution and tested by Student *t*-test. Data with non-normal distribution were presented with median with interquartile range and compared with independent sample Mann-Whitney *U* test. All tests in this study were two-sided, and $P < 0.05$

was deemed statistically significant. Data was analyzed using SPSS 23.0 software.

RESULTS

Study Patients

Of the 52 included case patients, 32 (61.54%) were male with a mean age of 45.61 ± 14.19 years. The duration of the disease course ranged from 1 to 10 days, with a mean of 5.61 ± 2.19 days. Chest CT examination was performed within 10 days of disease onset. Among patients with COVID-19, fever, fatigue, and dry cough were the major presenting symptoms; hypertension, chronic heart failure, and diabetes were the leading comorbidities, and lymphopenia with elevated CRP or lactate levels were the main laboratory findings. There was no significant difference between case and control in demographic, disease course, clinical symptoms, and laboratory findings (**Table 1**).

TABLE 1 | Patients characteristics between case and control patients [*n*(%)].

Characteristic	COVID-19 (*N* = 52)	Non-COVID-19 viral pneumonia (*N* = 80)	*P*-value
Age	45.61 ± 14.19	48.69 ± 14.62	0.12
Sex			
Female	32 (61.54%)	58 (72.50%)	0.44
Male	20 (38.46%)	22 (27.50%)	0.73
Course of disease	5.61 ± 2.19	8.32 ± 1.82	0.61
Onset of symptom to CT examination	4.32 ± 0.82	6.87 ± 1.89	0.06
Symptoms			
Fever	47 (90.38)	71 (88.75)	0.24
Fatigue	44 (84.61)	68 (85.00)	0.08
Dry cough	31 (59.62)	48 (60.00)	0.16
Dyspnea	15 (28.85)	27 (33.75)	0.67
Sore throat	28 (53.85)	49 (61.25)	0.48
Comorbidities			
Hypertension	24 (46.15)	39 (48.75)	0.89
Chronic heart failure	10 (19.23)	17 (21.25)	0.85
Diabetes	14 (26.92)	20 (25.00)	0.52
Cerebrovascular disease	9 (17.31)	19 (23.75)	0.37
COPD	11 (21.15)	16 (20.00)	0.94
Chronic kidney disease	4 (7.69)	8 (10.00)	0.30
Laboratory results			
White blood cell count (10^9/L)	13.60 ± 5.38	11.35 ± 5.88	0.39
Neutrophil percentage (%)	20.43 ± 5.86	18.20 ± 5.05	0.59
Lymphocyte percentage (%)	82.45 ± 14.54	85.69 ± 10.08	0.20
C-reactive protein (mg/dL)	93.20 ± 24.01	106.59 ± 29.29	0.10
Lactate (mmol/L)	2.71 ± 1.38	2.58 ± 1.20	0.16

Comparison of CT Interpretation by Radiologists

In both COVID-19 and non-COVID-19 viral pneumonia patients, ground-glass opacity (**Supplementary Figures 1A,B**) and patchy airspace infiltrates (**Supplementary Figures 2A,B**) were the major findings. The lesions could be found in multiple lobes or segments, more often bilateral. There was no statistical significance between the distribution of lesions identified by radiologists in the two groups. The majority of both categories of patients have ground glass opacities and patchy infiltration, followed by patchy consolidation. There is no statistically significant difference in any of the CT manifestations between the two groups (**Table 2**).

Comparison of Computer System Detected Infected Lung Volume

Compared between COVID-19 and non-COVID-19 viral pneumonia patients, there was no significant difference in the computer system detected infected lung volume/percentage in 5 lung lobes and in different lung segments (**Table 3**). Similarly, there was no significant difference in the absolute or relative infected lung volume between the two groups of patients (**Supplementary Table 1**).

Sensitivity Analysis Under Different Radiodensity Ranges for Machine Detection

The lesions' infection volume/percentage for COVID-19 and non-COVID-19 patients were stratified by four different CT number ranges (**Supplementary Table 2**). Lesions in the (−750, −300) Hounsfield units (HU) range were arbitrarily chosen to reflect ground-glass opacity, while those with CT numbers in the

TABLE 2 | Radiologist interpretation of chest CT before NAT results [*n*(%)].

	COVID-19 (*N* = 52)	Non-COVID-19 viral pneumonia (*N* = 80)	*P*-value
Ground glass opacities	38 (73.07)	57 (71.25)	0.27
Patchy infiltration	30 (57.69)	48 (60.00)	0.08
Patchy consolidation	15 (28.84)	28 (35.00)	0.15
Pleural effusion	4 (7.69)	9 (11.25)	0.54
Mediastinal lymphadenopathy	3 (5.78)	5 (6.25)	0.71
Air bronchogram	11 (21.15)	15 (18.75)	0.19
Pleural thickening	4 (7.69)	6 (7.50)	0.57
Interstitial change	10 (19.23)	16 (20.00)	0.29

TABLE 3 | Difference in lesions distribution between COVID-19 and non-COVID-19 viral pneumonia patients.

Affected Lung field	COVID-19		Non-COVID-19 viral pneumonia		P^a	P^b
	Infection volume (cm^3)	Percentage of infection (%)	Infection volume (cm^3)	Percentage of infection (%)		
Right upper lobe	29.55 ± 10.41	3.93 ± 1.11	40.76 ± 13.24	11.66 ± 3.67	0.70	0.71
Right middle lobe	11.61 ± 6.64	0.63 ± 0.20	25.35 ± 7.51	4.57 ± 1.90	0.22	0.20
Right lower lobe	45.93 ± 18.40	7.19 ± 3.41	80.61 ± 41.47	15.56 ± 7.38	0.37	0.93
Left upper lobe	22.17 ± 8.54	1.19 ± 0.11	34.60 ± 10.41	8.67 ± 2.00	0.90	0.79
Left lower lobe	48.42 ± 22.98	8.67 ± 1.45	105.76 ± 32.03	22.15 ± 8.02	0.52	0.75

[a] *Comparison of the infection volume in COVID-19 group and non-COVID-19 viral pneumonia.*
[b] *Comparison of the infection ratio of the affected sites in COVID-19 group and non-COVID-19 viral pneumonia group.*

(−300, 50) range were considered as denser airspace infiltrates and consolidation. Anything less than −750 becomes harder to differentiate from normal lung tissue. An extreme number 50 was chosen, beyond which infection becomes unlikely, though not impossible, as the number starts to get into the soft tissue mass range. No statistically significant difference in the infected volumes between the two groups was found in any of the four ranges.

DISCUSSION

In our case-control study, we selected controls in the real world settings, used both radiologist and computer system to evaluate the CT image difference between two groups of patients. We confirmed the main CT manifestations of patients with COVID-19 were multi-lobar/segmental peripheral ground-glass opacities and patchy airspace infiltrates. There was no significant difference between the two groups in radiologists' interpretation and the volume of software system detected lesions.

The Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) primarily invades the lung parenchyma (Lin and Li, 2020). Because of the long turnaround time and suboptimal sensitivity of NAT for SARS-CoV-2, whether chest CT can be used as a first line diagnostic rule to differentiate COVID-19 from other viral pneumonia is of critical importance in both clinical and public health perspectives. Chen et al. (2020) studied 29 patients with COVID-19 showed that the chest image lesions were mostly bilateral and multiple with patchy shadows and ground glass opacities (Chan et al., 2020; Fang et al., 2020a). Pan et al., reported CT images of COVID-19 are diverse in the early stages, which may present ground-glass opacities, pulmonary consolidation and nodules (Pan and Guan, 2020). These studies were case series without suitable control groups (Chen et al., 2020; Fang et al., 2020b; Kanne, 2020; Lei et al., 2020; Pan and Guan, 2020). We found Radiologists' interpretation alone or computer-based lesion detection cannot differentiate COVID-19 form other viral pneumonia. Such findings were robust under different anatomic sites or CT density ranges.

There are pros and cons using clinical diagnosis as a case definition. Clinical diagnosis allows early isolation with initial false negative NAT, slows transmission, and implements treatment early. However, the use of chest CT as a first line diagnostic method may miss early/mild disease, promote cross-infection in the CT room, increase radiation exposure, and consume enormous resources of disinfection.

Our findings have multiple implications. First, clinicians should not rely on initial chest CT findings to diagnose COVID-19 in the absence of laboratory confirmation. Second, patients who had been diagnosed with COVID-19 based on the clinical grounds should be re-evaluated with serum antibody tests. Third, patients who were isolated and cohorted with laboratory-confirmed cases in the temporary COVID-19 hospitals of Hubei province may need to be reevaluated aiming for further laboratory evidence including repeated NAT or serum antibody test.

Results of study should be interpreted in light of its shortfalls. Due to the constraints in time and patient number, we could not perform independent validation with an external sample. Comparison between radiologists and Artificial intelligence interpretations was not made as this was not within our study scope, but could be a point of interest for future study. We did not compare serial imaging changes. We used a commercial image analysis system based on deep learning. The system does not provide flexibility in adjustment of machine-learning model or hyperparameters selection. We do not exclude future tailor trained machine learning systems that can differentiate chest CT of COVID-19 from other suspected patients. Lastly, the sample size we collected could provide a mere statistical power of 70% to differentiate a 20% difference between two proportions. The calculation of the power is based on the Z-test with a statistical significance level of 0.0.5 and a sample size of 130 patients.

CONCLUSION

We documented the non-differentiating nature of initial chest CT between COVID-19 and other viral pneumonia with suspected symptoms. Our results do not support CT findings replacing microbiological diagnosis as a critical criterion for COVID-19 diagnosis.

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 the Institutional Review Board of the Second Affiliated Hospital of Shenzhen University. 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

QD and JL: conceptualization, data curation, and formal analysis. QD, JL, WZ, and YG: methodology. W-TH, HT, and WY: project administration. QD, JL, and C-CL: resources. QD and C-CL: supervision. QD, JL, WZ, YG, and C-CL: writing – original draft. QD, JL, W-TH, K-CH, HT, WY, and C-CL: writing – review and editing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We are indebted to Dr. Anuj Pareek at Department of Radiology, Stanford University for his constructive comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.614207/full#supplementary-material

REFERENCES

Ai, J.-W., Zhang, H.-C., Xu, T., Wu, J., Zhu, M., Yu, Y., et al. (2020). Optimizing diagnostic strategy for novel coronavirus pneumonia, a multi-center study in Eastern China. *medRxiv* [Preprint].

Bogoch, I. I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M. U., and Khan, K. (2020). Potential for global spread of a novel coronavirus from China. *J. Travel Med.* 27:taaa011.

Chan, J. F.-W., Yuan, S., Kok, K.-H., Wang To, K. K., Chu, H., Yang, J., et al. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 395, 514–523. doi: 10.1016/s0140-6736(20)30154-9

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 395, 507–513. doi: 10.1016/s0140-6736(20)30211-7

Coronavirus COVID-19 Global Cases by John Hopkins (2020). *The Center for Systems Science and Engineering at Johns Hopkins University, 2020*. Available online at: https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 (accessed February 16, 2020).

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* 20, 533–534. doi: 10.1016/s1473-3099(20)30120-1

Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., et al. (2020a). Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology* 296, E115–E117.

Fang, Y., Zhang, H., Xu, Y., Xie, J., Pang, P., and Ji, W. (2020b). CT manifestations of two cases of 2019 novel coronavirus (2019-nCoV) pneumonia. *Radiology* 295, 208–209. doi: 10.1148/radiol.2020200280

Kanne, J. P. (2020). Chest CT findings in 2019 novel coronavirus (2019-nCoV) infections from Wuhan, China: key points for the radiologist. *Radiology* 295, 16–17. doi: 10.1148/radiol.2020200241

Lei, J., Li, J., Li, X., and Qi, X. C. T. (2020). Imaging of the 2019 novel coronavirus (2019-nCoV) pneumonia. *Radiology* 295:18. doi: 10.1148/radiol.2020200236

Lin, L., and Li, T. (2020). Interpretation of" guidelines for the diagnosis and treatment of novel coronavirus (2019-nCoV) infection by the National Health Commission (Trial version 5)". *Zhonghua Yi Xue Za Zhi* 100:E001. rrr

Pan, Y., and Guan, H. (2020). Imaging changes in patients with 2019-nCov. *Eur. Radiol.* 30, 3612–3613. doi: 10.1007/s00330-020-06713-z

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 Physiology

METHODS
published: 27 April 2021
doi: 10.3389/fphys.2021.650714

Quantifying Vascular Density in Tissue Engineered Constructs Using Machine Learning

Hannah A. Strobel[1†], Alex Schultz[2†], Sarah M. Moss[1], Rob Eli[2] and James B. Hoying[1]*

[1] Tissue Modeling, Advanced Solutions Life Sciences, Manchester, NH, United States, [2] Innovations Laboratory, Advanced Solutions Life Sciences, Louisville, KY, United States

Given the considerable research efforts in understanding and manipulating the vasculature in tissue health and function, making effective measurements of vascular density is critical for a variety of biomedical applications. However, because the vasculature is a heterogeneous collection of vessel segments, arranged in a complex three-dimensional architecture, which is dynamic in form and function, it is difficult to effectively measure. Here, we developed a semi-automated method that leverages machine learning to identify and quantify vascular metrics in an angiogenesis model imaged with different modalities. This software, BioSegment, is designed to make high throughput vascular density measurements of fluorescent or phase contrast images. Furthermore, the rapidity of assessments makes it an ideal tool for incorporation in tissue manufacturing workflows, where engineered tissue constructs may require frequent monitoring, to ensure that vascular growth benchmarks are met.

Keywords: angiogenesis, machine learning, vascularity, neovessel growth, vessel quantification

OPEN ACCESS

Edited by:
Min Xu,
Carnegie Mellon University,
United States

Reviewed by:
Akanksha Bhargava,
Johns Hopkins Medicine,
United States
Jun Kong,
Georgia State University,
United States

****Correspondence:***
James B. Hoying
jhoying@advancedsolutions.com

† These authors have contributed equally to this work

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

Received: *07 January 2021*
Accepted: *06 April 2021*
Published: *27 April 2021*

Citation:
Strobel HA, Schultz A, Moss SM, Eli R and Hoying JB (2021) Quantifying Vascular Density in Tissue Engineered Constructs Using Machine Learning. Front. Physiol. 12:650714. doi: 10.3389/fphys.2021.650714

INTRODUCTION

A mature vascular network is essential for the viability of tissues, native or fabricated. Thus, making informative, quantitative measurements of a vascular network is important. This is especially relevant for a maturing, vascularized engineered tissue, which needs to meet certain vascular density benchmarks to remain viable. Making accurate vessel density measurements, particularly during angiogenesis, can be challenging due to the irregular features of both the individual vessels and the complex networks they form, and the surrounding tissue environments. Vasculatures exist in three dimensional environments with vessels extending across all dimensions, may vary in density, and can resemble other non-vascular tissue elements (e.g., ducts, cell bundles, etc.). A variety of imaging modalities are employed to visualize the vasculature in both the laboratory and the clinic (Spaide et al., 2015; Grüneboom et al., 2019; Bautista et al., 2020). Usually, a contrast agent that fills the blood space and/or labels vessel cells directly is involved, as inherent contrast between the vessel and the surrounding tissue is often low. Segmentation and quantification from these images is subsequently performed to assess the vasculature. While effective at visualizing vessel elements, non-uniform labeling of vessels by these agents can confound segmentation and feature detection. All of this complicates computer-based image analyses. Additionally, many labeling methods rely on contrast agents flowing through the vasculature and are, therefore, not useable for neovascular systems in which intravascular perfusion is not yet established.

Other pixel-based computational systems for quantifying do exist, designed to analyze images of vasculatures stained with fluorescent labels, but are often intended for very specific applications. The ImageJ plugin Angiogenesis Analyzer, for example, has been widely published for use in 2D endothelial cell assays (Yamamoto et al., 2014; Yang et al., 2016) and a handful of 3D tissues where the vasculature was very distinct and uniform (Samarelli et al., 2014). Rarely, however, in tissues, fabricated or native, is the vasculature so clear. Often clumps of cells or tissue, or irregularly shaped immature vessels, present challenges for computational methods of identifying blood vessels, even with highly specific stains.

Artificial intelligence and machine learning (AI/ML) are increasingly employed to make measurements in biomedical research that can be challenging for more traditional computational systems biological systems. With AI/ML, a software program can be "trained" to distinguish certain features. A handful of AI/ML programs exist that have been used to quantify the vasculature, although none have been designed specifically for high throughput analyses, especially involving phase contrast imaging. Additionally, all of these programs have only been tested on clearly defined, mature vasculatures, that lack the feature noise visible in growing and remodeling neovasculatures in a tissue space. For example, the program VesSAP has been used to map the vasculature in a whole mouse brain following the use of a perfused tag. While the program was able to accurately map the vasculature and produce an impressive amount of data, extensive clearing and staining protocols were needed to obtain clean, high resolution 3D images, which took an additional 24 h to segment (Todorov et al., 2020). While this may be ideal for some applications, it may not be as useful for high throughput analysis or as a routine-use tool in the laboratory. VesSAP also requires a background in computer software to operate. Another program, REAVER, has been used in similar applications, but also requires a MATLAB license and an understanding of software coding in order to use (Corliss et al., 2020). The open source software Ilastik has been used to quantify vascular density using ML (Bochner et al., 2020). However, it also can be challenging to use by the non-expert and requires more computing power to run than is available to a typical biomedical laboratory.

Perhaps because existing programs can be challenging to use, and are designed for very specific applications, the most common approach for measuring vessel density is still having an expert user manually trace vessels in each individual image (Holley et al., 2010; Top et al., 2011; Da et al., 2015; Cossutta et al., 2019). While manual annotations are an easy way to accurately quantify vessel density, it can be very tedious and time consuming, and is not a practical way to analyze large amounts of data.

Here, we leveraged modern ML to develop an easy-to-use analysis tool specifically designed to make measurements of vascularity in *in vitro* tissue environments. The focus of this application is assessing angiogenesis (new vessel growth) in which vessel morphology and network topology is highly variable. We focused on reporting vessel length density measurements from our experiments, although it may be possible for the tool to be trained to identify other features and corresponding metrics, as well. Ultimately, BioSegment will be incorporated into high-throughput tissue manufacturing workflows to monitor vascular growth within fabricated, tissue engineered products. Here, we demonstrate the application of this tool, the BioSegment software, in assessments of vascularity from confocal fluorescence and phase contrast images.

METHODS

Microvessel Culture

Whole, intact, microvessel fragments were isolated from adipose tissue, from either discarded human lipoaspirates or epididymal fat from male retired breeder Sprague Dawley rats (all animal procedures were approved by the Dartmouth College IACUC) and assembled into angiogenesis assays as previously described (Nunes et al., 2013; Strobel et al., 2020). Rat vessels were cultured at 60 k/ml in DMEM (Gibco) containing 20% fetal bovine serum (FBS; Thermo Fisher), 1% penicillin-streptomycin (Fisher), and 1% amphotericin B (Fisher). Human microvessels were cultured at 100 k/ml in RPMI (Corning) containing B27 (Gibco) and 50 ng/ml vascular-endothelial growth factor (VEGF; Peprotech). Assessments were made of four different treatments (Groups 1–4) promoting differing levels of angiogenesis. Experimental setups are reported in detail in Strobel et al. (2020). Briefly, for fluorescent 4× images, Group 1 contained microvessels with no additional stimuli, Group 2 was treated with vascular endothelial growth factor (VEGF), stromal cells were incorporated in Groups 3 and 4, and Group 4 was treated with a VEGF trap [R&D systems, described in Strobel et al. (2020)]. For 10× images, Group 1 contained microvessels alone, Group 2 contained stromal cells in a separate region of collagen surrounding the microvessels, Group 3 contained stromal cells mixed in with the microvessels, and Group 4 contained stromal cells both around and within microvessel containing regions [described in detail in Strobel et al. (2020)]. Phase contrast images were experiments comparing different microvessel donors or different culture medium types. The variation in treatment groups was intended to demonstrate the utility of the BioSegment software. The data presented and discussed in this manuscript pertains to the accuracy of BioSegment measurements, not the scientific findings of these experiments, which are already published (Strobel et al., 2020).

Lectin Staining and Imaging

Constructs were fixed overnight in 10% neutral buffered formalin. Rat microvessels were further processed by staining with a fluorescently labeled lectin. After rinsing in phosphate buffered saline (PBS), constructs were permeabilized for 20 min in 0.25% Triton X-100 and blocked for 4 h at RT in 5% bovine serum albumin. They were incubated overnight at 4°C in lectin stain at a dilution of 1:50 in blocking solution [Griffonia (Bandeiraea) Simplicifolia Lectin I (GSL I, BSL I), Vector Laboratories]. Constructs were rinsed multiple times, with one overnight wash, before imaging. Human constructs were imaged using a phase contrast filter on a standard benchtop upright microscope (Olympus). Lectin-stained rat constructs

were imaged with a confocal Olympus FV3000 or an INCell 6500 scanner (Cytiva, formerly GE Life Sciences), depending on the dataset. Confocal images were processed to create maximum z projections and saved as .png files before analysis.

Application of the BioSegment Software

BioSegment utilizes YOLOv4, a machine learning deep convolutional neural network (CNN) to detect and localize user-defined classes within images (Girshick et al., 2014). It's end-user facing front-end is desktop application written in C# and is used for image processing and annotation. Annotations and images are stored in the AWS cloud. Training is performed using AWS SageMaker. Trained models are retrieved from AWS cloud storage and transferred to the local storage for desktop inferencing via subordinate python process.

Prior to training, confocal images are transformed into a maximum projection if necessary and saved as a .png file using an ImageJ macro. After import into BioSegment, images can be further pre-processed through a histogram equalization to intensify features (**Supplementary Figure 1**). This can make the geometry easier for the algorithm to identify if the image has low contrast, and is done by selecting a "pre-process" option prior to training. Phase contrast images were converted to .png files and pre-processed with the same pre-processing function.

During implementation, a neural network was "trained" via expert annotation using polygonal chains (polylines) identifying user-defined features or classes of raw images within the BioSegment software environment (**Figure 1**). Polylines are generated by manually stepping over each vessel, and these lines are in turn bounded by a series of overlapping rectangles. These rectangles are the input regions for the model. Because a large number of rectangles are generated, YOLOv4 gains a large number (typically hundreds) of input regions from each image. Beyond this we utilized a set of pre-trained weights, thereby leveraging transfer learning to allow for generalization from a smaller data set. For the purposes of this application the YOLOv4-tiny.weights set of weights (trained on the MS-COCO dataset) were utilized.

When training is performed on a new dataset, training data is divided or "partitioned" into 3 sets (or groups): training, testing, and holdout. The "training partition" is the training group set for a specific fold. The "holdout" refers to the partition that was entirely segregated from the training process, to be used for later validation. After images were annotated ("train" data), the model was trained and then used to perform inference on new data ("test" data). With the BioSegment approach, images were annotated via the BioSegment interface by manually tracing each vessel to provide training sets for the machine learning engine, which then generated vessel measurements. For phase contrast images, the user also annotated "out of plane" vessels, which were too blurry for the user to tell if they were vessels or not, and "debris" (undigested pieces of tissue or other objects that are not microvessels). For confocal lectin-stained images, these extra parameters were not necessary, as confocal does not pick up out of plane objects, and the lectin stain will not label most debris.

When a trained model is used for object detection, YOLOv4 produces output detection regions ("raw ML detections") which are processed into polylines through a BioSegment specific process (**Figure 2**). First, the object detection regions are connected using a neighbor detection algorithm. Then, a contour detection algorithm is used to eliminate "false" connections that are suggested by proximity, but do not represent actual vessels (**Figure 3A**). Finally, a minimum spanning tree algorithm removes cycles that are generated through the neighbor detection region and allowed by contour detection but not representative of vessel structures (**Figure 3B**). The most up to date version of BioSegment includes a feature that enables users to correct false segmentations, although this was not available at the time the present data was analyzed.

The BioSegment models were compared to two existing measurement protocols. First, we compared BioSegment data to manual tracings of the vessels, an approach widely considered the gold standard. Other programs for vascular quantification do exist, but these are designed for very specific applications. Our target users, who are scientists looking for quick assessments of vascular density in *in vitro* tissues, are still using manual annotations, by an overwhelming majority. Thus, manual annotations were our primary validation method. In addition, we compared the BioSegment measurements to our in-house pixel-based protocol in which images are processed to improve contrast, thresholded, filtered to remove small cells and large clumps, and skeletonized to measure vessel lengths (Strobel et al., 2020). The data reported are from either combined train and test datasets, or test-only datasets.

Total vessel length was calculated for each image and used to calculate vessel length density. Measurements from 4 images per sample were averaged to obtain the length density for each sample. Then, samples within each of the four experimental groups were averaged to determine the average vessel length density within each treatment group. These average vessel length densities, calculated using both measurement methods, was used to calculate percent error for each experimental group. Percent accuracy was calculated by subtracting percent error (below) from 100.

$$Percent\ Error\ = \frac{|Manually\ annotated\ length - BioSegment\ calculated\ length|}{Manually\ annotated\ lengths} * 100$$

Accuracy is reported for each group of each experiment, as different groups sometimes had different densities and morphologies, which affect overall accuracy. An accuracy of "N/A" is assigned to any group where the percent error was greater than 100.

Machine Learning Methodology and Validation

The primary metric of concern related to vessel growth is vessel length density. Vessel lengths are measured manually by estimating the linear extents of visible vessels within a projected image (that is, one that is composed my merging images across a range of focal lengths), and then divided by image area to calculate density.

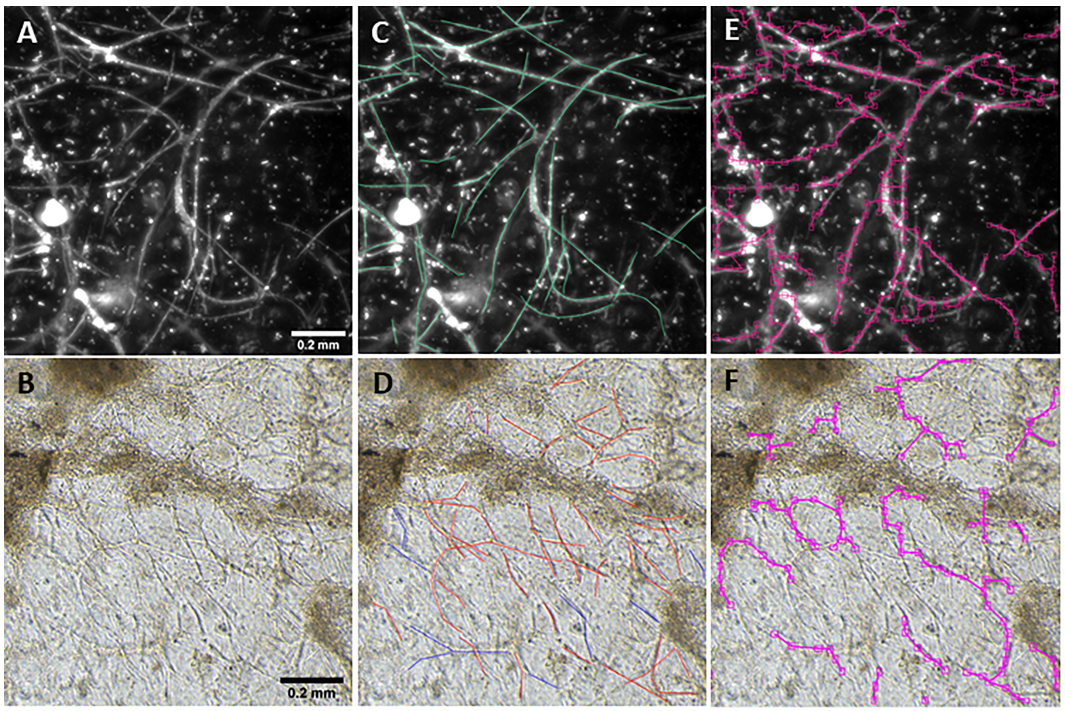


FIGURE 1 | Vessel length density measurement. Fluorescent **(A)** and phase contrast **(B)** images taken at a 10× magnification. Annotations are manually added to images to mark vessels **(C,D)**. After training, the BioSegment software automatically identifies vessels and calculates vessel pixel length **(E,F)**, which is used to calculate vascular length density. Scale = 0.2 mm.

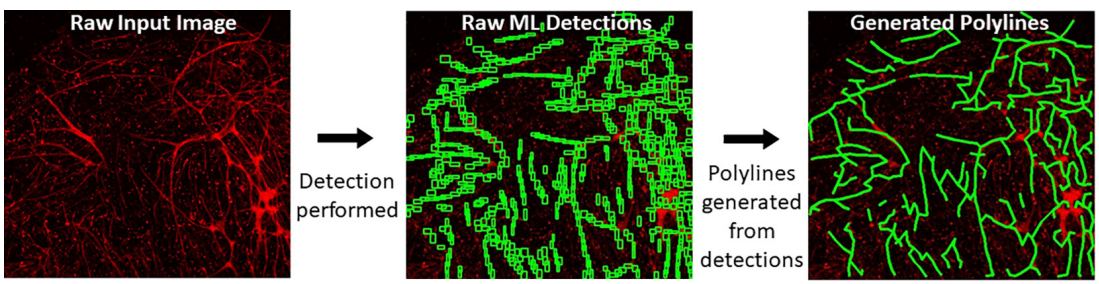


FIGURE 2 | Polyline generation workflow. Raw images are subject to ML detections by YOLOv4. Then, BioSegment generates polylines from these detections.

The approach presented here performs vessel length measurements through an application of object detection. The annotation process involves tracing the linear extents of vessels using polylines (joined line segments). Input object regions are programmatically constructed along the polylines and then used to train an YOLOv4 object detection model[1]. The input and output of object detection models such as YOLOv4 are rectangular bounding regions within a 2D image, which contain an object of interest along with bordering pixels. In most YOLOv4 applications, each input image within a training set will typically contain only 1 or a few examples of an object or objects which are the target of training, thus many images are required to produce enough input regions. In our case, there were often many microvessels in a given image. Thus, relatively small numbers of images (50–100 images) were used in a given training set.

The underlying model is an object-detection model and, as such, explanation algorithms such as Lime were not of value in performing feature importance/interpretability studies. Visual inspection was used to validate that both raw detections and constructed detection polylines represented vessels sufficiently well to produce valid linear extent measurements. This inspection showed that variance that is introduced by the process used to connect detection regions, that is, detection polylines have "kinks" or branches that would not be present in manual annotations, but such deviations tend to offset one another, and thus have a minimal impact on the final length measures.

[1]https://github.com/AlexeyAB/darknet

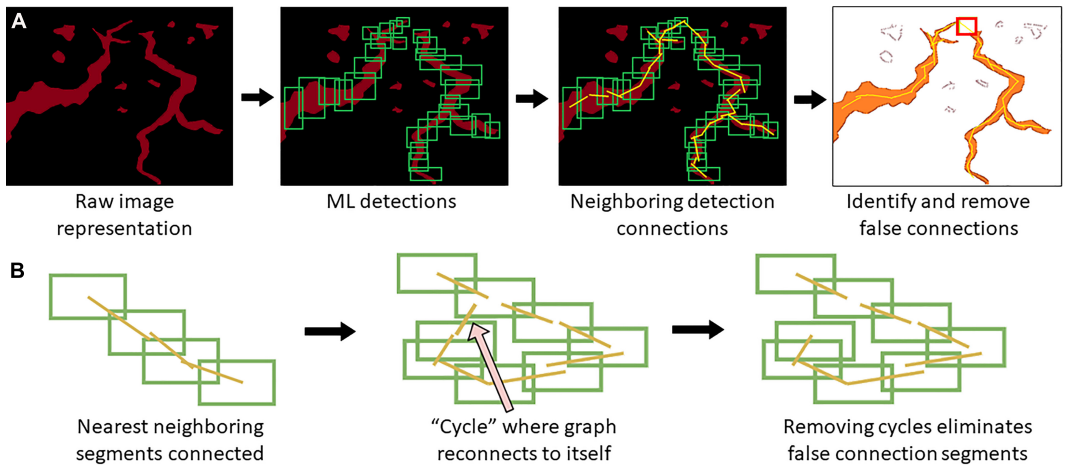


FIGURE 3 | Polyline segmentation and cycle removal. A neighbor detection algorithm is used to create polylines from ML detections **(A)**. This can result in "false" connections between vessel segments, which are then removed using a spanning tree algorithm **(B)**.

A five-fold cross-validation was performed to evaluate the consistency of the vessel measurement algorithm. Two datasets were constructed using lectin-stained images at different magnification levels (referred to as "lectin 4×" and "lectin 10×"). The lectin 4× dataset was comprised of 108 total images from 35 samples. Samples were included from 4 experimental groups. The lectin 10× dataset was comprised of 24 samples, with 2–10 images per sample and 172 total images. Images within each dataset and within each image groups were assigned to training, validation, and holdout partitions. Training partitions were used to train each fold. Validation partitions were utilized by the training process to report training metrics. Image groups assigned to holdout partitions were entirely excluded from the data passed to the model during the training process. Error for a partition overall was calculated by averaging the percent error, as calculated above, for the partition. Overall, data was validated using 2 experiments for the lectin 10× model, 3 for the lectin 4× model, and 3 for phase contrast.

Overall, the cross-validation shows that the vessel-measurement algorithm sufficiently generalized across each fold such that any set of trained weights can be used reliably (**Tables 1**, **2**). Error can be minimized by performing a cross-validation for a specific data set and selecting the best resultant weights for detection.

TABLE 1 | Cross validation results for the model trained to analyze 4× lectin images.

Fold	Partition	% Error
Lectin		
1	holdout	0.0845
1	validation	0.5605
1	train	2.7385
2	holdout	5.8329
2	validation	2.1858
2	train	4.619
3	holdout	12.8576
3	validation	7.2046
3	train	2.8441
4	holdout	0.8912
4	validation	3.6903
4	train	0.4749
5	holdout	1.5407
5	validation	5.5789
5	train	0.4428

RESULTS

BioSegment Accurately Measured Multiple Types of Images

Here, we created a program that leverages artificial intelligence and machine learning to identify tissue components, specifically microvessels. The interface was, subjectively, easy to use, and enabled users to upload and annotate images needed for training. **Figure 1** shows images of fluorescent (A) and phase contrast (B) images prior to processing. After manual annotation (C, D), the model is trained and then "learns" to identify vessels (E, F). Representative datasets show vessel length density as calculated by manual annotation, our previously published pixel-based method, and by BioSegment (**Figure 4**). The percent accuracy was calculated for each group within each dataset, as BioSegment compared to either manual annotations or the pixel-based method (**Figure 4D**). The fluorescent 10× images had the highest accuracies, with percentages ranging from 82.79 to 98.74% accurate. Most accuracies were above 90%. On the same images, the pixel-based measurements had an accuracy of 0–76%, though all but one point was above 44%. With lower magnification images (4×), BioSegment had accuracies of 53.4–99.74% (**Figure 4D**). In these images, subjectively, the lower accuracies occur in images with the highest vessel densities. For the 4× images, pixel-based accuracies were comparable to

TABLE 2 | Cross validation results for the model trained to analyze 10× lectin images.

Fold	Partition	% Error
Lectin 10×		
1	holdout	19.351
1	validation	11.3737
1	train	16.1774
2	holdout	11.0625
2	validation	15.2432
2	train	15.6101
3	holdout	13.4747
3	validation	14.6666
3	train	16.2374
4	holdout	14.8235
4	validation	11.2876
4	train	16.1631
5	holdout	23.9069
5	validation	16.7967
5	train	12.9337

BioSegment, ranging from 62.2 to 96.95% (**Figure 4D**). For phase contrast images, which were taken at a 10× magnification, accuracy ranged from 56.4 to 98.48% (**Figure 4D**).

Accuracy of BioSegment Improved With Time

Due to differences in vascular morphology from experiment to experiment, sometimes a model trained on a dataset from a single experiment failed to yield satisfactory results when tested on a different experiment. In these cases, another training set, from an additional experiment, was added, to add more variety in terms of vascular morphology to the model. After multiple training sets, the model demonstrated increased accuracy across subsequent datasets from multiple experiments, regardless of vascular morphology. Accuracies improved after 2 training sets for the 4× images (81 images total, **Figure 5A**). Here, black bars are annotated controls, gray bars are the BioSegment output after one round of training (Bio 1), and white bars are after 2 rounds of training (Bio 2). In groups 2, 3, and 4, one round of training resulted in inaccurate measurements compared to annotated controls. When a second training dataset was added, accuracy was improved and white BioSegment bars are similar in magnitude to black annotation bars. The exception to this is group one, where both rounds had different magnitudes compared to controls. For phase contrast images, 3 different training datasets were needed (155 images total; **Figure 5B**). Here, the software was highly accurate after a third round of training across all groups (dark gray bars compared to black bars). The 10× fluorescent images had a high accuracy after one large training set (79 images; **Figure 5C**). Statistics were performed using SigmaPlot 11.0 (Systat). A two-way ANOVA test was performed where applicable with Holm-Sidak *post hoc* analysis. Bars are mean ± standard deviation. A significant level of $\alpha = 0.05$ was used for all comparisons to determine statistical significance.

Use of BioSegment Saves Considerable Amounts of Time

It took an experienced scientist 7.26, 4.33, and 2.78 min to annotate a single 4× fluorescence, 10× fluorescence, and 10× phase contrast images (averaged over 3 images of varying density), respectively. The pixel-based method takes an experienced user approximately 80 s per image. Meanwhile, the BioSegment analyzed 45 images in 52 s, for an average of 1.15 s per image. For a single dataset of 45 images, the BioSegment therefore saves from 124 to 325 min of time (2–5.4 h), compared to manual annotations.

"Training and Test" Combined Data Are Comparable to "Test Only" Data

In all datasets used to train the BioSegment models, some images were used for training, while some were used to "test" the model. Whenever possible, the data is reported using "test" images only (**Figure 4C**, and 2 additional experiments plotted in **Figure 4D**). However, in some cases, data is presented as a combination of the training and test images in a dataset, to enable comparison of multiple groups within an experiment, as there were not quite enough "test" samples for meaningful comparisons (**Figures 4A,B**, and 3 additional experiments plotted in **Figure 4D**). To test the validity of the combined test and training data, the accuracy of combined data (compared to manual annotations) was compared to the "test only" images pulled from the same combined dataset. In all cases, average accuracy measurements of test only vs. combined test and train were within 2 percentage points (**Figure 6**). Thus, we can conclude that including training images in our datasets did not skew the calculated accuracy of the results.

DISCUSSION

A variety of applications benefit from or require quantitative assessments of vascularity. Whether in an experimental model of angiogenesis or vascularizing a tissue construct or organoid, knowing the extent and character of the vascular dynamics can be critical to successful outcomes. Furthermore, as tissue fabrication solutions continue to evolve matched by concomitant development of tissue manufacturing processes, real-time assessments of vascularity (perfused or otherwise) are critical. Yet, most methods of quantifying vascular growth are challenging to use without prior experience in computer programming, not conducive to high-throughput applications, or cannot accurately detect complex neovasculatures. Here, we describe a novel machine learning-based software program for semi-automatic quantification of microvessel length density that is simple to use yet leverages the power of machine learning. While directed at assessing vascularity, reflecting our research focus, BioSegment could be adapted to identify other cellular and tissue features as well. In addition to facilitating research investigations, we envision integrating BioSegment as part of a quality assurance program in a tissue manufacturing workflow.

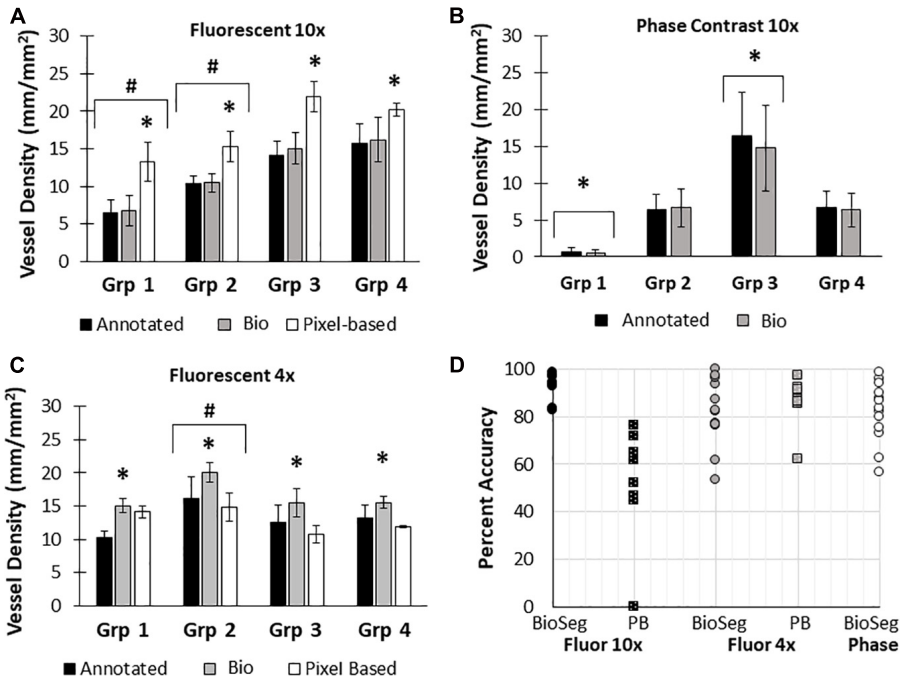


FIGURE 4 | Accuracy of BioSegment vessel measurements. BioSegment calculated vessel length densities (Bio) were compared to manual annotations (Annotated) and pixel-based (PB) measurements. Representative datasets are shown for 10× fluorescent confocal projections **(A)**, phase contrast images **(B)**, and 4× confocal stacks **(C)**. The percent accuracy of the BioSegment (circles) and PB (squares) compared to manual annotations is shown in **(D)**, across all groups from all experiments analyzed. Bars are mean ± SD. A Two-Way ANOVA with Holm Sidak *post hoc* analysis was performed on **(A–C)**. $^{*}P < 0.05$ with compared to all other quantification methods (Annotated, Bio 1, PB) within that group; $^{\#}P < 0.05$ compared to all other groups, regardless of quantification method.

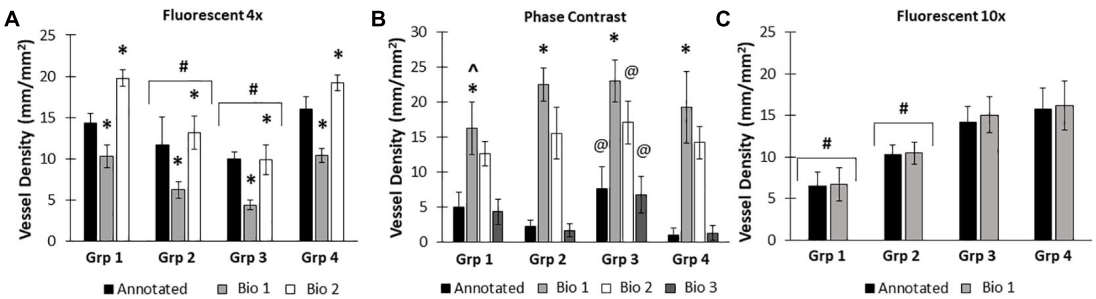


FIGURE 5 | Increasing the number of training datasets improves model accuracy. Representative datasets for 4× fluorescent images **(A)** and phase contrast images **(B)** after each iteration of the respective model. Annotated images are compared to multiple versions of the BioSegment model (Bio 1, Bio 2, Bio 3). Later versions, which contain more training sets, are more accurate. In (A), additional training data increased overall accuracy from 58.5 to 82.13%. In **(B)**, Bio 3 (model version 3) improved to 83% accurate, compared to 0% in Bio 1 and Bio 2 (model versions 1 and 2). Bars are mean ± SD. A Two-Way ANOVA with Holm Sidak post hoc analysis was performed. $^{*}P < 0.05$ compared to all other quantification methods (Annotated, Bio 1, Bio 2, Bio 3) within that group; $^{\#}P < 0.05$ compared to all other groups, regardless of quantification method. $^{@}P < 0.05$ compared to groups 2 and 4, within that quantification method. $^{\wedge}P < 0.05$ compared to other methods within that group.

Here, we compared measurements of vessel length density made by BioSegment to manual annotations made "by hand," and a pixel-based computational method previously developed by our lab. The BioSegment-based measurements of vessel length density from 10× fluorescent confocal images were overall highly accurate compared to manual assessments, both in overall vessel length density magnitude and the trends measured between groups (**Figure 4**). On the same images, the pixel-based approach was considerably less accurate (**Figure 4**). Pixel-based measurements tended to overestimate vessel density. This is because it cannot distinguish between microvessels and other objects with similar dimensions to microvessels, such as elongated endothelial cells and some pieces of undigested tissue. Both 4× fluorescent confocal images and 10× phase contrast images produced a broader range of accuracies across the different treatment groups, although most were above 80% accurate (**Figure 4D**). With 4× fluorescent images, the accuracy of both the pixel-based and BioSegment approaches both had

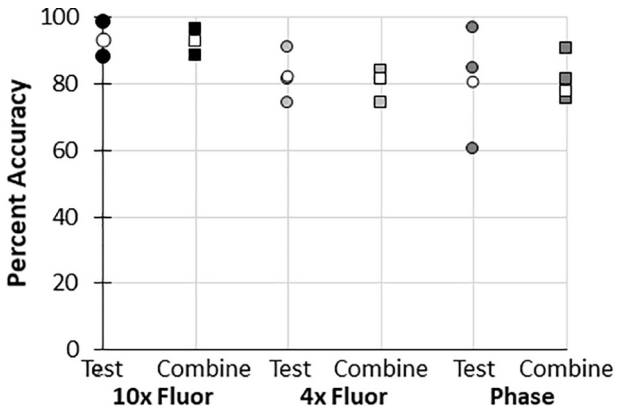


FIGURE 6 | Accuracy of "test and train" combined data compared to "test only" data. Percent accuracy of the BioSegment compared to manual annotations is shown for image sets containing both test and train images (combined; squares), compared to test only images (test; circles) from the same dataset. Individual experiments are shaded, while the average of all experiments is shown in white.

slightly lower reported accuracies (**Figure 4D**). The pixel-based approach was not able to derive vascular length density measurements from phase images due to the lower contrast. In all cases, data produced by the BioSegment within each experiment followed the same trends as annotated controls. For example, in **Figure 4C** Group 2 was highest regardless of quantification method. Thus, even though there is a small degree of error in the magnitude, the user is able to draw the same conclusions. Additionally, while this 20% error seems high, it is comparable to what others have observed in cases where machine learning was applied to biological systems (Liang et al., 2017; Tourlomousis et al., 2019). The observed error may improve with additional training, as is true of all machine-learning based applications. With our data, we observed that more training datasets improved performance (**Figure 5**). It should also be noted that, images contained a substantial amount of noise in the form of non-vessel elements, such as elongated endothelial cells and clumps of undigested tissue. Additionally, there were slight differences in vascular morphology from experiment to experiment. The reported accuracies reflect the ability of BioSegment to distinguish microvessels from this other noise, which further supports the robustness of the program. Cleaner images would likely produce better results.

It was not surprising that the higher magnification images yielded a high percent accuracy, as individual vessel features are more discreet, thereby making it easier both for the user to annotate and BioSegment to segment. For the low magnification confocal images, and particularly those with dense vasculatures, identifying individual vessel segments by both the user and the BioSegment model is relatively challenging. This is largely due to non-distinct boundaries between two or more vessel segments and the intrinsic variation in fluorescence intensities of different vessels, all of which are highly overlapping in the projections. In the future, magnifying images for annotating, in the absence of taking higher magnification images, may improve user annotations and thus BioSegment training. Additionally, extending the analysis to work with 3D image sets, as opposed to the 2D projections used in this study, might facilitate identification of discreet features. Future iterations of BioSegment will include the ability to measure vascular density and other parameters, such as branch points, directly from a 3D confocal image stack. This would enable measurements in Z, as well as X and Y. However, the ability to assess 2D projections makes the program more amenable to rapid, high-throughput screening. From past experience, microvessels in our models grow largely in the X and Y plane, rather than in the Z direction (due to collagen fibril orientation). Thus, the amount of data lost by eliminating growth in Z is minimal and the ability to analyze 3D confocal stacks is not necessary for our current application. Accuracies were determined by comparing BioSegment and pixel-based measurements to those values obtained by manual annotation, which we consider the "gold standard." However, measurements made by manual annotations, even by the most experienced user, can be inaccurate for a variety of reasons (Rudyanto et al., 2014), including subconscious bias, mood, etc. As mentioned above, overlapping vessels, and varying fluorescent intensities, can make it challenging to accurately annotate. Phase contrast images are even more challenging to annotate, as contrast between feature edges and background is often low creating ambiguity in object identification. If the software is trained with inaccurately annotated training sets, it may also have reduced accuracy. One potential way to overcome this challenge is to include training sets from multiple users, which may negate potential user bias and differences in mood/attentive states. Providing sufficient training to the BioSegment model in this way could potentially make it more accurate than a single human user. Furthermore, the ability to identify features from phase contrast images, which can be taken of live cultures quickly and easily without the risks associated with staining, presents an enormous advantage, even if there is some risk of quantitative error.

The amount of time saved by using the BioSegment was considerable, with potential to save users hours of time per experiment. This effect is magnified if users have multiple datasets. The rapid measurements possible with the BioSegment platform have implications for use in high throughput and high content screens and cell/tissue manufacturing efforts. Here, there may be hundreds of samples in culture that require imaging at frequent time points to assess tissue quality and maturation. This could generate thousands of images, which will be impossible to quantify without automated analysis.

Overall, we have demonstrated an innovative software system that utilizes machine learning to quantify microvessel length. It can rapidly and accurately measure vascular features from both fluorescent and phase contrast images. Such a tool saves users an enormous amount of time and has potential to be incorporated into automated processes such as assay screens or quality control in tissue manufacturing. While we focused on training the program to identify microvessels, it could potentially be used to identify any tissue element, including but not limited to cellular density, subcellular features, or contaminants. This could be done following the same procedures for identifying microvessels. The ability to make varied, accurate, and rapid

measurements via a customizable and flexible package may prove invaluable for automation of tissue fabrication, quality control, and real-time monitoring in an automated workflow.

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 animal study was reviewed and approved by the IACUC Committee at Dartmouth College.

AUTHOR CONTRIBUTIONS

AS wrote and edited the coding. HS collected the data, validated the software, and drafted the manuscript. JH and RE supervised code development and experimental design. All authors contributed to the initial conception, development of BioSegment, and edited the manuscript.

FUNDING

We acknowledge our funding sources, DOD grant W911NF-17-3-003/T0055 and NIH R01HL131856.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphys.2021.650714/full#supplementary-material

Supplementary Figure 1 | Histogram equalization. Example images before **(A)** and after **(B)** histogram equalization. The original histogram **(C)** has clip points identified using a defined threshold **(D)** to create a rescaled distribution **(E)**. Blue represents original histogram, and orange represents the new histogram.

REFERENCES

Bautista, M., Saleem, N., and Anderson, I. A. (2020). Current and novel non-invasive imaging modalities in vascular neurosurgical practice. *Br. J. Hosp. Med. (Lond.)* 81, 1–10. doi: 10.12968/hmed.2020.0550

Bochner, F., Mohan, V., Zinger, A., Golani, O., Schroeder, A., Sagi, I., et al. (2020). Intravital imaging of vascular anomalies and extracellular matrix remodeling in orthotopic pancreatic tumors. *Int. J. Cancer* 146, 2209–2217. doi: 10.1002/ijc.32759

Corliss, B. A., Doty, R. W., Mathews, C., Yates, P. A., Zhang, T., and Peirce, S. M. (2020). REAVER: a program for improved analysis of high-resolution vascular network images. *Microcirculation* 27:e12618.

Cossutta, M., Darche, M., Carpentier, G., Houppe, C., Ponzo, M., Raineri, F., et al. (2019). Weibel-palade bodies orchestrate pericytes during angiogenesis. *Arterioscler. Thromb. Vasc. Biol.* 39, 1843–1858. doi: 10.1161/atvbaha.119.313021

Da, W., Zhu, J., Wang, L., and Sun, Q. (2015). Curcumin suppresses lymphatic vessel density in an in vivo human gastric cancer model. *Tumour Biol.* 36, 5215–5223. doi: 10.1007/s13277-015-3178-8

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Columbus, OH, 580–587.

Grüneboom, A., Kling, L., Christiansen, S., Mill, L., Maier, A., Engelke, K., et al. (2019). Next-generation imaging of the skeletal system and its blood supply. *Nat. Rev. Rheumatol.* 15, 533–549. doi: 10.1038/s41584-019-0274-y

Holley, J. E., Newcombe, J., Whatmore, J. L., and Gutowski, N. J. (2010). Increased blood vessel density and endothelial cell proliferation in multiple sclerosis cerebral white matter. *Neurosci. Lett.* 470, 65–70. doi: 10.1016/j.neulet.2009.12.059

Liang, L., Liu, M., and Sun, W. (2017). A deep learning approach to estimate chemically-treated collagenous tissue nonlinear anisotropic stress-strain responses from microscopy images. *Acta Biomater.* 63, 227–235. doi: 10.1016/j.actbio.2017.09.025

Nunes, S. S., Maijub, J. G., Krishnan, L., Ramakrishnan, V. M., Clayton, L. R., Williams, S. K., et al. (2013). Generation of a functional liver tissue mimic using adipose stromal vascular fraction cell-derived vasculatures. *Sci. Rep.* 3:2141.

Rudyanto, R. D., Kerkstra, S., van Rikxoort, E. M., Fetita, C., Brillet, P. Y., Lefevre, C., et al. (2014). Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: the VESSEL12 study. *Med. Image Anal.* 18, 1217–1232.

Samarelli, A. V., Riccitelli, E., Bizzozero, L., Silveira, T. N., Seano, G., Pergolizzi, M., et al. (2014). Neuroligin 1 induces blood vessel maturation by cooperating with the alpha6 integrin. *J. Biol. Chem.* 289, 19466–19476. doi: 10.1074/jbc.m113.530972

Spaide, R. F., Klancnik, J. M. Jr., and Cooney, M. J. (2015). Retinal vascular layers imaged by fluorescein angiography and optical coherence tomography angiography. *JAMA Ophthalmol.* 133, 45–50. doi: 10.1001/jamaophthalmol.2014.3616

Strobel, H. A., LaBelle, S. A., Krishnan, L., Dale, J., Rauff, A., Poulson, A. M., et al. (2020). Stromal cells promote neovascular invasion across tissue interfaces. *Front. Physiol.* 11:1026. doi: 10.3389/fphys.2020.01026

Todorov, M. I., Paetzold, J. C., Schoppe, O., Tetteh, G., Shit, S., Efremov, V., et al. (2020). Machine learning analysis of whole mouse brain vasculature. *Nat. Methods* 17, 442–449. doi: 10.1038/s41592-020-0792-1

Top, A. P., Ince, C., de Meij, N., van Dijk, M., and Tibboel, D. (2011). Persistent low microcirculatory vessel density in nonsurvivors of sepsis in pediatric intensive care. *Crit. Care Med.* 39, 8–13. doi: 10.1097/ccm.0b013e3181fb7994

Tourlomousis, F., Jia, C., Karydis, T., Mershin, A., Wang, H., Kalyon, D. M., et al. (2019). Machine learning metrology of cell confinement in melt electrowritten three-dimensional biomaterial substrates. *Microsyst. Nanoeng.* 5:15.

Yamamoto, Y., Matsui, J., Matsushima, T., Obaishi, H., Miyazaki, K., Nakamura, K., et al. (2014). Lenvatinib, an angiogenesis inhibitor targeting VEGFR/FGFR, shows broad antitumor activity in human tumor xenograft models associated with microvessel density and pericyte coverage. *Vascular Cell* 6, 2–13.

Yang, Y., Zhang, J., Xia, T., Li, G., Tian, T., Wang, M., et al. (2016). MicroRNA-210 promotes cancer angiogenesis by targeting fibroblast growth factor receptor-like 1 in hepatocellular carcinoma. *Oncol. Rep.* 36, 2553–2562. doi: 10.3892/or.2016.5129

Conflict of Interest: JH and RE are partners of Advanced Solutions. All authors are employees of Advanced Solutions.

frontiers in Physiology

ORIGINAL RESEARCH
published: 22 June 2021
doi: 10.3389/fphys.2021.655556

An Interpretable Computer-Aided Diagnosis Method for Periodontitis From Panoramic Radiographs

Haoyang Li[1,2,3], Juexiao Zhou[2,4], Yi Zhou[5], Qiang Chen[6], Yangyang She[7], Feng Gao[8], Ying Xu[1,5]*, Jieyu Chen[7]* and Xin Gao[2]**

[1] Cancer Systems Biology Center, The China-Japan Union Hospital, Jilin University, Changchun, China, [2] Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, [3] College of Computer Science and Technology, Jilin University, Changchun, China, [4] Department of Biology, Southern University of Science and Technology, Shenzhen, China, [5] Department of Biochemistry and Molecular Biology and Institute of Bioinformatics, University of Georgia, Athens, GA, United States, [6] The Affiliated Stomatological Hospital of Soochow University, Soochow, China, [7] Department of Stomatology, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [8] Department of Colorectal Surgery, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

OPEN ACCESS

Edited by:
Yi Jiang,
Georgia State University,
United States

Reviewed by:
Jun Kong,
Georgia State University,
United States
Hiroshi Fujita,
Gifu University, Japan

***Correspondence:**
Feng Gao
gaof57@mail.sysu.edu.cn
Ying Xu
xyn@uga.edu
Jieyu Chen
chenjy335@mail.sysu.edu.cn
Xin Gao
xin.gao@kaust.edu.sa

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

Received: *19 January 2021*
Accepted: *31 May 2021*
Published: *22 June 2021*

Citation:
Li H, Zhou J, Zhou Y, Chen Q, She Y, Gao F, Xu Y, Chen J and Gao X (2021) An Interpretable Computer-Aided Diagnosis Method for Periodontitis From Panoramic Radiographs. Front. Physiol. 12:655556. doi: 10.3389/fphys.2021.655556

Periodontitis is a prevalent and irreversible chronic inflammatory disease both in developed and developing countries, and affects about 20–50% of the global population. The tool for automatically diagnosing periodontitis is highly demanded to screen at-risk people for periodontitis and its early detection could prevent the onset of tooth loss, especially in local communities and health care settings with limited dental professionals. In the medical field, doctors need to understand and trust the decisions made by computational models and developing interpretable models is crucial for disease diagnosis. Based on these considerations, we propose an interpretable method called Deetal-Perio to predict the severity degree of periodontitis in dental panoramic radiographs. In our method, alveolar bone loss (ABL), the clinical hallmark for periodontitis diagnosis, could be interpreted as the key feature. To calculate ABL, we also propose a method for teeth numbering and segmentation. First, Deetal-Perio segments and indexes the individual tooth via Mask R-CNN combined with a novel calibration method. Next, Deetal-Perio segments the contour of the alveolar bone and calculates a ratio for individual tooth to represent ABL. Finally, Deetal-Perio predicts the severity degree of periodontitis given the ratios of all the teeth. The Macro F1-score and accuracy of the periodontitis prediction task in our method reach 0.894 and 0.896, respectively, on *Suzhou* data set, and 0.820 and 0.824, respectively on *Zhongshan* data set. The entire architecture could not only outperform state-of-the-art methods and show robustness on two data sets in both periodontitis prediction, and teeth numbering and segmentation tasks, but also be interpretable for doctors to understand the reason why Deetal-Perio works so well.

Keywords: teeth segmentation and numbering, periodontitis diagnosis, panoramic radiograph, computer-aided diagnostics, interpretable model

INTRODUCTION

Periodontitis is a chronic inflammatory disease of periodontium resulting in inflammation within the supporting tissues of the teeth, progressive attachment, and bone loss (Lindhe et al., 1999). Periodontitis is prevalent in both developed and developing countries, and affects about 20–50% of the global population which makes it a public health concern (Nazir, 2017). Thus, the tool for automatically diagnosing periodontitis is highly demanded to provide the invaluable opportunity to screen at-risk people for periodontitis and its early detection could prevent the onset of tooth loss, especially in local community and health care settings where dentists are not easily accessible (Balaei et al., 2017). The form of periodontitis is characterized by periodontal ligament loss and resorption of the surrounding alveolar bone caused by severe inflammatory events (de Pablo et al., 2009). Cumulative alveolar bone loss (ABL) results in weakening of the supporting structures of the teeth, and predisposes the patient to tooth mobility and loss (Brunton, 2008; **Figure 1B**). Thus ABL is a hallmark of the periodontal disease (Yang et al., 2018). To calculate ABL of all teeth, it is necessary to gather the contours of the individual tooth and the alveolar bone. In this situation, teeth numbering and segmentation are essential and fundamental tasks for periodontitis diagnosis. In addition, dentists usually need to serve numerous patients and read a large number of panoramic radiographs daily. Thus an automatic tool for teeth numbering and segmentation to enhance efficiency and improve the quality of dental care is timely needed (Chen et al., 2019).

Several methods have been proposed to tackle the periodontitis prediction or teeth numbering and segmentation tasks. Joo et al. (2019) proposed a classification method for the periodontal disease based on convolutional neural network (CNN) by using periodontal tissue images. This method classified four states of periodontitis and the accuracy on validation data was 0.81. Ozden et al. (2015) tested three classification algorithms, artificial neural networks (ANN), support vector machine (SVM), and decision tree (DT) to diagnose periodontal diseases by using 11 measured variables of each patient as raw data. The results showed that DT and SVM were best to classify the periodontal diseases with high accuracy (0.98 of precision). It revealed that SVM and DT appeared to be practical as a decision-making aid for the prediction of periodontal disease. Lee et al. (2018) proposed a periodontitis prediction method using periapical radiographic images via deep CNN. The diagnostic accuracy for periodontally compromised teeth was 81.0% for premolars and 76.7% for molars. Li et al. (unpublished) proposed a method which could screen the existence of gingivitis and its irritants, i.e., dental calculus and soft deposits, from oral photos with a novel multi-task learning model. With 625 patients included in this study, the classification area under the curve for detecting gingivitis, dental calculus and soft deposits were 87.11, 80.11, and 78.57%, respectively.

As for the teeth numbering and segmentation, Wirtz et al. (2018) proposed a coupled shape model in conjunction with a neural network by using panoramic radiographs. The network provided a preliminary segmentation of the teeth region which is used to initialize the coupled shape model. Then the 28 individual teeth (excluding wisdom teeth) were segmented and labeled using gradient image features in combination with the model's statistical knowledge. The experimental result showed an average dice of 0.744. Chen et al. (2019) used faster regions with CNN features to detect and number teeth in dental periapical films. They proposed three post-processing techniques to improve the numbering performance. Results revealed that mean average precision (mAP) was 0.80 and the performance of this method was close to the level of junior dentists. Cui et al. (2019) used deep CNN to achieve automatic and accurate tooth instance segmentation and identification from cone-beam CT (CBCT) images. They extracted the edge map from the CBCT image to enhance image contrast along shape boundaries. Next, the edge map and input images were passed through 3D Mask R-CNN with encoded teeth spatial relationships. Their method produced accurate instance segmentation and identification results automatically.

The main limitations of methods mentioned above are as follows: (1) the bias of detecting and numbering teeth in some cases with severe periodontitis due to the disturbance of a large number of missing teeth, (2) the lacking capability of their methods on predicting the number of missing teeth in the shortage data volume of some individual classes, and (3) the lack of interpretability of predicting the severity degree of periodontitis.

In this paper, we try to overcome these limitations through the following contributions. (1) We propose an automatic and interpretable method called Deetal-Perio to predict the severity degree of periodontitis from dental panoramic radiographs. (2) As a subroutine of Deetal-Perio, we further propose a method for teeth numbering and segmentation which consists of a novel calibration algorithm. (3) Deetal-Perio outperforms state-of-the-art methods and shows the robustness on the two data sets from two hospitals. (4) Deetal-Perio uses ABL as the feature for periodontitis diagnosis and is thus fully interpretable.

We note that a shorter conference version of this paper appeared (Li et al., 2020b). This manuscript added more comprehensive details of data annotation, the calibration method for teeth numbering, implementations of experiments and measurements. We also collected more in-house data from the cooperative hospital to enhance the performance of our method.

MATERIALS AND METHODS

Data Sets

The Affiliated Stomatological Hospital of Soochow University supplied a total of 302 digitized panoramic radiographs (hereinafter referred to as the *Suzhou* data set). Each radiograph has a high resolution of 1,480 * 2,776 pixels and was annotated following the Fédération Dentaire Internationale (FDI) numbering system to get the contours of teeth and their labels as the ground truth (GT). FDI numbering system divides all teeth into four quadrants where teeth are labeled as 11–18, 21–28, 31–38, 41–48, respectively (**Figure 1A**). Among all radiographs, 298 were labeled with the severity degree of periodontitis by dentists, including four categories: 52 of no periodontitis, 189

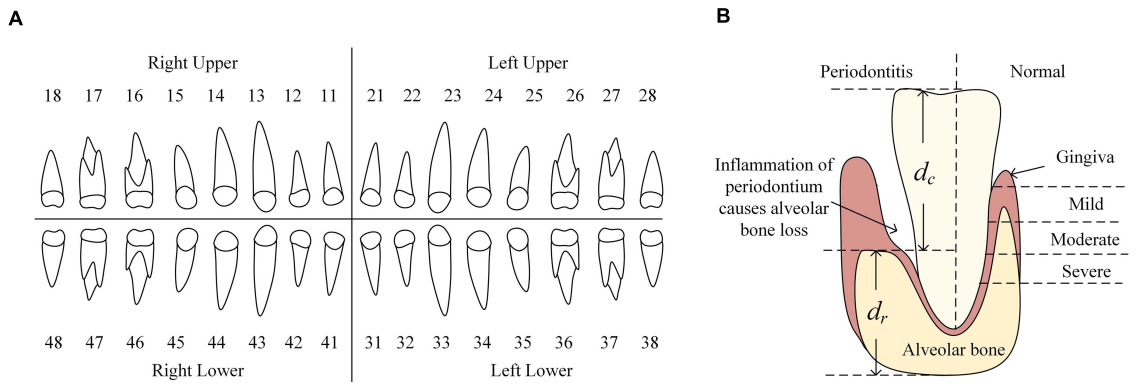


FIGURE 1 | (A) FDI numbering system. It divides all teeth into four quadrants where teeth are labeled as 11–18, 21–28, 31–38, 41–48, respectively. **(B)** The left of tooth shows the appearance of periodontitis and the representation of ABL. d_c represents the ABL and $d_c + d_r$ is used for normalize the ABL. Thus, each tooth has its ABL representation calculated as $d_c/(d_c + d_r)$. The right of tooth shows the appearance of a normal tooth. Different severities of alveolar bone loss reflect the severities of periodontitis (normal, mild, moderate and severe periodontitis).

of mild periodontitis, 43 of moderate periodontitis, and 14 of severe periodontitis. The 4 missing images contain the mislabeled images and unlabeled images.

We also collected another data set from the Sixth Affiliated Hospital, Sun Yat-sen University (hereinafter referred to as the *Zhongshan* data set). This data set includes 204 panoramic radiographs with high and various resolution which are categorized by four classes mentioned above (69 of no periodontitis, 54 of mild periodontitis, 42 of moderate periodontitis, and 39 of severe periodontitis).

These two data sets are both in-house data sets which are created by the machines called SOREDEX DIGORA Optime from KaVo Dental (North Carolina, United States) and GiANO from NewTom (Bologna, Itlay), respectively. Each data set was labeled by one expert dentist with more than 10 years of clinical experience, respectively. The pixel value in the radiograph reflected by the density (the quantity of darkness). The procedure of annotation is as follows. Given a panoramic radiograph, the dentist would draw a contour along with the boundary of each existing tooth and assign a corresponding tooth number to each existing tooth by following the FDI numbering system. Next, the dentist assigned the stage of periodontitis for each radiograph according to the visualizing of radiograph and the recognized standard from previous study (Papapanou et al., 2018). Here we also define a ratio c for each tooth, which is defined as the distance between cemento-enamel junction (CEJ) and alveolar bone divided by the distance between CEJ and the dental root, to indicate the level of bone loss for each tooth. No periodontitis is defined that none of teeth has bone loss. Mild periodontitis is defined that at least the c of one tooth is less than 15%. Moderate periodontitis is defined that at least the c of one tooth is less than 33% and larger than 15%. Severe periodontitis is defined that at least the c of one tooth is larger than 33%.

Methods

The architecture of Deetal-Perio is as follows. First, Deetal-Perio segments and numbers individual tooth via Mask R-CNN combined with a novel calibration method. Next, Deetal-Perio segments the contour of the alveolar bone and calculates the ratio for individual tooth which could represent ABL as the key feature to predict periodontitis. Finally, Deetal-Perio uses XGBoost to predict the severity degree of periodontitis by given a vector of ratios from all the numbered teeth. The entire architecture is shown in **Figure 2**.

Teeth Segmentation and Numbering

Inspired by the state-of-the-art architecture in object classification and segmentation called Mask R-CNN (Abdulla, 2017; He et al., 2017), we tried to segment the teeth with binary classification via Mask R-CNN. That was, we wanted to differentiate teeth from the background image. The result revealed that Mask R-CNN could detect almost all of the teeth given a radiograph. Then, we tried to segment and number the teeth with multi-class classification via Mask R-CNN and this time, we wanted to identify the labels of these teeth. The result showed that only a minority of teeth could be detected and numbered due to the limitation of data from individual classes, but compared with the GT, most detected teeth were numbered correctly. **Figure 3** shows these two results, which provide complementary information to each other. Thus, we proposed to combine the results from the binary and multi-class Mask R-CNN models together. We extracted the bounding boxes and masks from the binary classification results, and the labels of numbered teeth from the multi-class classification results. We further proposed a calibration method to integrate the results from the two classifiers, refine the labels of numbered teeth, and infer the labels of unnumbered teeth.

The calibration method is designed as follows: first, B=$\{B_1, B_2, ..., Bm\}$ and M=$\{M_1, M_2, ..., Mm\}$ represent m and n of center points of teeth' bounding boxes from the results of binary and multi-class classification Mask R-CNN, respectively. Then, we found the closest tooth to each tooth of M in B by calculating the Euclidean distance and assigned the labels of teeth from M to B. Next, we iterated each tooth in B to judge

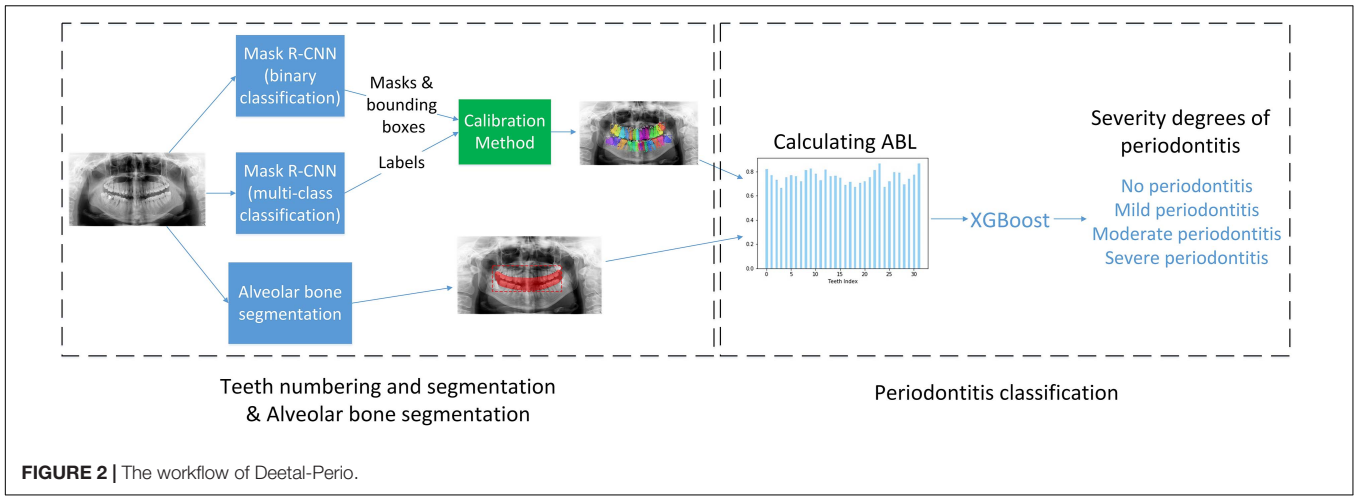


FIGURE 2 | The workflow of Deetal-Perio.

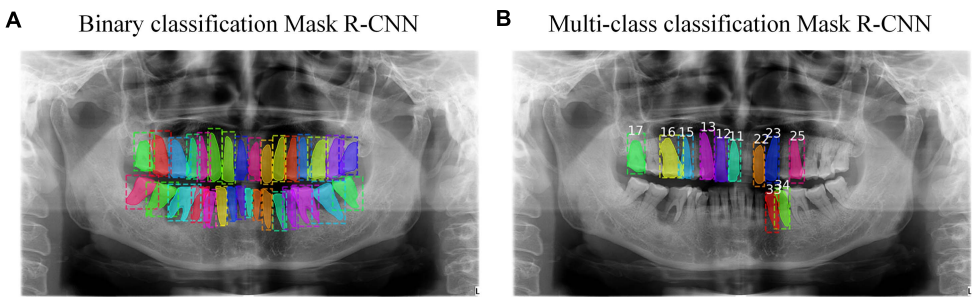


FIGURE 3 | (A) The result of binary classification Mask R-CNN. **(B)** The result of multi-class classification Mask R-CNN. Each color in these two results represents different instances.

whether neighboring teeth are labeled and calibrated its own label until all the teeth in B satisfied three conditions: 1. each tooth had been labeled; 2. no repeatedly labeled tooth; 3. all labeled teeth followed the rules of FDI numbering system. During the iteration of unlabeled teeth, there would be three circumstances. First, if neither of the neighboring teeth was labeled, we would skip. Second, if one of the neighboring teeth was labeled, we would use the ratio of distances between adjacent teeth to inference the label of selected tooth. For example, we selected tooth B_i after several iterations and the right neighboring of B_i, called B_{i+1}, had been numbered. We defined D_r as the horizontal distance from B_i to B_{i+1}, D_{rr} as the horizontal distance from B_{i+1} to the right neighboring tooth of B_{i+1}. Next, we calculated the rounded ratio D_r / D_{rr} which indicates the number of teeth between B_i and B_{i+1}. Then, this rounded ratio could be utilized to inference the label of B_i according to B_{i+1}. Third, when both of neighboring teeth were labeled, it is similar to the previous circumstance that we used the nearer neighbor of selected tooth and its adjacent tooth to inference the label of selected tooth. For example, we also defined the selected tooth as B_i, the nearer neighboring tooth of B_i as B_j, and the neighboring tooth of B_j as B_k (not B_i). Similarly, we utilized the rounded ratio, the distance from B_i to B_j divided by the distance from B_j to B_k, to indicate the number of teeth between B_i and B_j, which could be used to inference the label of B_i. The details of algorithm are given in **Figure 4**. Finally, all teeth are labeled in B which are considered as the results of teeth segmentation and numbering step.

The Representation of ABL

As previously introduced, ABL results in weakening of the supporting structures of the teeth, which makes it the hallmark of periodontal disease (Brunton, 2008; **Figure 1B**) and the annotation of degrees of periodontitis also depends on the severities of ABL from all of teeth. Thus, we were going to use the contours of individual teeth and their labels we have obtained and the contour of the alveolar bone to calculate the severity of ABL for each individual tooth. We used binary classification Mask R-CNN to segment the contour of alveolar bone and the confidence scores of detected boxes are high in the results. Crown-to-Root Ratio (CRR) is a kind of prognostic parameter in periodontology, which could be defined as,CRR=d_c/d_r where d_c denotes the vertical distance from the alveolar bone to the top of the dental crown and denotes the vertical distance from the bottom of the dental root to the alveolar bone (**Figure 1B**). We also defined the severity of ABL as the ratio $d=CRR/(1+CRR)$ for each individual tooth. Due to the smoothness of the contour of the alveolar bone, we randomly selected two points on this contour to draw a line. Then, we randomly chose 50% points on the contour of the dental crown to calculate the vertical distance from these

Algorithm 1 The calibration method for teeth numbering

Input: $B = \{B_1, \ldots, B_m\}$: m of central points of teeth's bounding boxes from binary classification

$M = \{M_1, \ldots, M_n\}$: n of central points of teeth's bonding boxes from multi-class classification

L_i : *the label of tooth* i

X_i : *the* x *coordinate of tooth* i

Output: *calibrated* $\{L_{B1}, \ldots, L_{Bm}\}$

$L_{B_1}, \ldots, L_{B_m}$ *are set to* 0
for $i \leftarrow 1$ *to* n **do**
  $B_k \leftarrow$ *the nearest point to* M_i *by calculating Euclidean distance*
  $L_{B_k} \leftarrow L_{M_i}$
end for
while exists teeth unlabeled or not follow FDI numbering system or has repeat numbered teeth **do**
  for *each unlabeled tooth* B_i **do**
    if $L_{B_{i-1}} = 0$ *and* $L_{B_{i+1}} = 0$ **then**
      continue
    else if $L_{B_{i-1}} \neq 0$ *and* $L_{B_{i+1}} \neq 0$ **then**
      choose one of B_{i-1} *and* B_{i+1} *to find nearer tooth* B_j *to* B_i
      $B_k \leftarrow$ *the neighboring tooth of* B_j *(not* B_i*)*
      $z \leftarrow round(|X_{B_i} - X_{B_j}|/|X_{B_j} - X_{B_k}|)$
      infer L_{Bi} *considering* $z - 1$ *of missing teeth according to* B_j
    else if *only one of* B_{i-1} *and* B_{i+1} *is labeled* **then**
      $B_j \leftarrow$ *labeled tooth between* B_{i-1} *and* B_{i+1}
      $B_k \leftarrow$ *the neighboring tooth of* B_j *(not* B_i*)*
      $z \leftarrow round(|X_{B_i} - X_{B_j}|/|X_{B_j} - X_{B_k}|)$
      infer L_{Bi} *considering* $z - 1$ *of missing teeth according to* B_j
    end if
  end for
end while

FIGURE 4 | The calibration algorithm for teeth numbering.

points to this line, respectively, and defined the largest vertical distance as d_c. We estimated d_r in a similar way. In our case, d_c could be a good estimation to represent ABL showing the level of destruction of the alveolar bone and to normalize d_c, we divided it by $d_c + d_r$ which is the estimation of the perpendicular height of the tooth. Thus, d could represent the ABL of each tooth.

Periodontitis Prediction

After acquiring the ratio d of individual tooth, each radiograph would output a vector of ratios D=$\{d_1, d_2 \ldots, d_{32}\}$ where each radio corresponds to a label of tooth. Apparently, some radiographs do not have all the 32 teeth. In such cases, the ratios of teeth which do not exist in the radiograph are set to be the mean value of its neighboring teeth' ratios. We then solved the periodontitis classification task by XGBoost (Chen and Guestrin, 2016) which has gained attention as the algorithm of choice for many winning teams of machine learning competitions (Volkovs et al., 2017). To tackle the class imbalance problem, we used Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002) for over-sampling the minority classes. After over-sampling, D was inputted as the feature to classify the four-class severity degree of periodontitis by XGBoost.

RESULTS

Experimental Setup

The binary and multi-class classification of Mask R-CNN were trained starting from pre-trained COCO weight and fine-tuned on our panoramic radiographs. We used ResNet-101 as the backbone network to extract features and Mask R-CNN was implemented for 30 epochs with a learning rate of 0.001 to avoid making the learning jump over minima and the batch size was set to 1. We also configured the minimum confidence score of detection as 0.7, because there were several teeth with abnormal shapes and we did not want to eliminate them. The threshold of non-maximum suppression was set to 0.3 to guarantee enough proposals retained and allow the existence of an overlap between teeth. During the data pre-processing, we resized all the panoramic radiographs to 1,024 * 1,024 pixels. We preserved the previous ratio, which means if an image was not square, we would pad it with zeros. All experiments were conducted by 2 NVIDIA QUADRO M6000 24GB GPU. The methods were running backend on TensorFlow version 1.9.0 and operating system was Ubuntu 16.04.

We randomly extracted 80% (i.e., 241) and 80% (i.e., 163) of *Suzhou* and *Zhongshan* data sets, respectively, as two training sets and the rest 20% (i.e., 61) and 20% (i.e., 41) were two testing sets, respectively. The reason why we didn't choose the validation test is to take full advantage of our collected data. We trained Mask R-CNN starting from pre-trained COCO weight on *Suzhou* training set and then we fine-tuned this model on *Zhongshan* training set. Finally, we tested the performance of our model on two testing sets, respectively. All training and testing processes have been repeated three times, and the final results were averaged from these three experiments. For comparison with other methods, we applied the same training and testing procedure on our dataset as we did with our method.

Performance

Figures 5A,B shows eight examples of teeth segmentation and numbering, the two panels include the results of four kinds of periodontitis from the *Suzhou* data set and *Zhongshan* data set, respectively, showing good performance and robustness in both datasets. **Table 1** shows the comparison of our teeth segmentation and numbering method with the baseline Mask R-CNN method and those proposed by Wirtz et al. (2018) and Chen et al. (2019) on our two datasets. The comparison is quantified by three metrics: (1) Dice (all) denotes the overall dice coefficient of teeth segmentation, which represents the capability of our method to

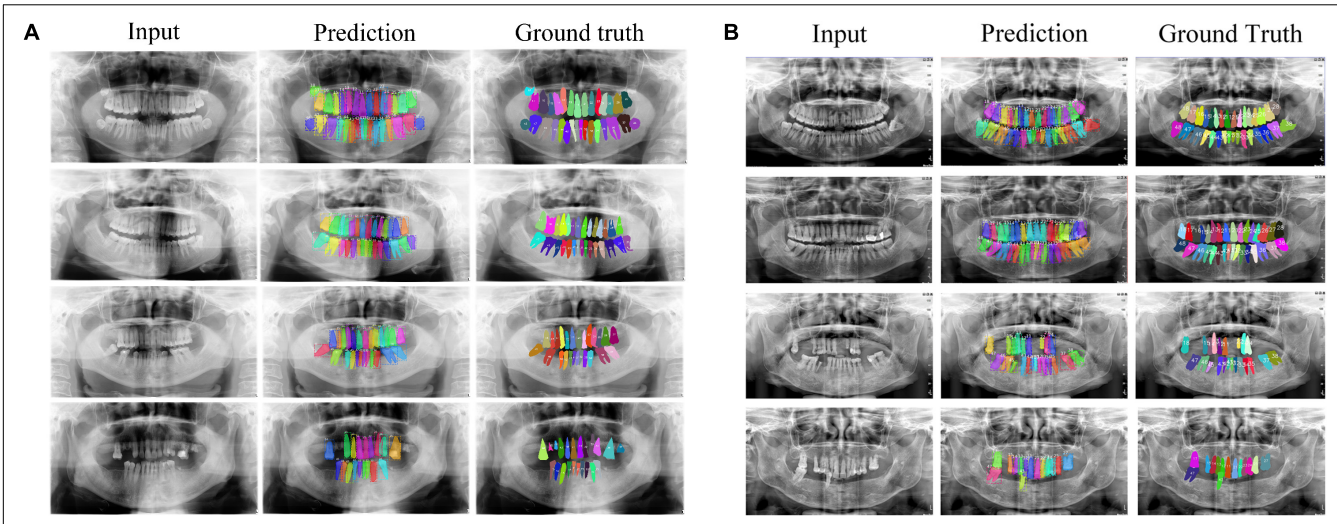


FIGURE 5 | Teeth numbering and segmentation results tested on *Suzhou* and *Zhongshan* data set are shown in subplot **(A,B)**, respectively. From top to bottom cases for each panel, they are no periodontitis, mild periodontitis, moderate periodontitis and severe periodontitis.

TABLE 1 | Performance comparison tested on the *Suzhou* data set and *Zhongshan* data set between Deetal-Perio and other methods on teeth numbering/segmentation task, by mAP, Dice (all) and Dice (single).

	***Suzhou* data set**			***Zhongshan* data set**		
Methods	**mAP**	**Dice (all)**	**Dice (single)**	**mAP**	**Dice (all)**	**Dice (single)**
Deetal-Perio	**0.863**	**0.892**	**0.809**	**0.927**	**0.903**	**0.819**
Multi-class Mask R-CNN	0.834	0.830	0.781	0.881	0.869	0.801
Wirtz et al. (2018)	0.435	0.765	0.502	0.409	0.648	0.428
Chen et al. (2019)	0.680	–	–	0.559	–	–

The method of Chen et al. (2019) only has tooth numbering, but not segmentation function. The best performance in each category is in bold.

TABLE 2 | Performance comparison tested on the *Suzhou* data set and *Zhongshan* data set between Deetal-Perio and five methods by F1-score and accuracy on the periodontitis prediction task.

	***Suzhou* data set**		***Zhongshan* data set**	
Methods	**Macro F1-score**	**Accuracy**	**Macro F1-score**	**Accuracy**
Deetal-Perio	**0.889**	**0.892**	**0.812**	**0.819**
SVM	0.693	0.711	0.449	0.590
Decision tree	0.745	0.758	0.643	0.665
Adaboost	0.701	0.742	0.670	0.688
CNN	0.591	0.611	0.669	0.729
Joo et al. (2019)	0.331	0.408	0.318	0.367

The best performance in each category is in bold.

detect and segment the teeth. We evaluated Dice (all) after the implementation of binary classification Mask R-CNN, thus Dice (all) could be defined as:

$$Dice(all) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

X represents the binary mask of prediction for all the teeth and the Y denotes the binary mask of GT for all the teeth. (2) The mAP denotes the capability of whether our method detects all numbers of teeth correctly, which is calculated as follows. First, we choose the proposal, whose intersection over union (IoU) with GT of a specific class is higher than the pre-defined threshold (normally set to 0.5) and the confidence score is the highest, as the true positive and the other proposals corresponding the same GT would be defined as false positive. Thus, we could calculate the precision, recall, and precision-recall curve for a particular class. The area under the precision-recall curve is the average precision (AP) of this specific class and the mAP could be calculated as the mean value of the APs for all object classes. (3) Dice (single) denotes the mean value of all dice coefficients from all labeled teeth, respectively, which could evaluate the performance of our method to segment each specific tooth. Dice (single) could be defined as:

$$Dice(single) = \sum_{i=0}^{n} Dice(P_i, G_i) \Big/ n \tag{2}$$

where n represents the number of teeth in the GT, P_i and G_i represent the i-th mask of a specific tooth from prediction and GT, respectively. Thanks to the calibration method in Deetal-Perio, we could number teeth much more correctly than other compared methods. Thus, the performance of Deetal-Perio in the segmentation of individual tooth is also better than other methods. In the compared methods, Chen et al. (2019) focus on the teeth numbering task, which output the mAP only. Their method is not trained on panoramic radiographs like ours, but the dental periapical films. Thus, their post-processing method may not work well on our data sets, which causes a worse performance than ours. The method of Wirtz et al. (2018) requires the fixed image size to inference the contour for each tooth which could not work on our data sets, which includes

FIGURE 6 | This case missed the filled tooth between number 17 and 16. Actually, the missing tooth should be number 17 and the tooth which is wrongly numbered as 17 should be 18. Considering about the filled teeth could enhance the performance of teeth numbering and segmentation tasks.

various sizes of panoramic radiographs. Less of robustness of their method for image size causes a worse performance than ours. Comparing the two datasets, we could also find that the results from *Zhongshan* data set are a little better than the results from *Suzhou* data set, and this could be caused by the different quality of the radiographs from the two hospitals. The radiographs from *Zhongshan* data set are more enhanced than *Suzhou*'s, including the darker background and the brighter foreground, which could make it easier for the network to extract dental features.

We then evaluated the performance of periodontitis prediction. **Table 2** compares our method with several baseline machine learning methods and the method proposed by Joo et al. (2019) on our two data sets. The input of Joo et al. (2019) is the resized panoramic radiograph and the other compared methods accept the ABL vector as the input. The metrics include Macro F1 score and accuracy over the four classes, which are defined as following equations:

$$Macro\ F1\ score = \sum_{i=0}^{3} \frac{2 \times (P_i \times R_i)}{P_i + R_i} \Big/ 4 \tag{3}$$

$$Accuracy = \frac{n_{correct}}{n_{total}} \tag{4}$$

where $\mathbf{P}_i$ and $\mathbf{R}_i$ denote the precision and recall of the i-th class, $n_{correct}$ indicates the number of prediction classified correctly, and n_{total} denotes the number of total data. Because the class-imbalanced issue has been solved by SMOTE, the usage of accuracy makes sense to evaluate the performance of our method. Considering the balance between precision and recall of each class, we use Macro F1 score to evaluate our method. The reason why we choose XGBoost as our classifier is that the boosting algorithm of XGBoost made it a strong learner to enhance the performance compared with the simple decision trees, and the regularization of XGBoost made it robust against the noise and thus outperforming Adaboost. The reason why CNN does not have great performance is that CNN may bring the overfitting issue when handling such low dimension input data. Compared to Joo et al. (2019), our method greatly reduces the feature dimension from a radiograph to a 1×32 vector instead of simply implementing a CNN for image classification with a large number of disturbing redundant features. In addition, the intermediate results of teeth segmentation and numbering as well as the geometrically calculated ABL provide dentists completely refined and interpretable information, so that they not only know that our method works, but also understand how it works.

DISCUSSION AND CONCLUSION

In this paper, we proposed a fully automatic and completely interpretable method, Deetal-Perio, for diagnosing the severity degrees of periodontitis from panoramic radiographs using ABL as the key feature. As the intermediate results, our method also accomplished teeth numbering and segmentation tasks. Comprehensive experiments on two data sets show that Deetal-Perio not only dramatically outperforms other compared methods in both teeth segmentation and numbering, and periodontitis prediction, but is also robust and generalizable on independent data sets, which makes Deetal-Perio a suitable method for periodontitis screening and diagnostics.

Despite the success of Deetal-Perio, the performance of teeth numbering relies on the numbering results from the multi-class classification Mask R-CNN model in Deetal-Perio. This can cause issues when there are radiographs with severe periodontitis which have only few or abnormal shapes of teeth. To overcome this limitation, more data, similar to these special cases, is needed to further improve the performance, employing few-shot learning might be a helpful way to deal with such special situations (Li et al., 2019, 2020a). We also found that there are filler materials in some teeth which could not be detected (**Figure 6**). Although these are a minority of cases, we need to focus on them in order to enhance the performance of teeth segmentation and numbering results in our model. We also plan to develop some functions, such as detecting the filled teeth automatically based on our model, which could assist the dentists during diagnosis at a higher level.

From the clinical perspective, the comprehensive process of periodontitis diagnosis should be also based on the measures like clinical attachment loss (CAL) which is costly for us to collect. For better performance and applicability of our method, we plan to collect this kind of data and integrate them with our method to make it more convincing for clinicians.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

HL implemented the pipeline, did the experiments, evaluated the performance, and wrote the manuscript. JZ designed the calibration method and did that part of experiments. YZ preprocessed raw data. QC, YS, JC, and FG supplied the in-house data sets. XG and YX supervised the study and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The research reported in this study was supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under award nos. BAS/1/1624-01, FCC/1/1976-23-01, FCC/1/1976-26-01, REI/1/0018-01-01, REI/1/4216-01-01, REI/1/4437-01-01, REI/1/4473-01-01, and URF/1/4098-01-01, the Fundamental Research Funds for the Central Universities (No. 20ykpy05 to FG), and the Sun Yat-sen University 100 Top Talent Scholars Program—China (No. P20190217202203617 to FG).

ACKNOWLEDGMENTS

We thank He Zhang, Yi Zhang, Yongwei Tan, Nana Liu, and Ying Zhao at The Affiliated Stomatological Hospital of Soochow University for providing the data. The content of this manuscript has been presented in part at the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2020) (Li et al., 2020b).

REFERENCES

Abdulla, W. (2017). *Mask R-CNNfor Object Detection and Instance Segmentation on Keras and TensorFlow, GitHub Repos*. Available online at: https://github.com/matterport/Mask_RCNN (accessed November 1, 2019).

Balaei, A. T., de Chazal, P., Eberhard, J., Domnisch, H., Spahr, A., and Ruiz, K. (2017). "Automatic detection of periodontitis using intra-oral images," in *Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* Jeju, 3906–3909. doi: 10.1109/EMBC.2017.8037710

Brunton, P. A. (2008). Prevention in clinical oral health care. *Br. Dent. J.* 204, 217.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357.

Chen, H., Zhang, K., Lyu, P., Li, H., Zhang, L., Wu, J., et al. (2019). A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Sci. Rep*. 9:3840. doi: 10.1038/s41598-019-40414-y

Chen, T., and Guestrin, C. (2016). "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, 785–794.

Cui, Z., Li, C., and Wang, W. (2019). "ToothNet: automatic tooth instance segmentation and identification from cone beam CT images," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, 6368–6377.

de Pablo, P., Chapple, I. L. C., Buckley, C. D., and Dietrich, T. (2009). Periodontitis in systemic rheumatic diseases. *Nat. Rev. Rheumatol.* 5, 218–224. doi: 10.1038/nrrheum.2009.28

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, 2961–2969.

Joo, J., Jeong, S., Jin, H., Lee, U., Yoon, J. Y., and Kim, S. C. (2019). "Periodontal disease detection using convolutional neural networks," in *Proceedings of the 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Okinawa, 360–362. doi: 10.1109/ICAIIC.2019.8669021

Lee, J. H., Kim, D. H., Jeong, S. N., and Choi, S. H. (2018). Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J. Periodontal Implant Sci.* 48, 114–123. doi: 10.5051/jpis.2018.48.2.114

Li, H., Tian, S., Li, Y., Fang, Q., Tan, R., Pan, Y., et al. (2020a). Modern deep learning in bioinformatics. *J. Mol. Cell Biol.* 12, 823–827.

Li, H., Zhou, J., Zhou, Y., Chen, J., Gao, F., Xu, Y., et al. (2020b). "Automatic and interpretable model for periodontitis diagnosis in panoramic radiographs," in *Proceedings of the Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, eds A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, et al. (Cham: Springer International Publishing), 454–463.

Li, Y., Huang, C., Ding, L., Li, Z., Pan, Y., and Gao, X. (2019). Deep learning in bioinformatics: introduction, application, and perspective in the big data era. *Methods* 166, 4–21.

Lindhe, J., Ranney, R., Lamster, I., Charles, A., Chung, C. P., Flemmig, T., et al. (1999). Consensus report: chronic periodontitis. *Ann. Periodontol.* 4:38.

Nazir, M. A. (2017). Prevalence of periodontal disease, its association with systemic diseases and prevention. *Int. J. Health Sci.* 11, 72–80.

Ozden, F. O., Özgönenel, O., Özden, B., and Aydogdu, A. (2015). Diagnosis of periodontal diseases using different classification algorithms: a preliminary study. *Niger. J. Clin. Pract.* 18, 416–421. doi: 10.4103/1119-3077.151785

Papapanou, P. N., Sanz, M., Buduneli, N., Dietrich, T., Feres, M., Fine, D. H., et al. (2018). Periodontitis: consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *J. Clin. Periodontol.* 45, S162–S170. doi: 10.1111/jcpe.12946

Volkovs, M., Yu, G. W., and Poutanen, T. (2017). "Content-based neighbor models for cold start in recommender systems," in *Proceedings of the Recommender Systems Challenge 2017*, Como, 1–6.

Wirtz, A., Mirashi, S. G., and Wesarg, S. (2018). "Automatic teeth segmentation in panoramic X-ray images using a coupled shape model in combination with a neural network," in *Proceedings of the Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, eds A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger (Cham: Springer International Publishing), 712–719.

Yang, M., Nam, G. E., Salamati, A., Baldwin, M., Deng, M., and Liu, Z. J. (2018). Alveolar bone loss and mineralization in the pig with experimental periodontal disease. *Heliyon* 4:e00589. doi: 10.1016/j.heliyon.2018.e00589

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 Molecular Biosciences

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fmolb.2021.614277

Multi-Task Classification and Segmentation for Explicable Capsule Endoscopy Diagnostics

Zishang Kong[1†], Min He[2†], Qianjiang Luo[2†], Xiansong Huang[3], Pengxu Wei[3,4], Yalu Cheng[1], Luyang Chen[5], Yongsheng Liang[6], Yanchang Lu[7], Xi Li[2] and Jie Chen[1,3]**

[1]School of Electonic and Computer Engineering, Peking University, Shenzhen, China, [2]Department of Gastroenterology, Peking University Shenzhen Hospital, Shenzhen, China, [3]Peng Cheng Laboratory, Shenzhen, China, [4]Sun Yat-sen University, Guangzhou, China, [5]Pennsylvania State University, Philadelphia, PA, United States, [6]Harbin Institute of Technology (Shenzhen), Shenzhen, China, [7]Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China

OPEN ACCESS

Edited by:
Lihua Li,
Hangzhou Dianzi University, China

Reviewed by:
Mohammad Shahid,
Children's National Hospital, United States
Long Xu,
Shenzhen University General Hospital, China
Sara Monteiro,
Centro Hospitalar de Trás os Montes e Alto Douro, Portugal
Qing Liu,
Central South University, China

***Correspondence:**
Jie Chen
chenj@pcl.ac.cn
Xi Li
1191@pkuszh.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Diagnostics and Therapeutics, a section of the journal Frontiers in Molecular Biosciences

Received: *05 October 2020*
Accepted: *23 July 2021*
Published: *19 August 2021*

Citation:
Kong Z, He M, Luo Q, Huang X, Wei P, Cheng Y, Chen L, Liang Y, Lu Y, Li X and Chen J (2021) Multi-Task Classification and Segmentation for Explicable Capsule Endoscopy Diagnostics. Front. Mol. Biosci. 8:614277. doi: 10.3389/fmolb.2021.614277

Capsule endoscopy is a leading diagnostic tool for small bowel lesions which faces certain challenges such as time-consuming interpretation and harsh optical environment inside the small intestine. Specialists unavoidably waste lots of time on searching for a high clearness degree image for accurate diagnostics. However, current clearness degree classification methods are based on either traditional attributes or an unexplainable deep neural network. In this paper, we propose a multi-task framework, called the multi-task classification and segmentation network (MTCSN), to achieve joint learning of clearness degree (CD) and tissue semantic segmentation (TSS) for the first time. In the MTCSN, the CD helps to generate better refined TSS, while TSS provides an explicable semantic map to better classify the CD. In addition, we present a new benchmark, named the Capsule-Endoscopy Crohn's Disease dataset, which introduces the challenges faced in the real world including motion blur, excreta occlusion, reflection, and various complex alimentary scenes that are widely acknowledged in endoscopy examination. Extensive experiments and ablation studies report the significant performance gains of the MTCSN over state-of-the-art methods.

Keywords: Capsule endoscopy, Multi-task learning, Explicable, Crohn's disease, Auxiliary diagnosis

1 INTRODUCTION

Deep learning and convolutional neural networks have recently shown outstanding performances for visual recognition and semantic understanding [Krizhevsky et al. (2012); Simonyan and Zisserman (2014); He et al. (2016); Huang et al. (2017); Long et al. (2015)]. The representation learning capacity of convolutional neural networks has also been successfully applied to medical image analysis and recognition in gastrointestinal endoscopy [Ronneberger et al. (2015); Le et al. (2019); Hwang et al. (2020)]. Crohn's disease [Podolsky (1991); Baumgart and Sandborn (2012)] is an inflammatory bowel disease (IBD), and its signs and symptoms range from mild to severe. It usually develops gradually but sometimes will come on suddenly, without warning. While there is not a known cure for Crohn's disease, early detection and preventative therapies will greatly reduce its signs and symptoms and even bring about long-term remission. Because the small intestine and colon can be affected by Crohn's disease, capsule endoscopy is the gold standard to examine the midsection of the gastrointestinal tract.

A major challenge in capsule gastroscopy is that the procedure will output a video of several hours which suffers from complicated gastrointestinal environmental challenges, such as excreta occlusion,

motion blur, and light scattering, wasting plenty of time for professional gastroenterologists to find out the location of lesions [Min et al. (2019)]. Although several software enhancements, including Quick-View (Medtronic, Minneapolis, MN, United States) and Express View (CapsoVision, Inc., Saratoga, CA, United States), attempt to overcome these drawbacks, their performance is insufficient for use in clinical practice because of their limited accuracy and unexplicable output [Hwang et al. (2020)]. To assist gastroenterologists to locate Crohn's lesions explicably and precisely, we introduce a dataset named the Capsule-Endoscopy Crohn's Disease dataset, a large-scale Crohn's gastrointestinal image dataset for clearness degree (CD) and tissue semantic segmentation (TSS) which will greatly help doctors understand the classification results. The proposed dataset covers 467 images in real-world scenarios.

In the meanwhile, we propose a multi-task learning (MTL) scheme, which combines pixel-level segmentation and global image-level category classification. The proposed architecture is based on a fully convolutional image-to-image translation scheme, which enables efficient feature sharing between image regions, and fast prediction. A novel cross fusion module is proposed to mitigate the gap between different foci of classification and segmentation tasks. We evaluate our model on the proposed dataset, with clearness degree classification and tissue segmentation with eight classes. We show that through joint training, the model is able to learn shared representations that are beneficial for both tasks. Our method can be seen as a generalization of approaches relying on detection annotations to pre-train the deep model for classification purposes. We show that our joint training of classification and segmentation enables a better cooperation between tasks.

2 RELATED WORK

2.1 Image Classification

Since AlexNet [Krizhevsky et al. (2012)], deep convolutional neural networks have dominated image classification. With this trend, research has shifted from engineering handcrafted features to engineering network architectures. VGG-Net [Simonyan and Zisserman (2014)] proposes a modular network design strategy, stacking the same type of network blocks repeatedly, which simplifies the workflow of network design and transfer learning for downstream applications. Built on the success of this pioneering work, He et al. (2016) introduced an identity skip connection which alleviates the difficulty of vanishing gradient in the deep neural network and allows for network learning deeper feature representations. Reformulations of the connections between network layers [Huang et al. (2017)] have been shown by DenseNet to further improve the learning and representational properties of deep networks. DenseNet has become one of the most successful CNN architectures which has been adopted in various computer vision applications.

2.2 Semantic Segmentation

With the great success of deep learning in high-level vision tasks, numerous semantic segmentation approaches [Long et al. (2015); Ronneberger et al. (2015); Zhao et al. (2017); Chen et al. (2018)] are beneficial for CNNs. Long et al. (2015) first introduced fully convolutional networks (FCNs) for semantic segmentation which conduct pixel-wise classification in an end-to-end fashion. While U-Net was introduced by Ronneberger et al. (2015), which concatenates the up-sampled feature maps with feature maps skipped from the encoder.

Due to the precise pixel-level representation, deep learning–based semantic segmentation has been widely adopted in lesion and tumor segmentation, helping doctors get an accurate and explicable diagnosis. Li et al. (2018) proposed H-DenseUNet for liver and liver tumor segmentation. A modification to U-Net was proposed by Zhou et al. (2019), named UNet++, which is applied to a variety of medical datasets for segmentation tasks.

2.3 Multi-Task Learning

Multi-task learning [MTL, Caruana (1997)] is often applied when related tasks can be performed simultaneously. Many MTL methods [Jalali et al. (2010); Misra et al. (2016); Gebru et al. (2017); Strezoski et al. (2019)] have achieved great success in a variety of computer vision tasks. In the medical domain, some recent works also focus on combining classification and segmentation into a joint framework. Yang et al. (2017) proposed a multi-task DCNN model for skin lesion analysis. Multi-task classification and segmentation was proposed by Le et al. (2019) for diagnostic mammography. In the recent COVID-19 pandemic, multi-task learning was applied in CT imaging analysis by Amyar et al. (2020). MTL schemes are based on the assumption that the difficulty of classification and segmentation tasks is the same. But in the real scenes, especially in the small intestine, classification is much simpler than segmentation tasks. Some pioneers have proposed a weighted loss design [Kendall et al. (2018)] and attention module [Liu et al. (2019)] to balance different types of tasks. As shown in **Figure 1**, the evolution of MTL tends to bring more precise control on fusion between different tasks. We dive into this problem and introduce our solution to it.

3 PROPOSED METHOD

To assist the gastroenterologists in capsule endoscopy examination, both precision and interpretability are necessary. Following the previous methods [Le et al. (2019)], we model the precision and interpretability tasks into classification and segmentation tasks. Our proposed multi-task framework shows that joint training of classification and segmentation enables a better cooperation between tasks.

In the following, we first describe the overall framework of our proposed multi-task classification and segmentation network (MTCSN), shown in **Figure 2**. Specifically, a backbone is adopted to extract the representations of the input image which are further used to generate the class label and segmentation map. Next, we introduce the cross fusion module, the key elements of the MTCSN, to alleviate the misalignment between classification and segmentation. Finally, we dive into the inherent problem in the multi-task learning training strategy and introduce our object function.

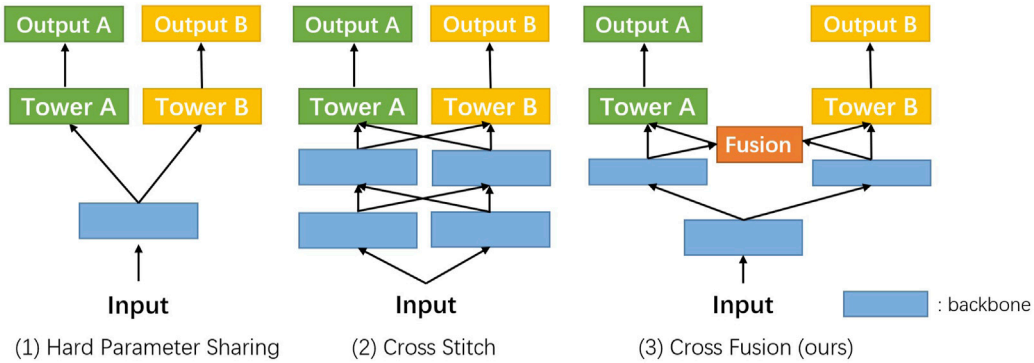


FIGURE 1 | Evolution routing map of the multi-task network structure. Blue rectangles represent the shared layers, like the backbone. Green and yellow ones denote the different task branches. The orange block is the cross fusion module introduced by us in this work.

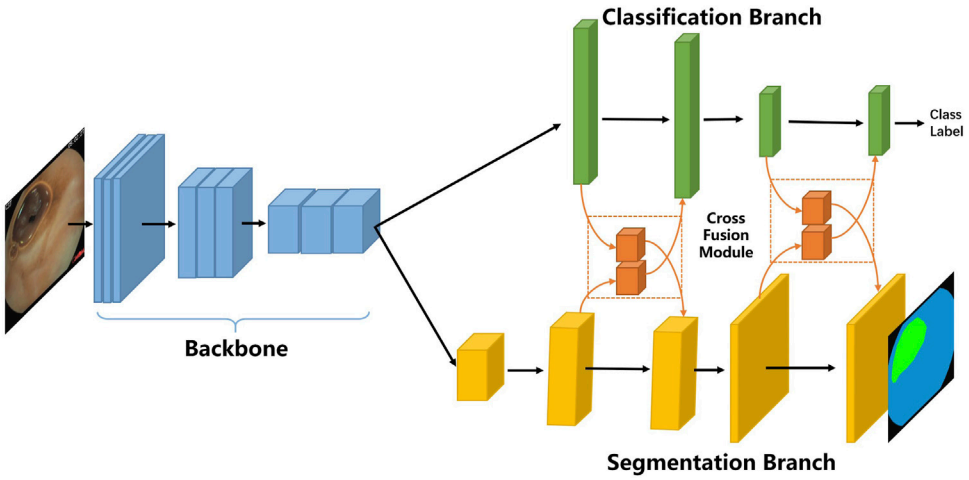


FIGURE 2 | Multi-task classification and segmentation network (MTCSN) architecture.

3.1 Network Architecture

As shown in **Figure 2**, the proposed multi-task classification and segmentation network first utilizes a backbone to extract local features. The backbone we adopted includes different depths of ResNet or DenseNet. Following feature extraction, we design two multi-task branches which are the classification branch for image clearness degree measuring usability and the segmentation branch for tissue segmentation producing explicable visualization to help doctors understand the whole image. The classification branch is mainly constructed by fully connected layers, and the segmentation branch is based on an image-to-image scheme enabling efficient feature computation in each region but also sharing computation from all regions in the whole image in a single forward pass. In addition, we can still process input images with high spatial resolution.

3.2 Cross Fusion Module

Our network mainly focuses on two tasks, classification and segmentation. In the prevailing pattern of MTL, two branches have been trained separately for these tasks following the shared backbone for feature extraction [**Figure 1**]. Because the classification task and segmentation task place different emphasis on feature extraction, performance degeneration is foreseeable and needs to be resolved.

Instead of designing two parallel backbones [Misra et al. (2016)], we set our sights on efficiently exploiting the interaction between the two tasks' branches. We introduce a novel non-linearity cross fusion module which learns the extent of sharing, as illustrated in **Figure 3**.

After global average pooling, the classification branch feature's usual shape is $[C_1, 1, 1]$, where C_1 denotes the number of channels. While the segmentation branch feature's shape is $[C_2, H, W]$, C_2 is usually not the same as C_1. First, we mold the classification feature into the same shape of segmentation. Then, we utilize a sharing parameter non-linearity transformation matrix M to learn the joint representations and extent of fusion automatically. In our experiment setting, M is formulated as a parameter matrix of the convolution layer. More

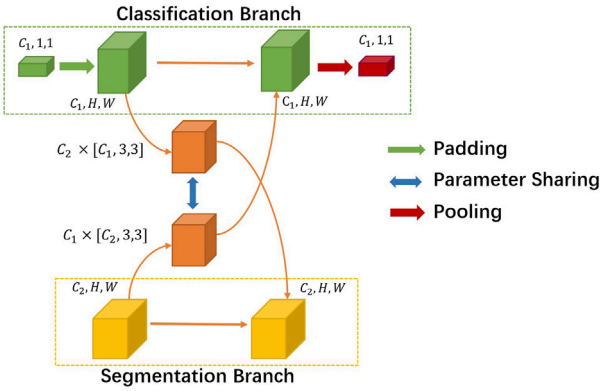


FIGURE 3 | Details of the cross fusion module.

precisely, the process of the cross fusion module can be formulated as

$$\begin{cases} \tilde{X}_{cls} = X_{cls} + Pool\left(M\left(X_{seg}\right)\right), \\ \tilde{X}_{seg} = X_{seg} + M^T\left(Pad\left(X_{cls}\right)\right), \end{cases} \tag{1}$$

where X_{cls} and X_{seg} denote the classification and segmentation feature inputs to cross fusion. M denotes the non-linearity transformation matrix, and M^T's dimension order is different. The output of cross fusion is $\tilde{X}_{cls}$ and $\tilde{X}_{seg}$. The network can automatically decide to make certain layers task-specific by setting a lower weight to the matrix or choosing a more shared representation by assigning a higher value to it.

3.3 Object Functions

In general multi-task learning with K tasks, input X, and task-specific labels Y_i, $i = 1, 2, \ldots, K$, the loss function is defined as

$$\mathcal{L}_{all} = \sum_{i=1}^{K} \lambda_i L_i\left(X, Y_i\right). \tag{2}$$

With task weightings λ_i, $\mathcal{L}_{all}$ is the linear combination of task-specific losses $\mathcal{L}_{all}$. We study the effect of different weighting methods on our multi-task learning approaches. The overall object function of the MTCSN is composed of two parts:

- For the classification task, we apply a class-wise cross-entropy loss for each predicted class label from a softmax classifier:

$$\mathcal{L}_{cls} = \Phi_{CE}\left(X'_{cls}, X_{cls}\right) + \alpha \mathcal{L}_{consistency}, \tag{3}$$

where

$$\mathcal{L}_{consistency} = \sum \Phi_{MSE}\left(X'_i, X_i\right). \tag{4}$$

Here, X'_{cls} is the predicted classification category. X'_i and X_i are the features before and after cross fusion in the classification branch. Φ_{CE} and Φ_{MSE} are the cross-entropy loss and MSE loss functions, respectively. We empirically set the weight $\alpha = 0.1$ in network training.

- For the segmentation task, we apply a pixel-wise cross-entropy loss for each predicted class label from a softmax classifier:

$$\mathcal{L}_{seg} = \Phi_{CE}\left(X'_{seg}, X_{seg}\right), \tag{5}$$

where X'_{seg} represents the predicted segmentation maps.

4 EXPERIMENTS AND DISCUSSION

4.1 Datasets and Tasks

Though Crohn's disease diagnosis is reliable using capsule endoscopy, there is no such open-sourced image dataset for further study so far. So, we build the first Capsule-Endoscopy Crohn's Disease dataset which includes 15 patients and 164 video clips. The dataset will improve the efficiency and accuracy of gastrointestinal endoscopy and help gain a better understanding of this disease.

We divide the annotation process into three stages, and the gastroenterologists are divided into three teams corresponding to these three stages, as shown in **Figure 4**.

In the first stage, gastroenterologists collect the source capsule endoscopy videos from the database center of the hospital, and all the 15 patients' capsule endoscopy videos are filmed by MOMO

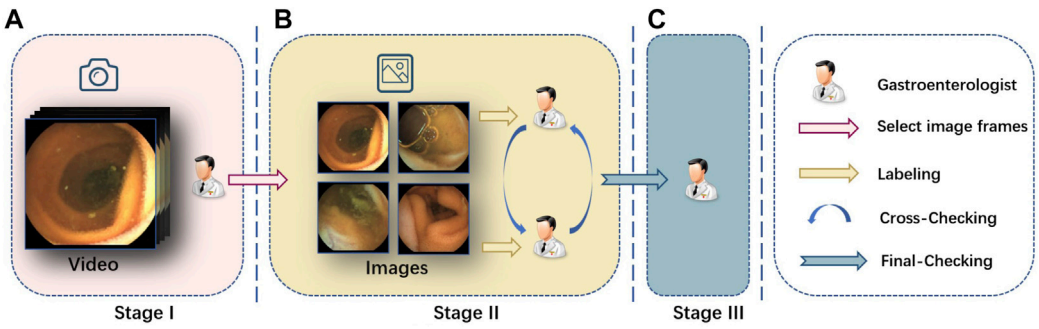


FIGURE 4 | Labeling pipeline we adopted. **(A)** In stage I, we invite several senior gastroenterologists to pick up the video clips of interest, and then we transform them into frames. **(B)** In stage II, we further invite another two gastroenterologists to label the clearness degree of the frames and semantic masks for each part. Cross validation is performed at the same time. **(C)** In stage III, a senior expert checks the labeling and makes the final decision on the annotations.

Wireless Capsule Endoscopy JS-ME-I. Then, we invite several gastroenterologists to pick up the video clip of interest from the full examined videos whose length normally lasts 3–4 h. Finally, we take screenshots from these video clips by a fixed frame rate and get images for follow-up stages.

In the second stage, two gastroenterologists are introduced to label the previous screenshots, respectively, at the pixel level and image level. They first classify the image into three clearness degrees according to adequacy assessment [Brotz et al. (2009)] and then segment the scenes into given categories. In the meantime, one gastroenterologist's annotations will be annotated by another doctor without knowing it, and divergence will be handed over to the third stage's chief to decide.

In the third stage, all revised images are submitted to the chief and expert gastroenterologist in stage III for final-checking. All the data are anonymized for privacy protection.

Here are the statics of the two tasks in our dataset:

1) Task 1: Clearness degree classification
2) Task 2: Tissue segmentation for precise understanding of the image

The total number of annotation images is 467, and we split the dataset into training, validation, and testing datasets strictly by stratifying the sampling in the clearness categories. There are 372 images in the training dataset, 47 images in the validation dataset, and 47 images in the testing dataset. The statistic of basic attribute of our proposed datasets have been shown in **Tables 1**, **2**.

4.2 Evaluation Metrics

The classification results are evaluated by accuracy, precision, recall, and F1 score. A classic classification problem has four possible outcomes, true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Accuracy is the fraction of predictions our model got right. Precision measures the proportion of actually correct positive identifications, and recall answers the proportion of actual positives identified correctly. F1 is an overall measure of a model's accuracy that combines precision and recall:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}, \\ \text{Precision} &= \frac{TP}{TP + FP}, \\ \text{Recall} &= \frac{TP}{TP + FN}, \\ F_1 &= 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \end{aligned} \tag{6}$$

The segmentation results are evaluated using the Jaccard index, also known as Intersection-over-Union (IoU). The IoU is a measure of overlap between the area of the automatically segmented region and that of the manually segmented region. The value of IoU ranges from 0 to 1, with a higher value implying a better match between the two regions. Pixel-wise accuracy is also used for evaluation.

TABLE 1 | Details about the classification category distribution.

Category	Number
Clearness	323
Blur	101
Invisible	42

TABLE 2 | Statistics of segmentation annotation in the dataset.

Category	Number	Category	Number
Clear tissue	361	Invisible by bubble	196
Blur tissue	128	Invisible by excreta	212
Lesion	91	Clear bubble	46
Hole	153		

TABLE 3 | Three-class clearness degree baseline classification results in the CECD dataset.

Classification method	Accuracy	Precision	Recall
ResNet-50	84.0	72.57	72.81
ResNet-101	81.9	69.67	71.41
DenseNet-121	86.7	73.48	73.72

TABLE 4 | Benchmark results in our dataset for the segmentation task.

Segmentation method	Backbone	Iteration	mACC	mIoU
FCN	ResNet-50	30 k	59.5	49.29
PSPNet	ResNet-50	30 k	65.37	54.11
GCNet	ResNet-50	30 k	62.96	53.29
DeepLabv3	ResNet-50	30 k	67.17	54.98

4.3 Experimental Results

In this section, we first evaluate several baselines in our Capsule-Endoscopy Crohn's Disease dataset, respectively, on classification and segmentation tasks. Then, we evaluate our proposed method on two types of tasks. The implementation of our method was done using PyTorch. The model was performed on an Nvidia RTX 2080Ti GPU with 11 gb. The batch size is set to 8, and all images are resized to 240 ∗ 240 to speed up training.

4.3.1 Baselines Results

- **Single Task, Classification Task**. We evaluate two different types of models on our classification problem. **Table 3** shows that existing CNN-based classification models already have an acceptable accuracy, precision, and recall score. On account of the scale of datasets and shape of the input image, a simpler and shallower classification model is preferred.
- **Single Task, Segmentation Task**. We evaluate four different models on our segmentation problem. Under the same backbone, **Table 4** shows that the state-of-the-art segmentation model can achieve competitive results on the CECD dataset. But as shown in **Figure 5**, the

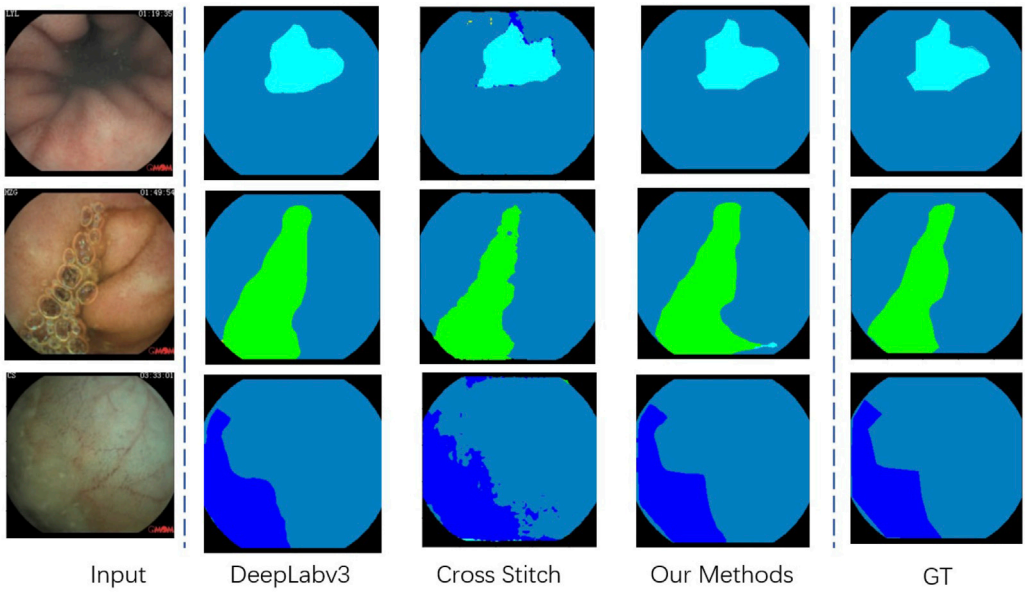


FIGURE 5 | Visualization of the segmentation result in our proposed Capsule-Endoscopy Crohn's Disease dataset.

TABLE 5 | Detailed analysis of our proposed MTCSN in comparison with others.

Our multi-task method	Backbone	Iteration	Accuracy	Precision	Recall	mACC	mIoU
Hard parameter sharing	ResNet-50	30 k	88.41	80.96	78.93	84.92	77.55
Hard parameter sharing	ResNet-101	30 k	83.3	77.43	77.31	83.08	77.46
Hard parameter sharing	DenseNet-121	30 k	87.5	77.66	78.12	82.08	73.79
Cross stitch	ResNet-50	30 k	80.21	68.92	67.71	81.22	73.33
Cross stitch	ResNet-101	30 k	78.13	73.32	72.51	77.94	69.98
Cross stitch	DenseNet-121	30 k	83.3	72.8	73.09	81.31	74.5
MTCSN	ResNet-101	30 k	**84.75**	**77.78**	**77.91**	**83.27**	75.43
MTCSN	DenseNet-121	30 k	**88.3**	**78.7**	**79.64**	**84.49**	73.75
MTCSN	ResNet-50	30 k	**89.23**	**81.54**	**80.14**	**85.50**	**77.62**

Bold values represents our experiment results suppress all the previous methods.

TABLE 6 | Ablation studies of the cross fusion module. The global max pooling (GMP) and global average pooling (GAP) denote the different implementation of the cross fusion module on the class fusion branch.

Segmentation method	Accuracy	Precision	Recall	mACC	mIoU
Global max pooling	85.1	73.52	76.1	84.04	72.9
Global average pooling	88.3	78.7	79.64	84.49	73.75

prediction of DeepLabv3 which performs best among them still has huge room for improvement.

4.3.2 Multi-Task Results

We employ the method described in **Section 3.1** and compare it with two widely used multi-task learning methods, and the results are shown in **Table 5**. Besides, we discuss some structure details when constructing the cross fusion module. We can see from **Table 6** that the GAP pooling method in the cross fusion module performs better than GMP. The reason is that the global max pooling may introduce outliers while emphasizing the maximum in cross features.

Table 6 shows that our proposed multi-task classification and segmentation network, described in **Section 3**, achieved the highest performance in both tasks. Because of the imbalance between the two tasks, if we simply apply a multi-task framework, the promotion of segmentation capacity is at the cost of classification performance. Our proposed cross fusion module elegantly fixes the imbalance between them. The qualitative segmentation can also be seen from **Figure 5**, and the proposed method achieved the best performance.

To the best of our knowledge, no one has previously attempted to utilize segmentation at the pixel level to assist the image-level clearness degree and provide explicable visual results for specialists in clinical practice. In practice, our proposed method will have inference on every

frame of the entire output video of capsule endoscopy. The high clearness frames or frames mostly occupied by tissue or lesions will be marked by our framework. In fact, the marked frames only account for 10% of all frames which significantly reduces the heavy work of gastroenterologists. Our pixel-level semantic segmentation results also provide an explicable reference for gastroenterologists to determine the confidence of the output.

5 CONCLUSION

In this work, we propose a multi-task learning framework named the multi-task classification and segmentation network (MTCSN). This framework combines tissue semantic segmentation and clearness degree classification for capsule endoscopy diagnosis. Our MTCSN achieves high performances on both clearness classification tasks and explicable tissue segmentation offering gastroenterologists visualization to understand the whole image. With explicable tissue segmentation, our framework significantly reduces the workload of gastroenterologists and provides steps forward for deep learning–based methods assisting gastroenterologists in clinical practice.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

ZK, MH, QL, and YC conceived and planned the experiments. ZK, YC, and LC carried out the experiments. XH, PW, and YLi contributed to sample preparation. YLu, XL, and JC contributed to the interpretation of the results. ZK took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

FUNDING

This work is supported by the Nature Science Foundation of China (No. 61972217, 62081360152), Guangdong Basic and Applied Basic Research Foundation (No.2019B1515120049) and Guangdong Science and Technology Department (No. 2020B1111340056).

REFERENCES

Amyar, A., Modzelewski, R., Li, H., and Ruan, S. (2020). Multi-Task Deep Learning Based CT Imaging Analysis for COVID-19 Pneumonia: Classification and Segmentation. *Comput. Biol. Med.* 126, 104037. doi:10.1016/j.compbiomed.2020.104037

Baumgart, D. C., and Sandborn, W. J. (2012). Crohn's Disease. *The Lancet.* 380, 1590–1605. doi:10.1016/s0140-6736(12)60026-9

Brotz, C., Nandi, N., Conn, M., Daskalakis, C., DiMarino, M., Infantolino, A., et al. (2009). A Validation Study of 3 Grading Systems to Evaluate Small-Bowel Cleansing for Wireless Capsule Endoscopy: a Quantitative Index, a Qualitative Evaluation, and an Overall Adequacy Assessment. *Gastrointest. Endosc.* 69, 262–270. doi:10.1016/j.gie.2008.04.016

Caruana, R. (1997). Multitask Learning. *Machine Learn.* 28, 41–75. doi:10.1023/a:1007379606734

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). "Encoder-Decoder With Atrous Separable Convolution for Semantic Image Segmentation," in Proceedings of the European conference on computer vision (ECCV), 801–818. doi:10.1007/978-3-030-01234-2_49

Gebru, T., Hoffman, J., and Fei-Fei, L. (2017). "Fine-Grained Recognition in the Wild: A Multi-Task Domain Adaptation Approach," in Proceedings of the IEEE International Conference on Computer Vision, 1349–1358. doi:10.1109/iccv.2017.151

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778. doi:10.1109/cvpr.2016.90

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely Connected Convolutional Networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 4700–4708. doi:10.1109/cvpr.2017.243

Hwang, Y., Lee, H. H., Park, C., Tama, B. A., Kim, J. S., Cheung, D. Y., et al. (2020). An Improved Classification and Localization Approach to Small Bowel Capsule Endoscopy Using Convolutional Neural Network. *Dig. Endosc.*

Jalali, A., Ravikumar, P., Sanghavi, S., and Ruan, C. (2010). "A Dirty Model for Multi-Task Learning," in Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1 (NIPS'10) (Red Hook, NY: Curran Associates Inc.), 964–972.

Kendall, A., Gal, Y., and Cipolla, R. (2018). "Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics," in Proceedings of the IEEE conference on computer vision and pattern recognition, 7482–7491.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet Classification With Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.*, 1097–1105.

Le, T.-L.-T., Thome, N., Bernard, S., Bismuth, V., and Patoureaux, F. (2019). "Multitask Classification and Segmentation for Cancer Diagnosis in Mammography," in International Conference on Medical Imaging with Deep Learning–Extended Abstract Track.

Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., and Heng, P.-A. (2018). H-Denseunet: Hybrid Densely Connected Unet for Liver and Tumor Segmentation from Ct Volumes. *IEEE Trans. Med. Imaging.* 37, 2663–2674. doi:10.1109/tmi.2018.2845918

Liu, S., Johns, E., and Davison, A. J. (2019). "End-to-end Multi-Task Learning with Attention," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1871–1880. doi:10.1109/cvpr.2019.00197

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully Convolutional Networks for Semantic Segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 3431–3440. doi:10.1109/cvpr.2015.7298965

Min, J. K., Kwak, M. S., and Cha, J. M. (2019). Overview of Deep Learning in Gastrointestinal Endoscopy. *Gut and liver.* 13, 388–393. doi:10.5009/gnl18384

Misra, I., Shrivastava, A., Gupta, A., and Hebert, M. (2016). "Cross-Stitch Networks for Multi-Task Learning," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3994–4003. doi:10.1109/cvpr.2016.433

Podolsky, D. K. (1991). Inflammatory Bowel Disease. *N. Engl. J. Med.* 325, 928–937. doi:10.1056/nejm199109263251306

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional Networks for Biomedical Image Segmentation," in International Conference on Medical image computing and computer-assisted intervention (Springer), 234–241. doi:10.1007/978-3-319-24574-4_28

Simonyan, K., and Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556

Strezoski, G., Noord, N. v., and Worring, M. (2019). "Many Task Learning With Task Routing," in Proceedings of the IEEE International Conference on Computer Vision, 1375–1384. doi:10.1109/iccv.2019.00146

Yang, X., Zeng, Z., Yeo, S. Y., Tan, C., Tey, H. L., and Su, Y. (2017). A Novel Multi-Task Deep Learning Model for Skin Lesion Segmentation and Classification. arXiv preprint arXiv:1703.01025

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). "Pyramid Scene Parsing Network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2881–2890. doi:10.1109/cvpr.2017.660

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2019). Unet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Trans. Med. Imaging.* 39, 1856–1867. doi:10.1109/TMI.2019.2959609

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