DATA-ENABLED INTELLIGENCE FOR MEDICAL TECHNOLOGY INNOVATION, VOLUME I

EDITED BY: Nianyin Zeng, Kathy Clawson and Yonghong Peng
PUBLISHED IN: Frontiers in Public Health, Frontiers in Medical Technology, Frontiers in Computational Neuroscience and Frontiers in Physiology

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-452-7
DOI 10.3389/978-2-88974-452-7

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

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

What are Frontiers Research Topics?

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

DATA-ENABLED INTELLIGENCE FOR MEDICAL TECHNOLOGY INNOVATION, VOLUME I

Topic Editors:
Nianyin Zeng, Xiamen University, China
Kathy Clawson, University of Sunderland, United Kingdom
Yonghong Peng, Manchester Metropolitan University, United Kingdom

Citation: Zeng, N., Clawson, K., Peng, Y., eds. (2022). Data-Enabled Intelligence for Medical Technology Innovation, Volume I. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-452-7

Table of Contents

frontiers in Medical Technology

EDITORIAL
published: 14 January 2022
doi: 10.3389/fmedt.2021.841150

Editorial: Data-Enabled Intelligence for Medical Technology Innovation, Volume I

Nianyin Zeng[1]*, *Kathy Clawson*[2] and *Yonghong Peng*[3]

[1] Department of Instrumental and Electrical Engineering, Xiamen University, Xiamen, China, [2] School of Computer Science, University of Sunderland, Sunderland, United Kingdom, [3] Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, United Kingdom

Keywords: medical data analysis, digital health, data-enabled intelligence, artificial intelligence, machine learning, med-tech

Editorial on the Research Topic

Data-Enabled Intelligence for Medical Technology Innovation, Volume I

OPEN ACCESS

Edited and reviewed by:
Yu-Dong Zhang,
University of Leicester,
United Kingdom

***Correspondence:**
Nianyin Zeng
zny@xmu.edu.cn

Specialty section:
This article was submitted to Medtech Data Analytics, a section of the journal Frontiers in Medical Technology

Received: *22 December 2021*
Accepted: *22 December 2021*
Published: *14 January 2022*

Citation:
Zeng N, Clawson K and Peng Y (2022) Editorial: Data-Enabled Intelligence for Medical Technology Innovation, Volume I. Front. Med. Technol. 3:841150. doi: 10.3389/fmedt.2021.841150

Artificial intelligence (AI) is increasingly being applied to solve real-world problems across a variety of domains and provides opportunities to reshape healthcare, the economy, science, and beyond. In this context, AI- enabled Medical Technology (Med-Tech) has been gaining a significant amount of attention and interest. The development of Med-Tec that incorporate AI, big data methods, and Internet of Things (IoT) can enhance quality of life and has aroused academic interest within the global science community. There is increasing research focusing on how data-driven methods can be utilized to gain greater understanding of medical data, enhance decision making, and leverage operational efficiencies.

Med-tech is of significant importance in this big-data era—it facilitates the development of more advanced devices and corresponding algorithms in order to address global healthcare challenges. This Research Topic seeks original research articles in data science and artificial intelligence that extend medical technology innovation and facilitate a better understanding of the frontiers of AI methods in medical areas.

We are pleased to see the quality and volume of research that was submitted to our topic in data-enabled intelligence for medical technology. We received a total of 28 international submissions from scholars in a variety of countries (China, United States, Switzerland, Netherlands, Sweden, Italy, India, and Germany) and accepted 16 quality and most relevant articles. We provide a brief introduction of accepted papers herein, and welcome readers to refer to these papers and their associated references for more details on the topic. Liangming et al. proposed an obstructive sleep apnea patient rescue monitoring system which provided data support via a Beidou satellite system. Jin et al. presented a breast cancer-specific database and analysis of corresponding applications based on big data techniques. Li M. et al. employed a subtype and stage inference model to identify subtypes of knee osteoarthritis. Liang, Shi, et al. designed a

Gaussian process model to predict knee joint angles via surface electromyography signals. Marini et al. developed a library of convolutional neural network (CNN) structures to extract multi-scale information from whole slide images. Chen B.-q. et al. proposed a separation method on the basis of adaptive linear neuron, where the intrinsic structure of data was depicted by an autoregressive model. Liang, Ye, et al. designed a computer-aided detection (CAD) system based on faster R-CNN for early diagnosis of lung cancer via CT images. Chen B. et al. proposed an adaptive sparse detector for reducing power line interference based on sparse representation. Zhao et al. employed machine learning techniques to build a prediction model for large-scale screening of fatty liver disease. Lakkamraju et al. used fault-tolerant features in a non-invasive medical diagnostic framework to enhance system reliability. Anumukonda et al. designed an ANN-based multi-channel phonocardiography system for extracting cardiac sound components. Song et al. applied a two-step model and a human capital method to assess the economic burden of disease (EBD) in China. Shi et al. proposed a multi-input convolutional neural network to diagnose patellofemoral pain syndrome. Valle et al. developed a standard psychometric platform that could be used for data analysis in the somatosensory neuroprosthetics domain. Li Y. et al. carried out research on time irreversibility of HRV in a hypoxic environment. Wang et al. established an attention mechanism integrated long short-term memory (LSTM) model to predict Kellgren/Lawrence (KL) grade for knee osteoarthritis patients.

We would like to thank all authors again for their contributions to this special topic, and we very much appreciate the efforts of the reviewers for ensuring manuscript quality. It is highly hoped that this special topic could effectively advance the state-of-the-art innovations in medical technology.

AUTHOR CONTRIBUTIONS

NZ wrote the editorial. KC and YP edited the editorial. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the Natural Science Foundation of China under Grant No. 62073271, International Science and Technology Cooperation Project of Fujian Province of China under Grant No. 2019I0003, in part by the UK-China Industry Academia Partnership Programme under Grant No. UK-CIAPP-276, in part by the Open Fund of Engineering Research Center of Big Data Application in Private Health Medicine of China under Grant No. KF2020002.

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.

frontiers in Public Health

ORIGINAL RESEARCH
published: 18 December 2020
doi: 10.3389/fpubh.2020.604654

Causal Discovery in Radiographic Markers of Knee Osteoarthritis and Prediction for Knee Osteoarthritis Severity With Attention–Long Short-Term Memory

Yanfei Wang[1], *Lei You*[1], *Jacqueline Chyr*[1], *Lan Lan*[1], *Weiling Zhao*[1], *Yujia Zhou*[1], *Hua Xu*[1], *Philip Noble*[2] *and Xiaobo Zhou*[1,2]*

[1] School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, United States, [2] McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, United States

The goal of this study is to build a prognostic model to predict the severity of radiographic knee osteoarthritis (KOA) and to identify long-term disease progression risk factors for early intervention and treatment. We designed a long short-term memory (LSTM) model with an attention mechanism to predict Kellgren/Lawrence (KL) grade for knee osteoarthritis patients. The attention scores reveal a time-associated impact of different variables on KL grades. We also employed a fast causal inference (FCI) algorithm to estimate the causal relation of key variables, which will aid in clinical interpretability. Based on the clinical information of current visits, we accurately predicted the KL grade of the patient's next visits with 90% accuracy. We found that joint space narrowing was a major contributor to KOA progression. Furthermore, our causal structure model indicated that knee alignments may lead to joint space narrowing, while symptoms (swelling, grinding, catching, and limited mobility) have little impact on KOA progression. This study evaluated a broad spectrum of potential risk factors from clinical data, questionnaires, and radiographic markers that are rarely considered in previous studies. Using our statistical model, providers are able to predict the risk of the future progression of KOA, which will provide a basis for selecting proper interventions, such as proceeding to joint arthroplasty for patients. Our causal model suggests that knee alignment should be considered in the primary treatment and KOA progression was independent of clinical symptoms.

Keywords: LSTM – Long Short-Term Memory, attention-LSTM, causal inference, prediction model, disease progression

OPEN ACCESS

Edited by:
Nianyin Zeng,
Xiamen University, China

Reviewed by:
Zichen Xu,
Nanchang University, China
Guozheng Rao,
Tianjin University, China

***Correspondence:**
Xiaobo Zhou
Xiaobo.Zhou@uth.tmc.edu

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *10 September 2020*
Accepted: *09 November 2020*
Published: *18 December 2020*

Citation:
Wang Y, You L, Chyr J, Lan L, Zhao W, Zhou Y, Xu H, Noble P and Zhou X (2020) Causal Discovery in Radiographic Markers of Knee Osteoarthritis and Prediction for Knee Osteoarthritis Severity With Attention–Long Short-Term Memory. Front. Public Health 8:604654. doi: 10.3389/fpubh.2020.604654

INTRODUCTION

Osteoarthritis (OA) is a common disease in older individuals, and the economic burden of OA rapidly increases with obesity prevalence and aging in the United States. Knee osteoarthritis (KOA) is the most prevalent type of osteoarthritis with around 22.7% (54.4 million) adults diagnosed with arthritis in the United States (1). The main symptoms of knee OA (KOA) are pain, stiffness, and swelling. These symptoms cause inconvenience to everyday life,

and some people may even lose their physical ability due to this. Once the symptoms appear, it is hard to be cured since damage to the joints cannot be reserved. There is no effective disease-specific drug for this irreversible degenerative disease currently (2). Medication only helps to relieve OA symptoms, primarily pain. Since OA is a slow-developing disease, it is often undiagnosed until symptoms appear, so the best treatment opportunity is missed. It is estimated that the number of total knee replacement surgery or knee arthroplasty (TKA) will reach 1.26 million by 2030 in the United States (3). Considering the rapid increase of KOA patients, it is important to detect KOA at an early stage and perform intervention treatment before knee condition deteriorates.

The most common standard of quantifying the severity of KOA is the five-grade Kellgren and Lawrence (KL) system (4). This grading system divides KOA into four stages, ranging from 0 to 4. Grade 0 indicates no evidence of KOA and grade 4 indicates severe KOA. Medically, incident radiographic KOA is defined when KL grade is $\geq$2 (5). A great number of works have been done to identify risk factors for the occurrence of KOA (6, 7). Zhang et al. used a logistic regression model and identified age, sex, body mass index (BMI), occupational risks, injury, and family history of KOA as risk factors (8). However, these studies only identified risk factors for the occurrence of KOA (whether KL $\geq$ 2), instead of the progression of the disease. There were also some attempts to quantify KOA severity. Du et al. employed support vector machine, random forest, and naive Bayes to predict the progression of KOA from 3-D magnetic resonance (MR) imaging (9). Although these specific models can quantify KOA severity, most attempts were built on image processing and were hard to interpret in the clinical setting. Therefore, our study was designed to build a predictive model with patients' assessment data such as symptoms, questionnaire data, and interpretable image features and identify the key risk factors during the disease progression. Besides the predictive model, we also adopted the causal inference to verify the risk factors identified in predictive models. In order to get higher accuracy, the predictive model may include some unnecessary predictors since it only considers the association between the dependent variable and predictors. However, if physicians take all predictors under consideration in intervention treatment, it may cause overtreatment, because some predictors are not the cause of the disease. In order to eliminate the effect of unknown factors (also called confounders) and avoid unnecessary therapies, we adopted a causal analysis to estimate the causal effect relationship between clinical factors and radiographic markers. The directed graphical causal models (DGCM) can identify the causal relationship instead of association (10). This causal relation is the result of multiple hypothetical experiments by measuring how much an error score will change when the values of a variable are randomly permuted (11). These experimental predictions are computed from the probability distribution. Therefore, a causal relation is defined when variable A and variable B co-vary if we only changed variable A (12). In this study, we used the fast causal inference (FCI) algorithm (12) for causal inference. FCI is designed to test conditional independence. It first generates a complete undirected graph and then deletes recursively edges based on the conditional independence decisions.

The purpose of this study is to use short-term data to predict long-term KOA progression by inputting observed time series into an attention–long short-term memory (LSTM) model and outputting the likelihood of patients' KL grade. Additionally, we built a causal model that evaluates the causal structure of potential predictors and identified the primary contributors to KOA progression and pain progression. To characterize the OA progression, we used 5-year data from the Osteoarthritis Initiative (OAI), specifically patients' clinical assessment data such as symptoms, questionnaire data, and interpretable radiographic image features.

MATERIALS AND METHODS

Study Population

The OAI is a multicenter observational cohort study with longitudinal clinical and image data. This database includes MRI/CT imaging data, genotyping data, and clinical data for evaluating potential biomarkers and characterizing OA incidence and progression. Individuals from this study were between the ages of 45 and 79 and were at high risk for KOA. We selected patients with at least five visits over the 5-year longitudinal study and excluded patients who underwent knee replacement surgery. In total, 518 patients were included in our study. Among these patients, 394 patients remained in the same level of KL while the knee conditions for other 124 patients had worsen during the 5-year following-up period. We used predictive mean value to impute the missing values. For patients with OA in both knees, we selected the knee with a higher KL grade. We also extracted data from the Cerner Health Facts database to cross-validate our predictive model. This database contains clinical health records from over 500 health care centers across the United States. This clinical dataset uses ICD-9/10 diagnosis codes instead of KL grades.

Predictors

To capture the full picture of disease progression, we extracted clinical data, questionnaires, and radiographic markers from the OAI. In order to assess the functional status of patients, we adopted the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). The WOMAC questionnaire evaluates the level of pain, stiffness, and function of the knee. The American College of Rheumatology considers this questionnaire as the gold standard for KOA functional status (13). For the clinical data, we chose patients' characteristics that could potentially predict future KOA, such as age, BMI, and measurements of physical ability. The radiographic markers, such as joint space width and knee alignment (the hip–knee–ankle angle), were extracted from X-ray. The minimal joint space width (mJSW) has been considered as a proxy for cartilage thickness (14). We first analyzed different independent predictors of KL grade in a multivariate logistic regression model, which is assumed as the basic standard in KOA analysis. BMI and age were analyzed as a categorical variable. BMI is divided into four groups: underweight (BMI 10–19), normal (BMI 20–26),

overweight (BMI 26–34), and obese (35+). Age was divided into 5-year intervals, namely 50–54, 55–59, 60–64, 65–69, 70–74, and 75–79. We set the normal BMI and age between 45 and 49 as the reference for the odds ratio. For the questionnaire, we marked "no symptom" as the reference and the highest score as the worst condition of the individual in that question. **Table 1** shows the independent predictors of KL grade in a multivariate logistic regression model.

Predictive Model

The traditional prediction methods based on time series primarily comprise the autoregressive integrated moving average model (ARIMA), hidden Markov model (HMM), and recurrent neural networks (RNN) (15). However, ARIMA models are usually applied where data shows evidence of non-stationarity and suitable for numerical sequence (16). Since we included some questionnaires as inputs, ARIMA is not a good option. A Markov model is a useful tool to characterize the moving process where an individual transits among multiple states. In this study, we used Markov processes to estimate KOA disease stage transition probabilities. However, one limitation of the Markov model is that it assumes that the future state only depends on the current state (17). Unlike the Markov model, RNN allows the future states to depend on all past states (18).

RNN is a special neural network which could efficiently pass forward information to the next cell at each point (19). For each hidden layer at time point t, it not only includes the input layer at time point t but also considers the output of the hidden layer at the previous time point $t-1$(20). Although RNN did a very good job in dynamically combining the sequential information based on its internal recurrence (21), RNN may have some issues of gradient vanishing when dealing with long-term data. Therefore, an advanced type of RNN named LSTM was designed to solve the gradient disappearing (22). Unlike other RNNs, LSTM introduced a forget gate (22) to decide whether to keep or drop a cell state based on the previous hidden state and the current input variables. Another important component of LSTM is cell state which is used to control whether to add or remove the information in the previous cell state. It has been proven that LSTM did a better job in long-term time series data (23).

The attention mechanism focuses on certain time points in the time series when processing the data. For example, BMI is very important in the early stage of the disease, while joint space width is more important in the late stage. It allows the model to pay more attention to the most important time point based on what it has learned so far. We adopted the attention mechanism from Nauta's repository and calculated the context vector as a weighted sum of each input vector instead of each time series so that an attention vector learns weights corresponding to input features (11). The attention mechanism assigns a different weight to different variables based on its ability to forecast. All initial attention scores are set to one and updated in every training epoch.

Causal Inference

In this study, we used FCI with directed acyclic graphs (DAGs). DAGs are commonly used to represent causal relationships, where vertices denote variables and the edges represent causal relationships between variables. The PC algorithm uses statistical tests to find conditional independence and constructs the structure of DAGs based on the results (12). The FCI algorithm is a generalization of the PC algorithm, except that it allows the existence of confounder variables (12). The FCI algorithm is able to detect a Markov equivalence class of DAGs with latent variables based on conditional independence information from the observed variables (12).

There are two important structures in FCI: the "V" structure and the "Y" structure (24). The "V" structure is defined when A and C are independent but dependent conditionally on B, marked as A→B←C. The "Y" structure is defined when A and C are independent of D conditional on B, marked as A→B←C and B→D. The FCI starts with a complete, undirected graph and removes recursively edges based on conditional independence decisions. After finding the skeleton of DAG, edges are oriented by identifying the "V" and "Y" structures, and further orientation rules given by Zhang (25) are applied.

Validation and Data Integration

We used 10-fold cross-validation for performance evaluation and compared true KL grade vs. the KL grade predicted. **Figure 1** reports areas under the receiving operating characteristic curves (ROC). The x-axis presents sensitivity (true-positive rate) and the y-axis represents specificity (false-positive rate). The area under the curve is defined as AUC, which is a standard of performance of classification. The higher the AUC is, the better the classifier.

One limitation of the OAI dataset is that not every hospital measures KOA-specific features. In order to include these clinical indicators, we chose 30 clinical features from the Cerner database, which are commonly used for KOA diagnosis, but not included in the OAI, such as pulse popliteal of the knee. Previous studies identified age and BMI as the most significant risk factors in the development of KOA (26); therefore, we assumed that patients may have similar clinical indicators with those who shared the same BMI and age. Under this assumption, we used values from the Cerner dataset to estimate the values of variables that are missing in the OAI dataset. For each patient in the OAI dataset, we extracted clinical features from age- and BMI-matched patients from the Cerner database. OAI patients who are age- and BMI-matched with only one patient from the Cerner dataset are directly assigned the matched patient's Cerner clinical values. For patients who matched with multiple patients from the Cerner dataset, we took the average of matched patients and assigned averages to corresponding OAI patients. However, the prediction accuracy of LSTM after this imputation reduced to 85%. Therefore, we applied autoencoder and principal components analysis (PCA) as the denoising feature extractor. The performance of autoencoder is better than PCA and achieved a prediction accuracy of 93% in the dataset combined with the OAI and Cerner data.

RESULTS

Characterizing Disease Progression

Before we built the predictive model, we used logistic regression for feature selection. Compared with the other predictive models, such analysis can avoid confounding effects by considering the

TABLE 1 | Multivariable logistic regression model demonstrating independent predictors of KL >1.

Predictors	Odds ratio	[95% conf. interval]		$p > z$	p (>Chi)
JSPAINPRG		JSL and pain progressor			0.3391265
1	0.63	0.39	1.01	0.054881	
LKALNMT	1.02	0.95	1.1	0.623373	0.3143121
RKALNMT	0.96	0.89	1.03	0.263646	0.56715
JSONLYPRG		JSL only progressor			0.287
1	0.5	0.29	0.87	0.013476	
PAINONLYPRG		Pain only progressor			0.3808985
1	0.73	0.44	1.21	0.212488	
XRJSM	4.1	2.9	5.86	<0.001	<0.001
XRJSL	1.04	4.38	2.67	0.934219	0.5478206
MCMJSW	0.84	6.83	1.04	0.11842	0.19541
BMI		Body mass index			<0.001
Overweight	3.35	2.01	5.58	<0.001	
Obese	5.44	3.23	9.2	<0.001	
Morbidly obese	13.81	3.49	93.9	0.001	
WOMKP	1.16	1.03	1.3	0.012959	<0.001
WOMADL	0.97	0.94	1.01	0.174	0.9597197
P02KPN		Either knee pain, aching or stiffness: any, in the past 12 months			<0.001
1	0.2	0.08	0.43	<0.001	
P01KPACT30		Whether either knee, limit activities due to pain, aching or stiffness, past 30 days			0.490686
1	0.89	0.6	1.33	0.582108	
SF2	How much health limit involvement in moderate activities (e.g., moving a table, pushing a vacuum cleaner...)				0.1215677
1: Yes, limited	Ref				
2: Limit a little bit	0.32	0.1	0.87	0.03652	
3: Not limited at all	0.57	0.18	1.49	0.284351	
WSRKN1		Right knee stiffness: in the morning, the last 7 days			0.151677
1	0.6	0.37	0.95	0.028428	
2	0.5	0.25	0.97	0.039967	
3	0.55	0.16	2.11	0.353752	
WSRKN2		Right knee stiffness: later in the day, the last 7 days			0.8830752
1	0.96	0.59	1.58	0.85981	
2	1	0.49	2.07	0.989272	
3	0.93	0.24	4.21	0.913856	
KSXRKN1		Right knee symptoms: swelling, the last 7 days			<0.001
1	2.1	1	4.88	0.063434	
2	2.06	0.89	5.21	0.105648	
3	2.57	0.66	13.29	0.206253	
4	0.67	0.22	2.3	0.495412	
KSXRKN2	Right knee symptoms: feel grinding, hear clicking or any other type of noise when knee moves, the last 7 days				0.1097449
1	1.8	0.96	3.52	0.075166	
2	1.49	0.88	2.57	0.141885	
3	2.28	0.98	5.63	0.062791	
4	2.52	0.77	9.27	0.141467	
KSXRKN3		Right knee symptoms: knee catch or hang up when moving, the last 7 days			0.0368252
1	2.9	1.38	6.8	0.008418	
2	1.62	0.72	3.86	0.256756	
3	0.93	0.23	4.88	0.920198	
KSXRKN4		Right knee symptoms: straighten knee fully, the last 7 days			0.1337019
1	3.86	1.21	14.23	0.029485	
2	6.09	0.85	132.68	0.130066	

(Continued)

TABLE 1 | Continued

KSXRKN5		Right knee symptoms: bend knee fully, the last 7 days			0.790488
1	0.64	0.28	1.55	0.307365	
2	0.6	0.18	2.44	0.433231	
3	0.33	0.08	1.52	0.124832	
4	0.32	0.08	1.46	0.114924	
KSXLKN1		Left knee symptoms: swelling, the last 7 days			0.1417692
1	2.32	1.04	5.73	0.050612	
2	1.35	0.63	3.13	0.460365	
3	3.93	0.87	30.51	0.117698	
4	2.67	0.69	14.19	0.193381	
KSXLKN2	Left knee symptoms: feel grinding, hear clicking or any other type of noise when knee moves, the last 7 days				0.8641942
1	0.84	0.45	1.61	0.58539	
2	0.73	0.42	1.26	0.251584	
3	0.82	0.35	2.01	0.649514	
4	0.76	0.24	2.61	0.651474	
KSXLKN3		Left knee symptoms: knee catch or hang up when moving, the last 7 days			0.7839969
1	0.97	0.49	1.99	0.924445	
2	1.1	0.5	2.57	0.81326	
3	0.33	0.064	2.62	0.22145	
KSXLKN4		Left knee symptoms: straighten knee fully, the last 7 days			0.1494063
1	0.2	0.07	0.59	0.00364	
2	0.44	0.08	3.62	0.380026	
3	0.04	0.005	0.49	0.006329	
KSXLKN5		Left knee symptoms: bend knee fully, the last 7 days			0.0377456
1	3.45	1.18	11.47	0.03219	
2	3.66	1.15	13.96	0.039426	
3	7.66	1.31	70.22	0.039477	
4	5.76	1.17	48.2	0.057999	
Age					<0.001
50–54	1.62	0.85	3.17	0.161995	
55–59	3	1.47	6.1	0.002337	
60–64	2.09	1.02	4.21	0.039225	
65–69	3.23	1.51	6.9	0.00243	
70–74	5.61	2.43	13.21	<0.001	
75–79	2.32	1.02	5.29	0.044134	

association of all variables (27). The result of multivariable logistic regression is presented in **Table 1**. Odds ratio (OR) is the constant effect of a predictor on the occurrence of outcome. Susceptible risk factors, including BMI and age, were associated independently with increased risk of KOA severity. Compared with normal BMI (18.5–24.9), BMI was associated with increased likelihood of KOA severity and the odds of KOA severity increase as BMI increases (overweight: OR: 3.35; 95% CI: 2.01–5.58, $p < 0.001$; obese: OR: 5.44; 95% CI: 3.23–9.20, $p < 0.001$; morbidly obese: OR: 13.81; 95% CI: 3.49–93.9, $p = 0.001$). Age is also a risk factor for KOA. The OR for people between ages 70 and 74 has the highest OR (OR: 5.61; 95% CI: 2.43–13.21, $p < 0.001$). Susceptible joint factors, including joint space narrowing (XRJSM) (OR: 4.10; 95% CI: 2.9–5.86, $p < 0.001$) and knee pain, aching, or stiffness in the past 12 months (P02KPN) (OR: 0.2; 95% CI: 0.08–0.43, $p < 0.001$), were associated with increased risk of occurrence of KOA independently. Swelling in the knee (KSXRKN1) was associated with successively increased odds of occurrence of KOA. Based on the result, we removed three variables from the predictor list for the predictive models, including right knee stiffness status later in the day, the ability to bend right knee fully in the last 7 days, and whether left knee feels grinding.

Predicting Disease Progression With Attention–LSTM

We investigated the performance of LSTM algorithms in predicting KL grade using clinical data spanning 1 year. The predicted accuracy of LSTM achieved 90%. We compared the LSTM model against the previous models, namely, random forest, support vector machine, and naive Bayes. Random forest

(RF) constructs a set of multiple decision trees on data and then averages the prediction from each of them. Compared with a single decision tree, RF reduces the chance of overfitting. Support vector machine (SVM) adopts kernels, which transform a lower dimensional input space into a higher dimensional space. This conversion made SVM more flexible and accurate. Naive Bayes is another classification technology based on Bayes' theory, which assumes that all the predictors are independent to each other. Our LSTM model performs better than RF, SVM, and naive Bayes in predicting KL grade. ROC curves are shown in **Figure 1**. A value of 0.5 in AUC represents as a random guessing and a value higher than 0.8 is considered quite good. Based on the results, LSTM and RF both did a good job in KL classification. The AUC value of LSTM is higher than that of SVM and naive Bayes for

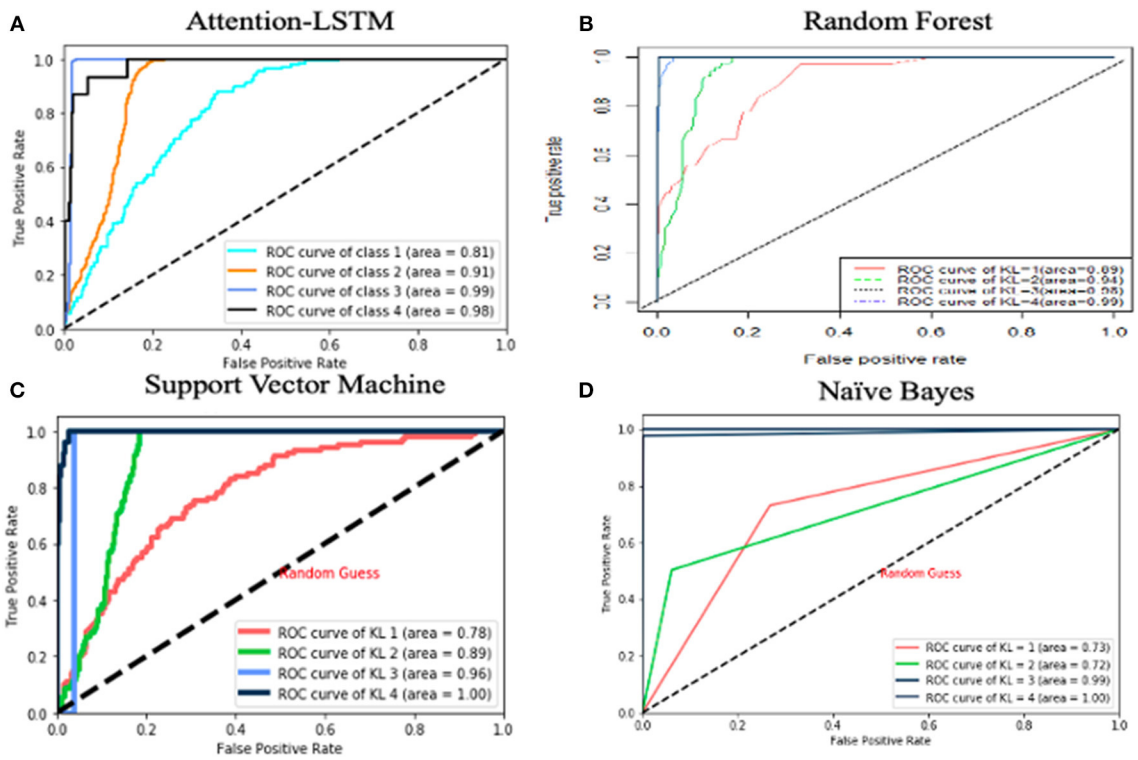


FIGURE 1 | AUC curve for predictive models. The x-axis represents sensitivity and the y axis represents specificity.

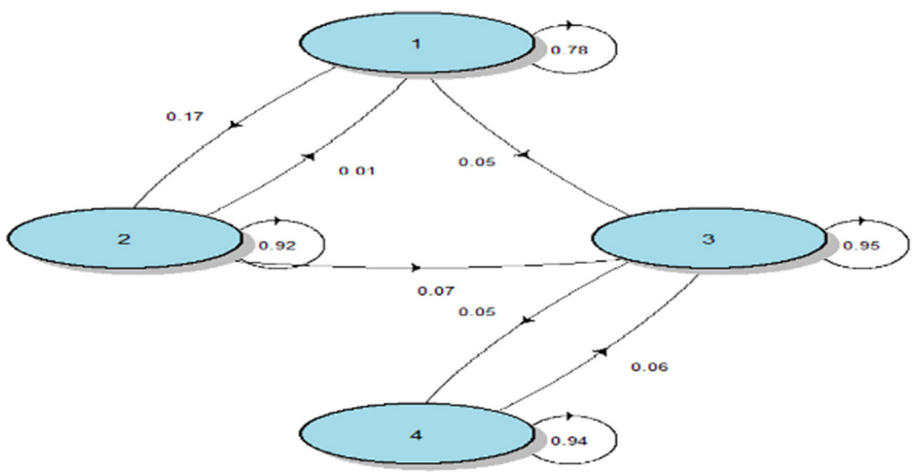


FIGURE 2 | Markov model for KOA progression. Rates of transition between four stages of OA. The number 1, 2, 3, 4 represent four stages of KOA. The arrows are the transitions rates between these states.

all KL grades, and LSTM has better performance than random forest for KL = 3/4. In other words, random forest did a better job in diagnosing KOA, while LSTM did a better job in predicting KOA severity.

To better describe disease progression, we consider Markov states as disease stages. The Markov chain result is shown in **Figure 2**. For example, the patients at high risk (KL = 1) have 17% chance to move to the next stage (KL = 2) and 78% chance to maintain their KL grade after the 1-year follow-up visit. It should be noted that patients diagnosed with knee osteoarthritis (KL $\geq$ 2) are more likely to remain at this stage and very less likely to revert back to previous disease states.

Importance Analysis of Variables

Figure 3 graphically depicts the important indexes of random forest. We used the mean decrease in accuracy index to evaluate the importance of variables in classification. It shows that the variable "XRJSM" (joint space narrowing) stands out among all the variables with the largest mean decrease in accuracy. Variables "P01BMI" (BMI), "MCMJSW" (medial minimum joint space width), "V00AGE" (age), and "WOMADL" (WOMAC disability score) are also relatively important for predicting KOA severity based on the indexes of variable importance.

The attention maps in **Figure 4** demonstrate how attention scores can identify the dynamics. Although it was hard to distinguish which visit would have the model attended the most, XRJSM (OARSI joint space narrowing grade) was significantly important in predicting KL grade. The attention mechanism identified that joint space narrowing leads to the worsening status of KOA, which causes pain. The finding also explained why joint space narrowing is the most important factor in predicting KL progression in the LSTM model.

Causal Relationship of Variables

To understand the dependency and independency of important features in predicting KL progression, we employed FCI. We found that misalignments of the left and right knee are the reason for joint space narrowing. The angles of tibiofemoral and patellofemoral joints affected the alignment of the knees and caused joint space narrowing. This is consistent with previous reports that various knee alignment is associated with the radiographic measures of KOA severity (28). The FCI output is presented in **Figure 5**. The bidirected edge between left alignment and right alignment indicates that there exists at least one unmeasured confounder of left alignment and right alignment. The "o" symbols at alignment and joint space narrowing (JSM) indicate that it is difficult to distinguish whether

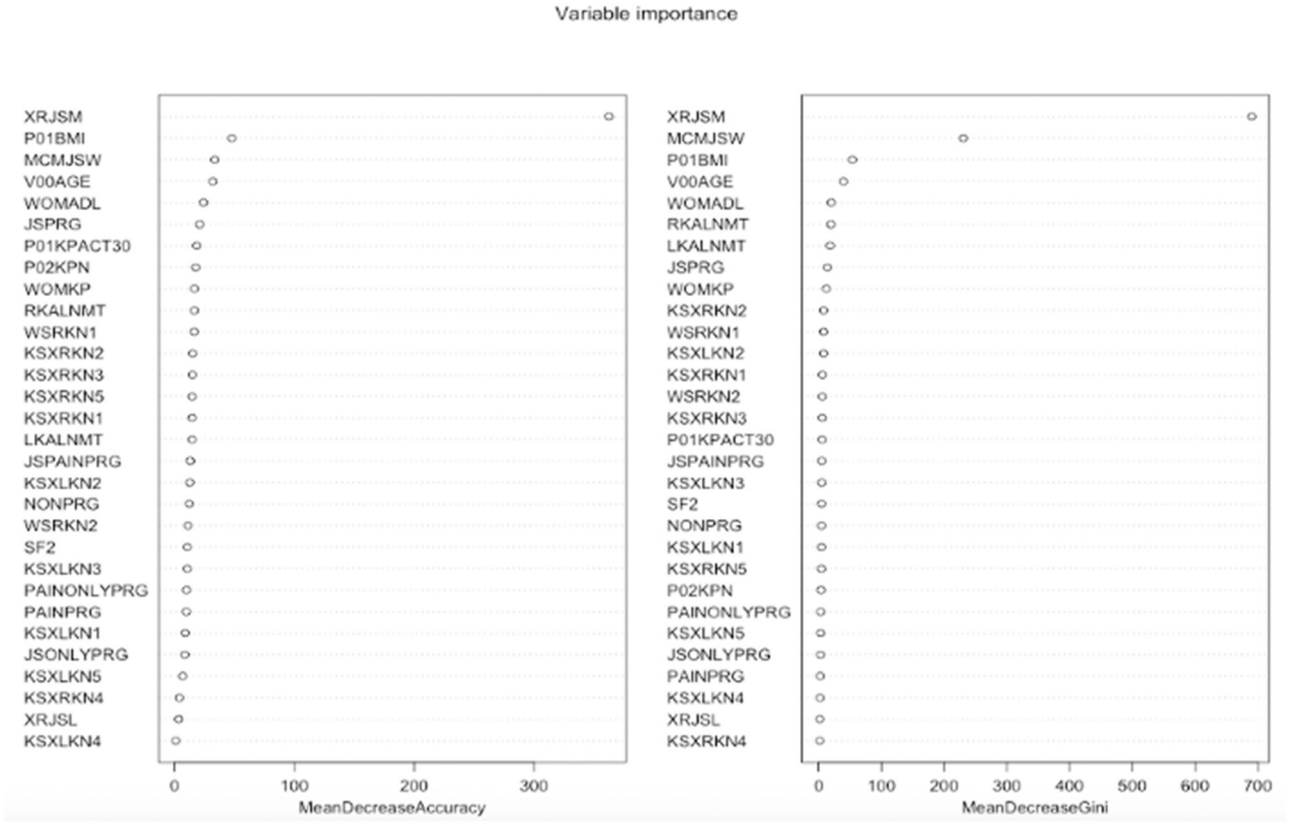


FIGURE 3 | Features importance of random forest. The left figure is mean decrease in accuracy which measures misclassification of removing the given variable. The right figure is mean decrease in Gini which measures the average gain by splitting the given variable.

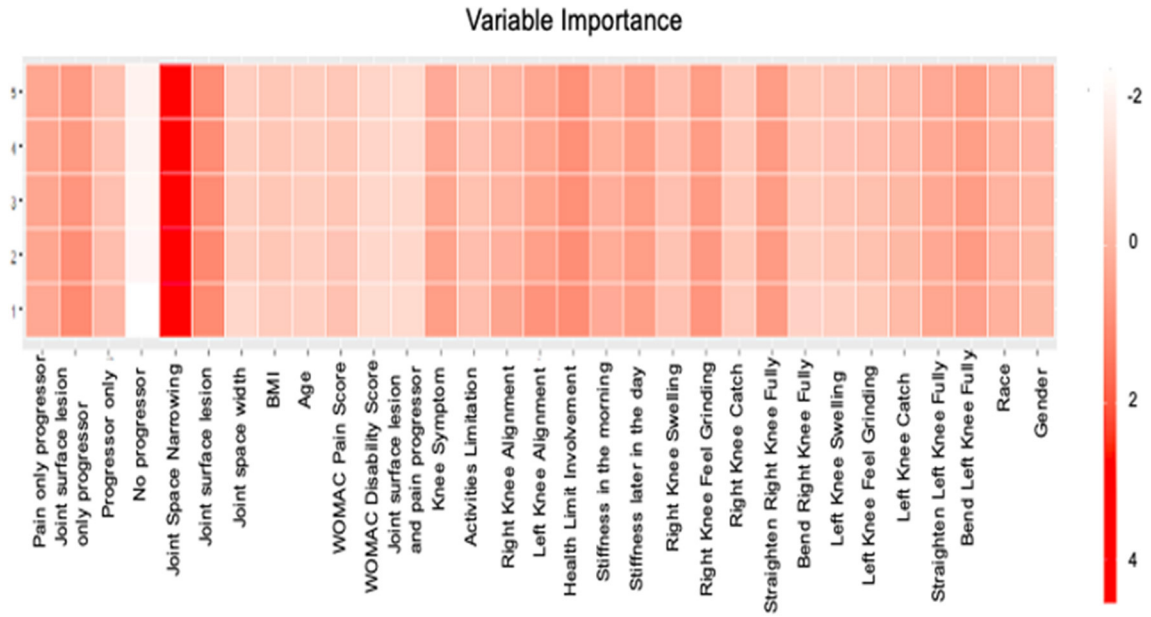


FIGURE 4 | Features Importance of LSTM. The x-axis represents the name of variable and the y axis represents the number of visit. The color represents the attention score.

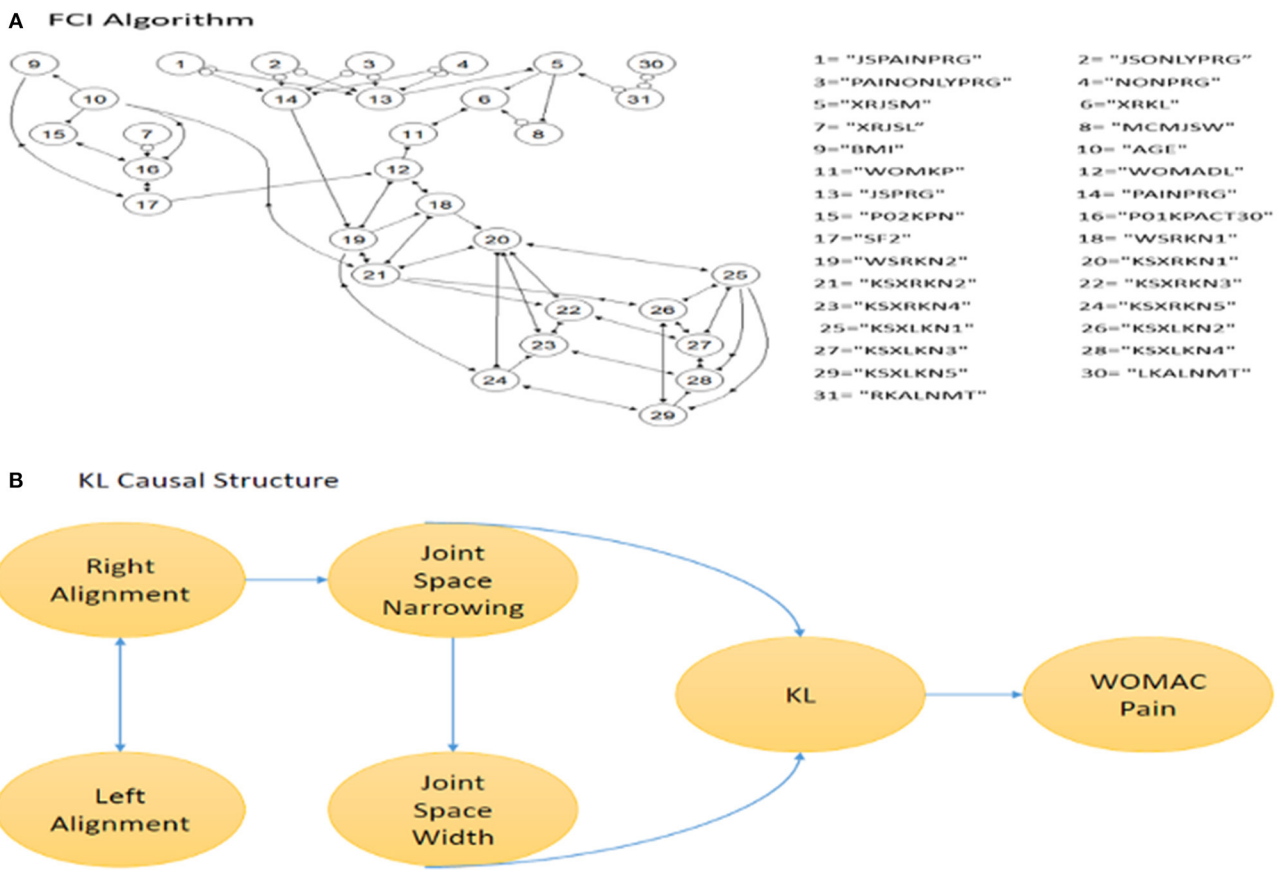


FIGURE 5 | Causal Inference. **(A)** the result of FCI in DAGs. The number denotes the variable and the edges represent the causal relationship. **(B)** is snapshot of a small part of **(A)**.

the connection between alignment and JSM connection is a directed edge or an unmeasured confounder. This result also eliminated some unnecessary predictors from the LSTM model. Although the symptoms, such as swelling, feel grinding, knee catch, straighten knee full, and bend knee full, may be good predictors for KL grade, they are not the reason causing KOA progression. Symptoms seem to be a good predictor of pain progression. Using FCI, we find that the change in joint space is the real reason for disease progression.

DISCUSSION

KOA is a slowly progressive disease with irreversible joint damage. With obesity prevalence and aging in the United States, KOA becomes the most frequent disease. However, an accurate predictive model with diagnostic criteria is still unavailable. Therefore, the goal of this study was to develop a mathematic model to predict OA severity and the key risk factors associated with disease progression. In this study, we used logistic regression for feature selection. After removing the irrelevant variables, we employed the attention–LSTM predictive model to predict OA severity and compared this approach against the existing methods. RNN, especially LSTM, did a good job in modeling long-term dependency in time series data. The AUC values of LSTM for KL = 1, 2, 3, and 4 are 0.81, 0.91, 0.99, and 0.98, respectively. The attention mechanism allows us dynamically to detect the feature importance across multiple time steps for predicting KOA progression. Meanwhile, the attention scores extracted from the attention mechanism would help to discover the direction of causal relationship. Finally, we use causal inference to interpret the inside connection of variables. To validate our predictive model, we used the OAI database in conjunction with the Cerner Health Facts database. Our experiments showed that the attention–LSTM with the scaled autoencoder resulted in 3% increased accuracy (from 90 to 93%). Hence, this method can be a promising tool for patients and doctors to prescreen for possible osteoarthritis to prevent deterioration of OA, thereby supporting for clinical decision-making.

Our attention–LSTM model not only reliably and accurately predicts KL grade and progression but also identifies the key factors among different time points. Using our model, clinicians can predict the possible KOA progression of patients and take preventative measures in advance. The attention mechanism dynamically shows the importance of variables at different disease stages and indicates that joint space narrowing is the primary factor in KOA progression at all time points. This method will provide strong support in clinical decision-making, including diagnosis, appropriate treatments, and preventive care. Using causal inference analysis, we identified the real cause of joint space narrowing and pain. They are misalignments of tibiofemoral and patellofemoral joints. Current primary preventive intervention is limited to weight loss (29). This finding recommended another possible intervention in clinical practice. The causal discovery also helped providers to avoid overtreatment. For example, treatment for symptoms such as swelling, grinding, catching, and inability to straighten or fully flex the knee may not be able to prevent the worsening status of knee OA. Our work had two major clinical contributions. (1) We evaluated a broad spectrum of potential risk factors (clinical variables such as age, BMI, clinical symptoms, WOMAC questionnaires, and image measurements from X-rays) and investigated the performance of RNN algorithms in predicting KL grade. By using Markov hidden states as the disease stages, we have gained more knowledge about the stage transition, which enables a deeper understanding of the temporal progression of OA. (2) Considering the existence of hidden confounding variables, we built a causal structure of candidate risk factors and identified the preventable factors for treatment.

This study has several limitations to be considered. First, the research target of the OAI is individuals at high risk, with expected overweight and aging population. Thus, our model may not be applicable to the general population. The continuing work will focus on testing its generalizability of the models to different populations. To improve the expansion ability, we would consider transfer learning and generative-adversarial-network-based method in further studies. Second, when interpreting the findings, we found that symptoms of right knee and left knee have different impacts on KOA progression. This finding required some external validation. An additional limitation of this study is the small sample size. We have included prominent OA symptoms such as swelling joint pain, stiffness, and bending (30) in this study. However, we were unable to adjust for factors that may affect long-term outcomes, such as other symptomatic joint diseases.

In conclusion, we used attention scores in LSTM to describe feature importance at different time points and compared our model with previous works. In addition, we used causal inference to identify the key diagnostic criteria in disease progression. Our study has illustrated that clinical symptoms are important in predicting disease severity but may not be essential in disease progression. With the help of causal inference, LSTM is a better tool to help physicians in decision-making.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

YW: data preparation, predicting models, causal inference, statistical analysis, and manuscript preparation. LY: study design and manuscript editing. JC: manuscript editing. LL: manuscript review. PN: clinical support and manuscript editing. XZ: study design, manuscript editing, and responsible for the integrity of this work. WZ: final version manuscript review. YZ: data extract and data management from Cerner Heath Facts database. HX: responsible for Cerner Health Facts and manuscript review. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Institutes of Health (NIH) under Grant 1R01DE027027-02 and Grant 1U01 AR069395-03 (XZ).

ACKNOWLEDGMENTS

This was a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors. We thank the members of the Center for Computational Systems Medicine (CCSM) for valuable discussion. National Institutes of Health (NIH) (1R01DE027027-02 and Grant 1U01 AR069395-03 to XZ). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Funding for open access charge: Dr & Mrs Carl V. Vartian Chair Professorship Funds to YZ from the University of Texas Health Science Center at Houston.

REFERENCES

1. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, et al. Estimates of the prevalence of arthritis other rheumatic conditions in the United States. Part II. *Arthritis Rheum.* (2008) 58:26–35. doi: 10.1002/art.23176
2. Losina E, Daigle ME, Suter LG, Hunter DJ, Solomon DH, Walensky RP, et al. Disease-modifying drugs for knee osteoarthritis: can they be cost-effective? *Osteoarthritis Cartilage.* (2013) 21:655–67. doi: 10.1016/j.joca.2013.01.016
3. Sloan M, Premkumar A, Sheth NP. Projected volume of primary total joint arthroplasty in the U.S., 2014 to 2030. *JBJS.* (2018) 100:1455–60. doi: 10.2106/JBJS.17.01617
4. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis.* (1957) 16:494–502. doi: 10.1136/ard.16.4.494
5. Kohn MD, Sassoon AA, Fernando ND. Classifications in brief: kellgren-lawrence classification of osteoarthritis. *Clin Orthop Relat Res.* (2016) 474:1886–93. doi: 10.1007/s11999-016-4732-4
6. Sharma L, Hochberg M, Nevitt M, Guermazi A, Roemer F, Crema MD, et al. Knee tissue lesions and prediction of incident knee osteoarthritis over 7 years in a cohort of persons at higher risk. *Osteoarthritis Cartilage.* (2017) 25:1068–75. doi: 10.1016/j.joca.2017.02.788
7. Neogi T, Zhang Y. Epidemiology of osteoarthritis. *Rheum Dis Clin North Am.* (2013) 39:1–9. doi: 10.1016/j.rdc.2012.10.004
8. Zhang W, McWilliams DF, Ingham SL, Doherty SA, Muthuri S, Muir KR, et al. Nottingham knee osteoarthritis risk prediction models. *Ann Rheum Dis.* (2011) 70:1599. doi: 10.1136/ard.2011.149807
9. Du Y, Almajalid R, Shan J, Zhang M. A novel method to predict knee osteoarthritis progression on MRI using machine learning methods. *IEEE Trans NanoBiosci.* (2018) 17:228–36. doi: 10.1109/TNB.2018.2840082
10. Glymour C, Zhang K, Spirtes P. Review of causal discovery methods based on graphical models. *Front Genet.* (2019) 10:524. doi: 10.3389/fgene.2019.00524
11. Nauta M, Bucur D, Seifert C. Causal discovery with attention-based convolutional neural networks. *Mach Learn Knowl Extr.* (2019) 1:312–40. doi: 10.3390/make1010019
12. Spirtes P, Glymour C, Scheines R. *Causation, Prediction, and Search*. Cambridge, MA: The MIT Press (2001). doi: 10.7551/mitpress/1754.001.0001
13. Fernandes MTP, Fernandes KBP, Marquez AS, Cólus IMS, Souza MF, Santos JPM, et al. Association of interleukin-6 gene polymorphism (rs1800796) with severity and functional status of osteoarthritis in elderly individuals. *Cytokine.* (2015) 75:316–20. doi: 10.1016/j.cyto.2015.07.020
14. Castaño-Betancourt MC, Evans DS, Ramos YFM, Boer CG, Metrustry S, Liu Y, et al. Novel genetic variants for cartilage thickness and hip osteoarthritis. *PLoS Genet.* (2016) 12:e1006260. doi: 10.1371/journal.pgen.1006260
15. Almqvist O. *A comparative Study Between Algorithms for Time Series Forecasting on Customer Prediction: An Investigation Into the Performance of ARIMA, RNN, LSTM, TCN and HMM.* (2019). Available online at: http://www.diva-portal.org/smash/get/diva2:1321224/FULLTEXT01.pdf
16. Boshnakov G, Introduction to Time Series Analysis Forecasting, 2nd Edition, Wiley Series in Probability Statistics, by Douglas C.Montgomery, Cheryl L.Jennings MuratKulahci (eds). Published by John Wiley Sons, Hoboken, et al. Total number of pag: INTRODUCTION TO TIME SERIES ANALYSIS AND FORECASTING, 2ND EDITION, WILEY SERIES IN PROBABILITY AND STATISTICS, by Douglas Montgomery C, Cheryl L. Jen. *J Time Ser Anal.* (2016) 37:864. doi: 10.1111/jtsa.12203
17. Awad M, Khanna R. *Hidden Markov Model, Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, Apress, Berkeley, CA*. (2015). p. 81–104. doi: 10.1007/978-1-4302-5990-9_5
18. Alaa AM, van der Schaar M, *Forecasting Individualized Disease Trajectories Using Interpretable Deep Learning*. (2018). Available online at: https://arxiv.org/pdf/1810.10489.pdf
19. Russell S, Norvig P. *Artificial Intelligence: A Modern Approach*. New Jersey, NJ: Prentice Hall Press (2009). Available online at: https://www.cin.ufpe.br/~tfl2/artificial-intelligence-modern-approach.9780131038059.25368.pdf
20. Graves AJA, *Generating Sequences With Recurrent Neural Networks*. (2013). Available online at: https://arxiv.org/pdf/1308.0850.pdf
21. Murtagh F, Starck J-L, Renaud O. On neuro-wavelet modeling. *Decis Support Syst*. (2004) 37:475–84. doi: 10.1016/S0167-9236(03)00092-7
22. Hochreiter S, Schmidhuber JA. Long short-term memory. *Neural Comput*. (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735
23. Palangi H, Deng L, Shen Y, Gao J, He X, Chen J, et al. Deep sentence embedding using long short-term memory networks: analysis and application to information retrieval. *IEEE.* (2016) 24:694–707. doi: 10.1109/TASLP.2016.2520371
24. Kalisch M, Mächler M, Colombo D, Maathuis MH, Bühlmann P. Causal inference using graphical models with the R Package pcalg. *J Stat Software*. (2012) 47:26. doi: 10.18637/jss.v047.i11
25. Zhang J. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artif Intell*. (2008) 172:1873–96. doi: 10.1016/j.artint.2008.08.001
26. Driban JB, McAlindon TE, Amin M, Price LL, Eaton CB, Davis JE, et al. Risk factors can classify individuals who develop accelerated knee osteoarthritis: data from the osteoarthritis initiative. *J Orthop Res.* (2018) 36:876–80. doi: 10.1002/jor.23675
27. Sperandei S, Understanding logistic regression analysis. *Biochem Med (Zagreb).* (2014) 24:12–8. doi: 10.11613/BM.2014.003
28. Sharma L, Song J, Felson DT, Cahue S, Shamiyeh E, Dunlop DD. The role of knee alignment in disease progression and functional decline in knee osteoarthritis. *JAMA*. (2001) 286:188–95. doi: 10.1001/jama.286.2.188
29. Vincent HK, Heywood K, Connelly J, Hurley RW. Obesity and weight loss in the treatment and prevention of osteoarthritis. *PM R.* (2012) 4:S59–67. doi: 10.1016/j.pmrj.2012.01.005
30. Heidari B, Knee osteoarthritis prevalence, risk factors, pathogenesis features: Part I. *Caspian J Intern Med*. (2011) 2:205–12.

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 Public Health

ORIGINAL RESEARCH
published: 11 February 2021
doi: 10.3389/fpubh.2021.643191

Diagnosis of Patellofemoral Pain Syndrome Based on a Multi-Input Convolutional Neural Network With Data Augmentation

Wuxiang Shi[1,2], *Yurong Li*[2], *Baoping Xiong*[2,3]* *and Min Du*[1,2,4]*

[1] College of Physics and Information Engineering, Fuzhou University, Fuzhou, China, [2] Fujian Key Laboratory of Medical Instrumentation & Pharmaceutical Technology, Fuzhou University, Fuzhou, China, [3] Department of Mathematics and Physics, Fujian University of Technology, Fuzhou, China, [4] Fujian Provincial Key Laboratory of Eco-Industrial Green Technology, Wuyi University, Wuyishan, China

Patellofemoral pain syndrome (PFPS) is a common disease of the knee. Despite its high incidence rate, its specific cause remains unclear. The artificial neural network model can be used for computer-aided diagnosis. Traditional diagnostic methods usually only consider a single factor. However, PFPS involves different biomechanical characteristics of the lower limbs. Thus, multiple biomechanical characteristics must be considered in the neural network model. The data distribution between different characteristic dimensions is different. Thus, preprocessing is necessary to make the different characteristic dimensions comparable. However, a general rule to follow in the selection of biomechanical data preprocessing methods is lacking, and different preprocessing methods have their own advantages and disadvantages. Therefore, this paper proposes a multi-input convolutional neural network (MI-CNN) method that uses two input channels to mine the information of lower limb biomechanical data from two mainstream data preprocessing methods (standardization and normalization) to diagnose PFPS. Data were augmented by horizontally flipping the multi-dimensional time-series signal to prevent network overfitting and improve model accuracy. The proposed method was tested on the walking and running datasets of 41 subjects (26 patients with PFPS and 15 pain-free controls). Three joint angles of the lower limbs and surface electromyography signals of seven muscles around the knee joint were used as input. MI-CNN was used to automatically extract features to classify patients with PFPS and pain-free controls. Compared with the traditional single-input convolutional neural network (SI-CNN) model and previous methods, the proposed MI-CNN method achieved a higher detection sensitivity of 97.6%, a specificity of 76.0%, and an accuracy of 89.0% on the running dataset. The accuracy of SI-CNN in the running dataset was about 82.5%. The results prove that combining the appropriate neural network model and biomechanical analysis can establish an accurate, convenient, and real-time auxiliary diagnosis system for PFPS to prevent misdiagnosis.

Keywords: patellofemoral pain syndrome, convolutional neural network, data preprocessing, data augmentation, biomechanical analysis

OPEN ACCESS

Edited by:
Nianyin Zeng,
Xiamen University, China

Reviewed by:
Ni Youcong,
Fujian Normal University, China
Yong Xu,
Fujian Agriculture and Forestry University, China

***Correspondence:**
Baoping Xiong
xiongbp@fjut.edu.cn
Min Du
dm_dj90@163.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *17 December 2020*
Accepted: *19 January 2021*
Published: *11 February 2021*

Citation:
Shi W, Li Y, Xiong B and Du M (2021) Diagnosis of Patellofemoral Pain Syndrome Based on a Multi-Input Convolutional Neural Network With Data Augmentation. Front. Public Health 9:643191. doi: 10.3389/fpubh.2021.643191

INTRODUCTION

Patellofemoral pain syndrome (PFPS), also known as patellofemoral pain and chondromalacia patellae, often presents as a gradual onset of knee pain behind or around the patella (1–3). PFPS is a common chronic knee disease, especially among women and athletes (4, 5). It can cause pain in patients climbing up and down the stairs or squatting, thereby affecting their activities of daily living (6). According to the survey, PFPS may eventually evolve into patellofemoral osteoarthritis (7–9). If not treated in time, it can cause joint deformities and disability. Thus, the early and accurate diagnosis of PFPS is highly important.

Despite the high prevalence of PFPS, the etiology and gender differences of this disease remain unclear (10, 11). Two main difficulties are encountered in its diagnosis. One is the multifactorial etiology of PFPS (12, 13). It may be due to excessive extension of the knee joint, quadriceps weakness, valgus or varus of the knee joint, medial femoral muscle weakness, or gastrocnemius muscle tension. The other is the similarity of PFPS to many knee joint disease symptoms, such as bursitis, patellar tendinitis, and rheumatoid arthritis, causing misdiagnosis. The previous diagnosis of PFPS generally depends on the subjective judgment of doctors; thus, doctors should have very rich experience in patellar tracking, patellar apprehension, Waldron test, and squatting test (14, 15). However, the diagnosis results for the same patient may be inconsistent because of the different diagnostic criteria (14, 16).

The objective auxiliary diagnosis methods of PFPS include X-ray, magnetic resonance imaging, computed tomography, and arthroscopy (17–19). Among them, arthroscopy has the highest accuracy in diagnosing PFPS. However, arthroscopy is an invasive operation and requires a professional arthroscopy doctor (19). Magnetic resonance imaging has high diagnostic accuracy and non-invasiveness (20). However, its detection time is long, and some patients have claustrophobia, preventing them from actively cooperating with the examination. These imaging auxiliary diagnostic techniques require expensive equipment and professional doctors who are familiar with patellar abnormalities to correctly diagnose PFPS. Inexperienced personnel are prone to misdiagnosis, missed diagnosis, and other medical accidents. Subjective factors, such as the psychology and physiology of experts, can greatly reduce the diagnosis and medical effect, thus affecting the stability of the diagnosis.

In recent years, biomechanical research has been a hot spot in disease diagnosis, and PFPS is no exception (21, 22). Besier et al. used the lower limb joint angle and surface electromyography (sEMG) signals of 10 muscles around the knee joint as the input of the musculoskeletal model to explore the changes in muscle forces in patients with PFPS (23). Ferrari et al. discussed the diagnostic value of sEMG signals of the vastus medialis (VM) and the vastus lateralis (VL) for PFPS by an independent *t*-test (15). Briani et al. used linear regression models to diagnose PFPS through the time-domain and frequency-domain variables of sEMG and compared the results (24). However, the results of these traditional analysis methods are inaccurate, and experienced doctors are needed to select the classification features.

With the development of machine learning, the combination of machine learning and biomechanical analysis has become increasingly popular (25, 26). In recent years, machine learning algorithms have been improved and applied in various fields (27–29). Many studies have shown that it is also suitable for disease diagnosis (30, 31). The artificial neural network model is widely used in machine learning because of its good non-linear adaptive information processing ability. Wang et al. trained a deep neural network using electroencephalography to diagnose neonatal encephalopathy (32). Cho et al. used an artificial neural network model with a single hidden layer to distinguish normal and abnormal knee joints, thereby assisting in the treatment of unstable patella and anterior knee pain (33). These neural network models have shown good results in the diagnosis of various diseases. However, the selection of a suitable neural network model is a problem worth considering, and the result of the network model is often related to the method of data processing. A general rule for the selection of the preprocessing method for biomechanical data is currently lacking.

To solve the above problems, we propose an improved multi-input convolution neural network (MI-CNN) model to diagnose PFPS. Compared with the single-input convolutional neural network (SI-CNN), MI-CNN simultaneously extracts data information from two mainstream data preprocessing perspectives of normalization and standardization. Given that biomechanical time-series data are different from image data, MI-CNN adopts the 1D convolution kernel, that is, it only slides on the time axis. The model was tested on the walking and running datasets of 41 subjects (26 patients with PFPS and 15 pain-free controls). Meanwhile, data augmentation was performed in the training set to prevent model overfitting. Compared with SI-CNN and previous methods, MI-CNN achieved higher accuracy (89.0%) on the running dataset. This method can be used as a computer-aided diagnosis method to prevent doctors from misdiagnosing.

METHODS

Dataset

All experimental data in this paper were obtained from the database published by the website https://www.sciencedirect.com/science/article/pii/S0021929009000396?via%3Dihub. The database collected 10 types of the biomechanical characteristic of 41 subjects (26 patients with PFPS and 15 pain-free controls) during walking and running, including three joint angles and seven sEMG signals: hip flexion angle (HF), knee flexion angle (KF), ankle dorsiflexion angle (ADF), semimembranosus (SEB), rectus femoris (REF), VL, VM, biceps femoris (BIF), medial gastrocnemius (MG), and lateral gastrocnemius (LG). The sampling frequency of angle data was 60 Hz, and the sampling frequency of the sEMG signal was 2400 Hz. These conditions were set because the effective sEMG signal spectrum distribution is between 10 and 500 hz. Thus, the sampling frequency of the sEMG signal should be large enough to ensure the quality of the sampling signal. Each biomechanical characteristic contains 100 time-series values. The detailed gathering process of the whole dataset can be seen in the reference (23).

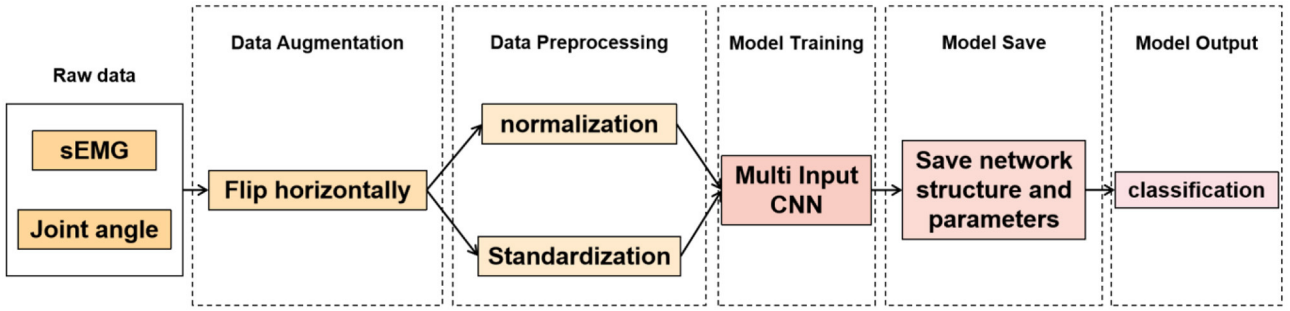


FIGURE 1 | Overall algorithm flow chart.

The overall algorithm flow is shown in **Figure 1**.

Data Augmentation

At present, a large number of experiments have proven that data size directly affects the performance of neural networks. PFPS involves many types of physiological signals of the lower limbs, but the number of samples in the dataset is relatively small, which easily leads to model overfitting. Data augmentation can prevent overfitting to some extent. Many methods of data augmentation for image data are available, such as rotation, horizontal flipping, vertical flipping, and random scaling. However, biomechanical data are different from image data. They are interrelated in the time dimension. Thus, many data augmentation methods are not applicable. We used the data of each subject to form a 100 $\times$ 10 2D matrix, with 100 rows representing time series values and 10 columns representing biomechanical characteristics. It has the same format as the image data to facilitate data augmentation. Hence, we can flip it horizontally because no strong correlation exists between these characteristics, thus doubling the training set.

Data Preprocessing

Before data input into the neural network, data preprocessing is an important link because it can accelerate the convergence speed of the neural network and improve the accuracy of the model. PFPS involves a variety of lower limb biological signals, and the ways to select these signals are different. Evaluating PFPS only based on a single index is usually insufficient. The problem from multiple indexes should be considered comprehensively. However, given their different nature, various evaluation indicators usually have different data scales. The level of each index differs greatly if the original data is directly used for analysis, highlighting the role of the index with a high numerical value in the comprehensive analysis and relatively weakening the role of the index with a low numerical level. Therefore, the original data must be preprocessed to ensure the reliability of the results. Different preprocessing methods have different effects on the evaluation results of the system. Unfortunately, a general rule to follow in the selection of data preprocessing methods is lacking. To improve data preprocessing, we need to view the distribution of data. Thus, we plotted the data distribution for one of the subjects, as shown in **Figure 2**, which from top to bottom are three joint angle values and seven sEMG signals: HF, KF, ADF, SEB, REF, VL, VM, BIF, MG, and LG.

Figure 2 shows that the range and distribution of the three joint angle values largely differ from those of the seven sEMG signals. Feeding such data into the neural network leads to poor results, and preprocessing is needed. The two most used pretreatment methods are standardization and normalization, which have their advantages and disadvantages.

Standardization can scale the data distribution of different characteristic dimensions to near 0, with the mean value of 0 and the variance of 1, which is comparable. The formula is as follows:

$$X_i = \frac{X_i - \overline{X}}{X_{std}}, \tag{1}$$

where $\overline{X}$ is the mean value of each column characteristic in the original data X, and X_{std} is the variance of each column characteristic in the original data X.

Normalization can limit the range of values of different characteristic dimensions within (0, 1), but it changes the distribution of the original data. The formula is as follows:

$$X_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, \tag{2}$$

where $X_{\min}$ is the minimum value of each column characteristic in the original data X, and $X_{\max}$ is the maximum value of each column characteristic in the original data X.

These two preprocessing methods can improve the convergence speed of the neural network but also have some shortcomings. The standardized results are related to each data point, and a specific scope limit is absent, causing the data to lack mean and variance information. The normalized result is mainly related to the maximum and minimum values but not much to the intermediate value. Moreover, the scaling range of normalization is mandatory and cannot be exceeded, causing the loss of some abnormal value information. Therefore, standardization and normalization were used for the data, and data features were extracted from these two aspects to maximize mining of data information.

In the process of data preprocessing, we preprocessed the training set and test set separately. More specifically, each column

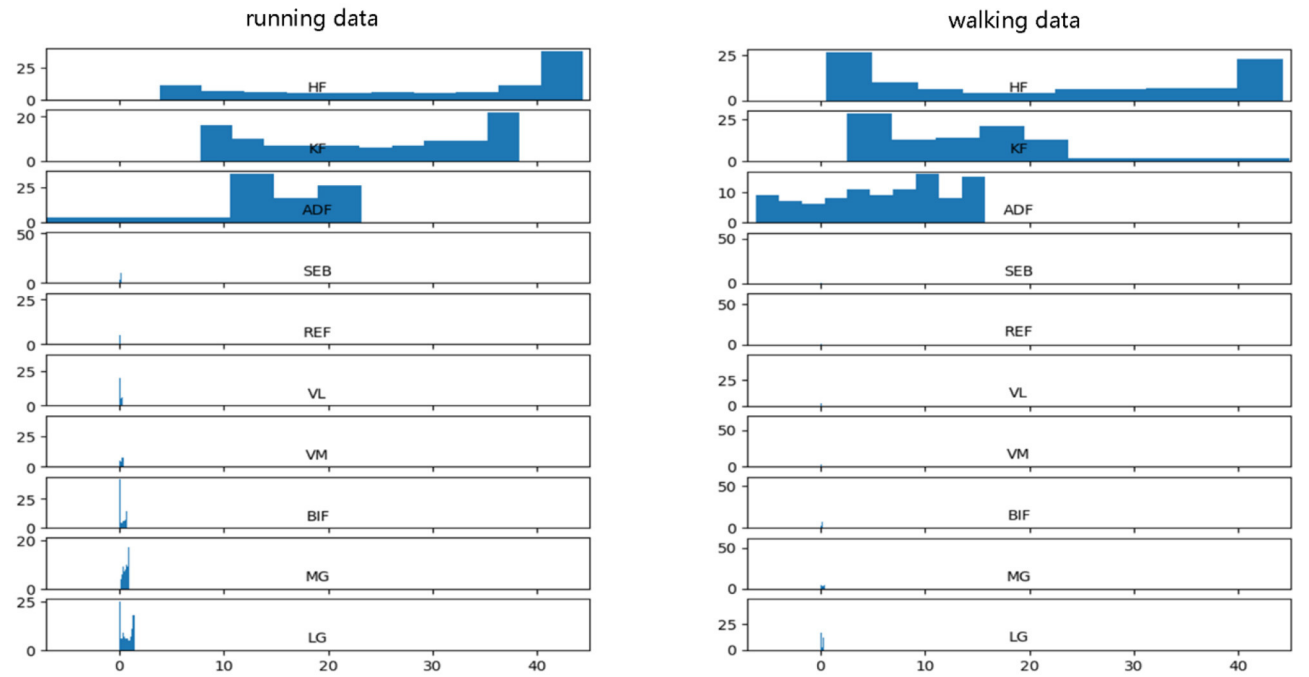


FIGURE 2 | Raw data distribution in the walking and running datasets.

feature of each subject in the dataset was standardized and normalized separately.

Multi-Input Convolution Neural Network

CNN has achieved excellent results in various classification and recognition tasks (34–36). The advantage of using CNN for time series classification is that it can learn directly from the original time series data without requiring domain experts to design input features manually. Thus, after data augmentation, MI-CNN with the 1D convolution kernel is proposed in this paper to extract the features of normalized and standardized data from two input channels simultaneously. Its model structure is shown in **Figure 3**.

All convolutional layers in the model have 16 filters with a convolution kernel size of 3 and a sliding step size of 1. A Relu activation function is added after each convolutional layer to perform a nonlinear mapping on the output of the convolutional layer. The calculation formula is as follows:

$$\boldsymbol{Relu}\,(\boldsymbol{X}) = \boldsymbol{max}\,(\boldsymbol{X}, \boldsymbol{0}) \tag{3}$$

Thus, if the data value transmitted to the neuron is <0, the value of the neuron will be changed to 0.

The size of the max-pooling layer is 2 and the sliding step is 1, which only keeps the maximum value in the window to reduce the complexity of the model and expand the receptive field. After the last convolutional layer, a dropout layer (rate = 0.3) is added to randomly ignore 30% of the neurons when training the network model. It can prevent the neural network from overfitting and reduce the training time. Then, the flattening layer will flatten the features extracted by the convolutional layer into a 1D vector and concatenate the output of the two convolutional channels by the fusion layer. At the end of the model is a fully connected neural network to interpret all the feature information extracted by the convolutional layer and map it to the category value. The first fully connected layer has 50 neurons and uses the Relu activation function. The output layer has two neurons, representing two categories. The activation function of the output layer is SoftMax, which maps the output into two types of probability values. The formula for the SoftMax function is as follows

$$\boldsymbol{y_{out}} = \boldsymbol{SoftMax}\,(\boldsymbol{Z_i}) = \frac{\boldsymbol{e^{Z_i}}}{\sum_{\boldsymbol{p=1}}^{\boldsymbol{2}} \boldsymbol{e^{Z_p}}},\ \ \boldsymbol{for\ p = 1, 2,} \tag{4}$$

where $\mathbf{Z_i}$ is the output value of the ith neuron in the output layer.

After setting the model structure, we use the Adam optimization algorithm to update the network parameters by backpropagation. The learning rate is set to 0.00001, the number of iterations is set to 4000, and the Cross-Entropy function is selected as the loss function.

RESULTS

Test Environment and Evaluation Index

Two datasets of 41 subjects in walking and running states were used as experimental data to verify the effectiveness of the proposed algorithm. The dataset is described in detail in Section 2.1. The neural network model in this paper was constructed using the Keras

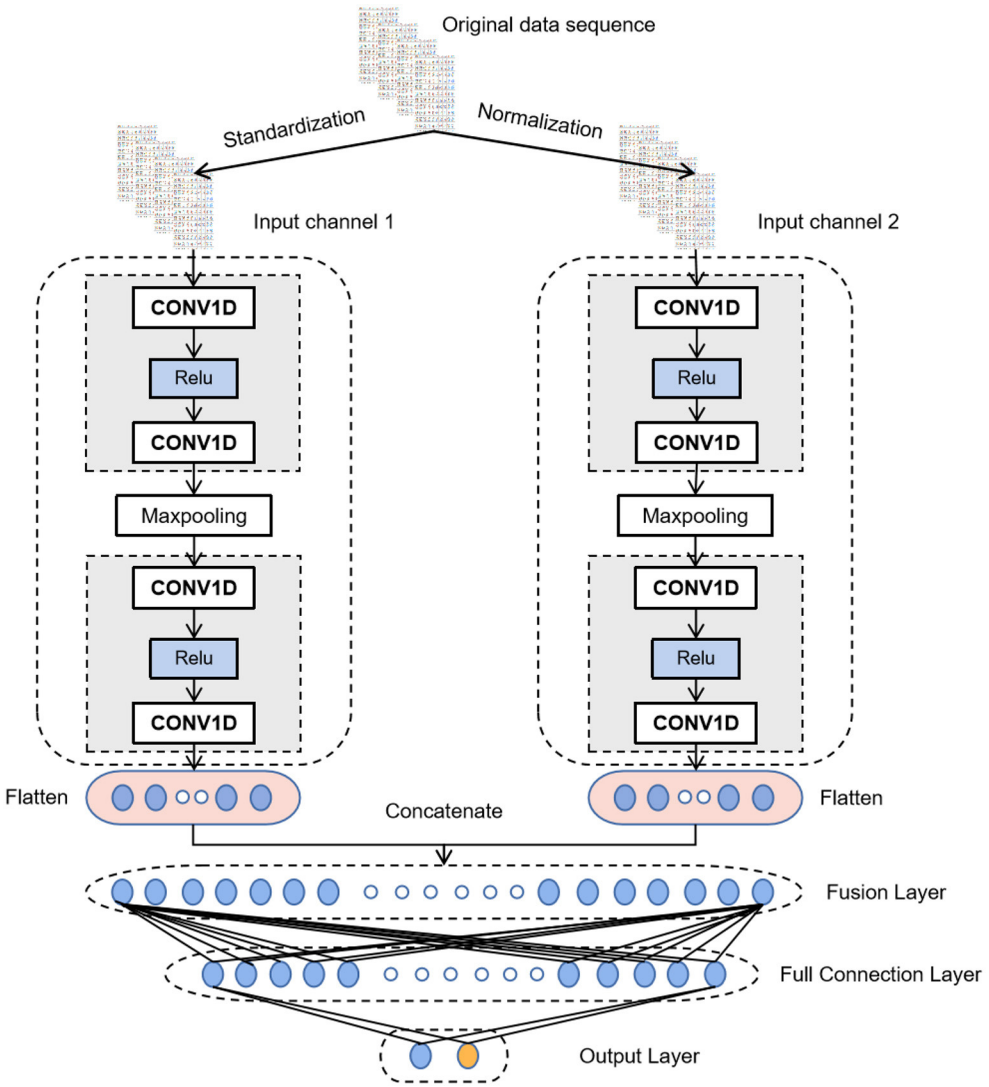


FIGURE 3 | Network structure of MI-CNN.

TABLE 1 | Configuration of test environment.

Parameters	Version or value
Operating system	Windows 10 (*64)
CPU	Intel Core i7-8700
GPU	GTX 1080
RAM	16.0 GB
Tensorflow	1.13.1
Keras	2.2.4
Python	3.7

framework based on Tensorflow. The configuration parameters of this experimental environment are shown in **Table 1**.

We randomly selected 70% of the dataset as the training set and the remaining 30% as the test set. The training and test sets were subjected to the same preprocessing, namely, normalization and standardization, and the data enhancement was only used on the training set. Meanwhile, given that our dataset is not large, the training batch size of the model was set to the whole training set to reduce the training time and improve the stability of training.

In this paper, accuracy, sensitivity, specificity, and training time were selected as the evaluation indexes of the test results. The definitions of these indexes are as follows

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN}, \tag{5}$$

$$sensitivity = \frac{TP}{TP + FN}, \tag{6}$$

$$specificity = \frac{TN}{TN + FP}, \tag{7}$$

where *TP* is the number that is correctly classified as PFPS, *TN* is the number that is correctly classified as pain-free controls, *FN* is the number that is wrongly classified as pain-free controls, and *FP* is the number that is wrongly classified as PFPS.

The best parameters of each model were selected by ten-fold cross-validation, which is equivalent to training 10 models and makes up for the disadvantage of a small amount of training data. Each experiment was repeated 10 times independently, and the average value was taken as the evaluation result.

We conducted four experiments on the running and walking datasets, respectively, including SI-CNN without any data preprocessing, SI-CNN with standardization processing, SI-CNN with normalization processing, and MI-CNN with standardization processing and normalization processing. Then, their results were compared.

Comparison of Test Results on the Walking Dataset

To clarify the comparison results, we randomly created the accuracy and loss curves of four data preprocessing methods in one of the experiments, as shown in **Figure 4**. Meanwhile, the average results of the 10 repeated tests on the running dataset are shown in **Table 2**.

Comparison of Test Results on the Running Dataset

In the same way, we randomly created the accuracy and loss curves of four data preprocessing methods in one of the

TABLE 2 | Results of neural networks with different data preprocessing methods on the walking dataset.

Algorithm	Accuracy	Sensitivity	Specificity	Training time (s)
SI-CNN (raw data)	0.415	0.548	0.2	28.5
SI-CNN (Normalization)	0.615	0.908	0.167	29.8
SI-CNN (Standardization)	0.639	0.824	0.32	29.8
MI-CNN	0.692	0.88	0.4	38.8

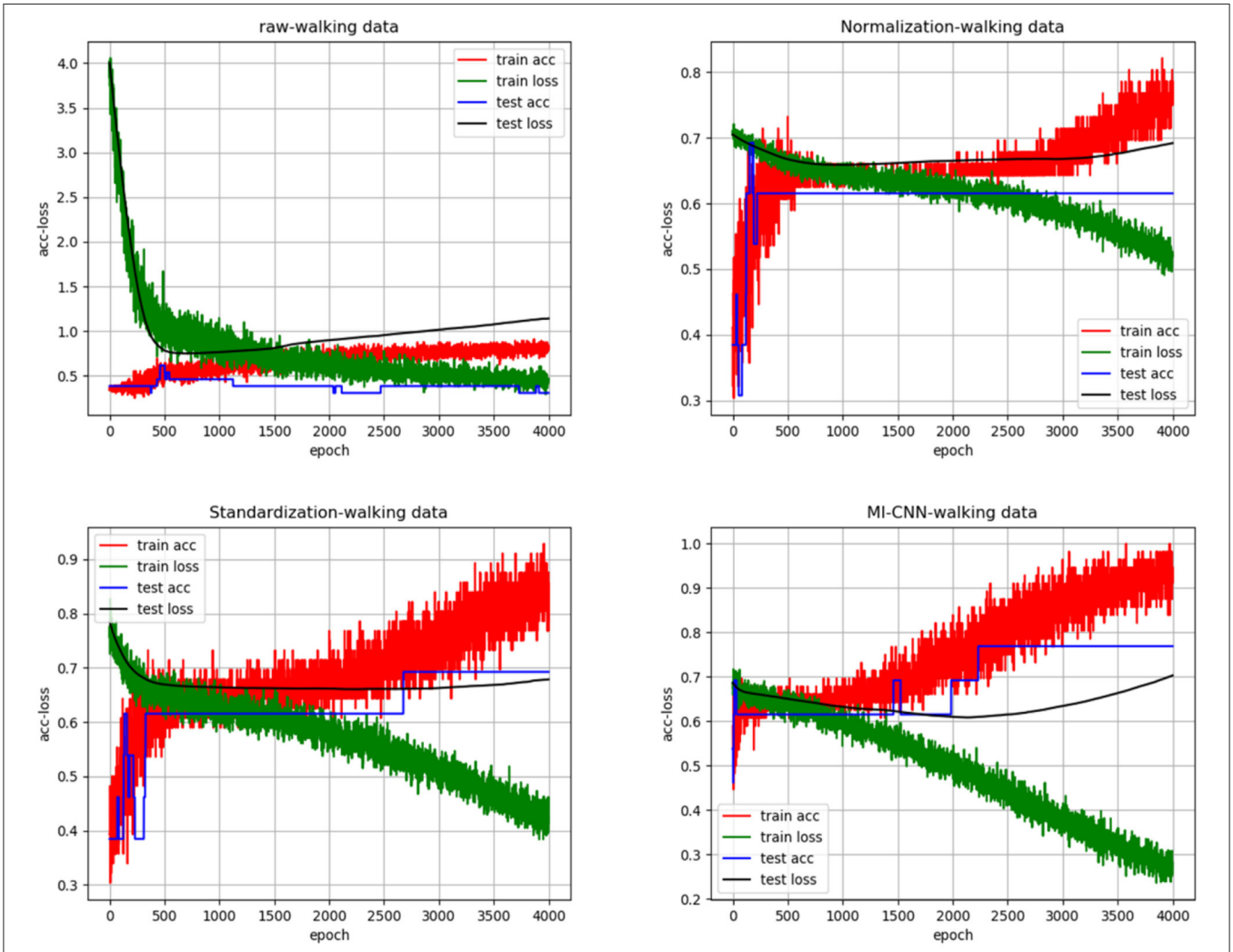


FIGURE 4 | "acc-loss" curves with different data preprocessing on the walking dataset.

experiments, as shown in **Figure 5**. The average results of the 10 repeated tests on the running dataset are shown in **Table 3**.

DISCUSSION

As shown in **Figures 4**, **5**, when the neural network model does not carry out any data preprocessing, the loss curve of its training set drops rapidly, but the accuracy of the test set is not effectively improved. This result can be ascribed to the considerably different data range and data distribution of the joint angle and sEMG. The decrease in the loss curve of the training set is mainly affected by the joint angle value. Network learning is very one-sided, leading to poor results. The loss and accuracy curves of the training set jitter because of the added dropout layer, which only acts on the training of the neural network.

Tables 2, **3** show that the results of the neural network after data standardization are slightly better than those after data normalization possibly because sEMG has no negative value after rectification, but the joint angles have some negative values, whereas the normalization completely limits the range of values, resulting in the loss of some outlier information. Thus, unlike image data, joint angle data are more suitable for standardized processing.

TABLE 3 | Results of neural networks with different data preprocessing methods on the running dataset.

Algorithm	Accuracy	Sensitivity	Specificity	Training time (s)
SI-CNN (raw data)	0.596	0.75	0.35	29.1
SI-CNN (Normalization)	0.781	0.868	0.63	29.2
SI-CNN (Standardization)	0.825	0.84	0.72	29.5
MI-CNN	0.89	0.976	0.76	38.5

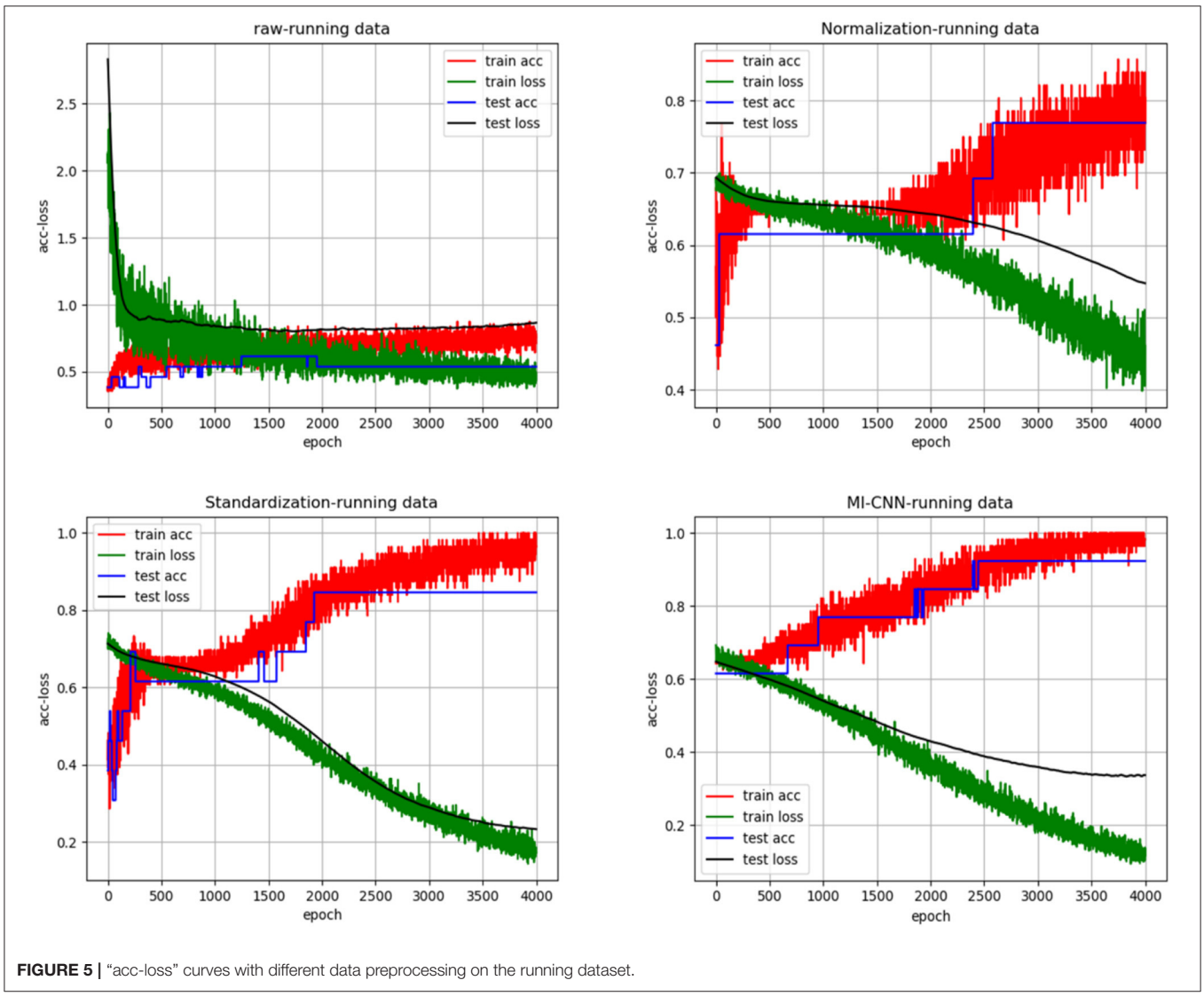


FIGURE 5 | "acc-loss" curves with different data preprocessing on the running dataset.

TABLE 4 | Comparison results with previous methods.

Methods	Sensitivity	Specificity
Ferrari' method on VM and VL	70%	87%
Briani' method on VM	72%	69%
Briani' method on VL	68%	62%
Squatting test	91%	50%
VM coordination test	16%	93%
MI-CNN on running data	97.6%	76.0%

Tables 2, **3** also show that the results of MI-CNN are better than those of traditional SI-CNN on the walking and running datasets. The accuracy rate of MI-CNN has increased by nearly 6%, but its training time is longer, because it has two input channels. Thus, the network parameters of MI-CNN are almost double that of SI-CNN. Given that this model is mainly used for the auxiliary diagnosis of diseases, accuracy is more important. Moreover, as long as the neural network model is saved after training, it can be used for real-time diagnosis.

Comparison of **Tables 2**, **3** shows that the convolutional neural network model does not perform well on the walking dataset, and the results on the running dataset are better, which indicates that the biomechanical data of patients with PFPS in walking state is not much different from that of the pain-free controls, but large differences exist in the running state.

Finally, the method proposed in this paper is compared with the previous methods. Ferrari et al. obtained 70% sensitivity and 87% specificity through the sEMG signals of VM and VL. Its dataset contains 51 subjects (22 patients with PFPS and 29 pain-free controls) (15). Briani et al. obtained 72% sensitivity and 69% specificity through the sMEG signals of VM and 68% sensitivity and 62% specificity through the sEMG signals of VL. Its dataset includes 59 subjects (31 patients with PFPS and 28 pain-free controls) (24). According to the survey (14), in the previous methods, the squatting test has the highest sensitivity (91%), but its specificity is only 50%. The VM coordination test has the highest specificity (93%), but its sensitivity is only 16%. To clarify the comparison results, we prepared **Table 4**. The MI-CNN method proposed in this paper has a sensitivity of 97.6% and a specificity of 76.0% on the running dataset, which is better than the previous methods in general.

CONCLUSION

PFPS is a common knee joint disease, but its specific etiology remains unclear. An accurate, convenient, and real-time PFPS detection system must be established for clinical auxiliary diagnosis. MI-CNN is proposed to diagnose PFPS. Compared with the musculoskeletal model, this model is more convenient and more versatile without considering the differences between subjects. Compared with the linear regression model, this model is more suitable for non-linear biomechanical data. Compared with the traditional 1D convolution neural network model, this model can fully mine data information from standardization and normalization at the same time to improve the accuracy of the model. The multi-dimensional biomechanical data are also augmented to prevent the neural network model from overfitting and further improve the accuracy of the model.

In sum, the biomechanical analysis technology based on real and objective gait data of patients can effectively reduce the influence of subjective factors and improve the stability of diagnosis and medical treatment. Combining it with the neural network model can make the biomechanical analysis more convenient and accurate.

This work is a preliminary study, and its applicability needs to be cautious. The next research work will focus on two aspects. One is to try to test the dataset with multiple gait diseases to obtain multi-classification models. The other is to continue to optimize the network structure to improve the accuracy of diagnosis.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

WS conceived the layout, the rationale, and the plan of this manuscript. WS wrote the first draft of the manuscript. BX, MD, and YL edited the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by Natural Science Foundation of Fujian Province (2020J01890).

ACKNOWLEDGMENTS

The authors are grateful to all the participants for proofreading this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.643191/full#supplementary-material

REFERENCES

1. Crossley KM, Callaghan MJ, Linschoten R van. Patellofemoral pain. *Br J Sports Med.* (2016) 50:247–50. doi: 10.1136/bjsports-2015-h3939rep
2. Smith BE, Selfe J, Thacker D, Hendrick P, Bateman M, Moffatt F, et al. Incidence and prevalence of patellofemoral pain: a systematic review and meta-analysis. Screen HR, editor. *PLoS ONE.* (2018) 13:e0190892. doi: 10.1371/journal.pone.0190892
3. Fagan V, Delahunt E. Patellofemoral pain syndrome: a review on the associated neuromuscular deficits and current treatment options. *Br J Sports Med.* (2008) 42:489–95. doi: 10.1136/bjsm.2008.046623
4. Taunton JE. A retrospective case-control analysis of 2002 running injuries. *Br J Sports Med.* (2002) 36:95–101. doi: 10.1136/bjsm.36.2.95
5. Csintalan RP, Schulz MM, Woo J, Mcmahon PJ, Lee TQ. Gender differences in patellofemoral joint biomechanics. *Clin Orthop Relat Res.* (2002) 402:260–9. doi: 10.1097/00003086-200209000-00026
6. Witvrouw E, Lysens R, Bellemans J, Cambier D, Vanderstraeten G. Intrinsic risk factors for the development of anterior knee pain in an athletic population. A two-year prospective study. *Am J Sports Med.* (2000) 28:480–9. doi: 10.1177/03635465000280040701
7. Thomas MJ, Wood L, Selfe J, Peat G. Anterior knee pain in younger adults as a precursor to subsequent patellofemoral osteoarthritis: a systematic review. *BMC Musculoskelet Disord.* (2010) 11:201. doi: 10.1186/1471-2474-11-201
8. Wyndow N, Collins N, Vicenzino B, Tucker K, Crossley K. Is There a Biomechanical Link Between Patellofemoral Pain and Osteoarthritis? A Narrative Review. *Sports Med.* (2016) 46:1797–808. doi: 10.1007/s40279-016-0545-6
9. Hinman RS, Lentzos J, Vicenzino B, Crossley KM. Is patellofemoral osteoarthritis common in middle-aged people with chronic patellofemoral pain? Patellofemoral OA in chronic patellofemoral knee pain. *Arthritis Care Res.* (2014) 66:1252–7. doi: 10.1002/acr.22274
10. Jensen R, Kvale A, Baerheim A. Is pain in patellofemoral pain syndrome neuropathic? *Clin J Pain.* (2008) 24:384–94. doi: 10.1097/AJP.0b013e3181658170
11. Witvrouw E, Callaghan MJ, Stefanik JJ, Noehren B, Bazett-Jones DM, Willson JD, et al. Patellofemoral pain: consensus statement from the 3rd International Patellofemoral Pain Research Retreat held in Vancouver, September 2013. *Br J Sports Med.* (2014) 48:411–4. doi: 10.1136/bjsports-2014-093450
12. Powers CM, Witvrouw E, Davis IS, Crossley KM. Evidence-based framework for a pathomechanical model of patellofemoral pain: 2017 patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester, UK: part 3. *Br J Sports Med.* (2017) 51:1713–23. doi: 10.1136/bjsports-2017-098717
13. Tramontano M, Pagnotta S, Lunghi C, Manzo C, Manzo F, Consolo S, et al. Assessment and management of somatic dysfunctions in patients with patellofemoral pain syndrome. *J Am Osteopath Assoc.* (2020) 120:165. doi: 10.7556/jaoa.2020.029
14. Nunes GS, Stapait EL, Kirsten MH, De Noronha M, Santos GM. Clinical test for diagnosis of patellofemoral pain syndrome: Systematic review with meta-analysis. *Phys Ther Sport.* (2013) 14:54–9. doi: 10.1016/j.ptsp.2012.11.003
15. Ferrari D, Kuriki HU, Silva CR, Alves N, Mícolis de Azevedo F. Diagnostic accuracy of the electromyography parameters associated with anterior knee pain in the diagnosis of patellofemoral pain syndrome. *Arch Phys Med Rehab.* (2014) 95:1521–6. doi: 10.1016/j.apmr.2014.03.028
16. Cook C, Mabry L, Reiman MP, Hegedus EJ. Best tests/clinical findings for screening and diagnosis of patellofemoral pain syndrome: a systematic review. *Physiotherapy.* (2012) 98:93–100. doi: 10.1016/j.physio.2011.09.001
17. Drew BT, Redmond AC, Smith TO, Penny F, Conaghan PG. Which patellofemoral joint imaging features are associated with patellofemoral pain? Systematic review and meta-analysis. *Osteoarthr Cartil.* (2016) 24:224–36. doi: 10.1016/j.joca.2015.09.004
18. Kettunen JA, Harilainen A, Sandelin J, Schlenzka D, Hietaniemi K, Seitsalo S, et al. Knee arthroscopy and exercise vs. exercise only for chronic patellofemoral pain syndrome: a randomized controlled trial. *BMC Med.* (2007) 5:38. doi: 10.1186/1741-7015-5-38
19. Denti M, Arosio A, Trevisan C. Comparison of "catheter" and conventional arthroscopy in the diagnosis of knee derangements. *Arthroscopy.* (1994) 10:614–7. doi: 10.1016/S0749-8063(05)80057-X
20. Zelenski N, Falk DP, D'Aquilla K, Borthakur A, Zgonis M. Zone- and layer-specific differences in proteoglycan content in patellofemoral pain syndrome are detectable on T1ρ MRI. *Skelet Radiol.* (2020) 49:1–6. doi: 10.1007/s00256-020-03418-8
21. Chen S, Chang W-D, Wu J-Y, Fong Y-C. Electromyographic analysis of hip and knee muscles during specific exercise movements in females with patellofemoral pain syndrome: An observational study. *Medicine.* (2018) 97:e11424. doi: 10.1097/MD.0000000000011424
22. Dag F, Dal U, Altinkaya Z, Erdogan AT, Ozdemir E, Yildirim DD, et al. Alterations in energy consumption and plantar pressure distribution during walking in young adults with patellofemoral pain syndrome. *Acta Orthop Traumatol Turc.* (2019) 53:50–5. doi: 10.1016/j.aott.2018.10.006
23. Besier TF, Fredericson M, Gold GE, Beaupré GS, Delp SL. Knee muscle forces during walking and running in patellofemoral pain patients and pain-free controls. *J Biomech.* (2009) 42:898–905. doi: 10.1016/j.jbiomech.2009.01.032
24. Briani RV, de Oliveira Silva D, Pazzinatto MF, de Albuquerque CE, Ferrari D, Aragão FA, et al. Comparison of frequency and time domain electromyography parameters in women with patellofemoral pain. *Clin Biomech.* (2015) 30:302–7. doi: 10.1016/j.clinbiomech.2014.12.014
25. Xiong B, Zeng N, Li Y, Du M, Huang M, Shi W, et al. Determining the online measurable input variables in human joint moment intelligent prediction based on the hill muscle model. *Sensors.* (2020) 20:1185. doi: 10.3390/s20041185
26. Keijsers NLW, Stolwijk NM, Louwerens JWK, Duysens J. Classification of forefoot pain based on plantar pressure measurements. *Clin Biomech.* (2013) 28:350–6. doi: 10.1016/j.clinbiomech.2013.01.012
27. Zeng N, Wang Z, Liu W, Zhang H, Hone K, Liu X. A dynamic neighborhood-based switching particle swarm optimization algorithm. *IEEE Trans Cybern.* (2020) 99:1–12. doi: 10.1109/TCYB.2020.3029748
28. Zeng N, Wang Z, Zhang H, Kim K-E, Li Y, Liu X. An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips. *IEEE Trans Nanotechnol.* (2019) 18:819–29. doi: 10.1109/TNANO.2019.2932271
29. Zeng N, Wang Z, Zineddin B, Li Y, Du M, Xiao L, et al. Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach. *IEEE Trans Med Imaging.* (2014) 33:1129–36. doi: 10.1109/TMI.2014.2305394
30. Zeng N, Qiu H, Wang Z, Liu W, Zhang H, Li Y. A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease. *Neurocomputing.* (2018) 320:195–202. doi: 10.1016/j.neucom.2018.09.001
31. Li L, Chen Y, Shen Z, Zhang X, Sang J, Ding Y, et al. Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. *Gastric Cancer.* (2020) 23:126–32. doi: 10.1007/s10120-019-00992-2
32. Wang J, Ju R, Chen Y, Liu G, Yi Z. Automated diagnosis of neonatal encephalopathy on aEEG using deep neural networks. *Neurocomputing.* (2020) 398:95–107. doi: 10.1016/j.neucom.2020.01.057
33. Cho K, Müller JH, Scheffer C, Erasmus PJ. Application of an artificial neural network for the quantitative classification of trochlear dysplasia. *J Mech Med Biol.* (2013) 13:1350059. doi: 10.1142/S0219519413500590
34. Demir F, Abdullah DA, Sengur A. A new deep CNN model for environmental sound classification. *IEEE Access.* (2020) 8:66529–37. doi: 10.1109/ACCESS.2020.2984903
35. Zeng N, Li H, Li Y, Luo X. Quantitative analysis of immunochromatographic strip based on convolutional neural network. *IEEE Access.* (2019) 7:16257–63. doi: 10.1109/ACCESS.2019.2893927
36. Vasan D, Alazab M, Wassan S, Safaei B, Zheng Q. Image-Based malware classification using ensemble of CNN architectures (IMCEC). *Comput Secur.* (2020) 92:101748. doi: 10.1016/j.cose.2020.101748

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: 05 March 2021
doi: 10.3389/fphys.2021.607356

Variations of Time Irreversibility of Heart Rate Variability Under Normobaric Hypoxic Exposure

Yang Li[1], Jianqing Li[1], Jian Liu[1], Yong Xue[2], Zhengtao Cao[2]* and Chengyu Liu[1]**

[1] School of Instrument Science and Engineering, Southeast University, Nanjing, China, [2] Air Force Medical Center, Beijing, China

In the field of biomedicine, time irreversibility is used to describe how imbalanced and asymmetric biological signals are. As an important feature of signals, the direction of time is always ignored. To find out the variation regularity of time irreversibility of heart rate variability (HRV) in the initial stage of hypoxic exposure, the present study implemented 2 h acute normobaric hypoxic exposure on six young subjects who have no plateau or hypoxia experiences; oxygen concentration was set as 12.9%. Electrocardiogram (ECG) signals were recorded in the whole process and RR interval sequences were extracted. Mathematical operations were executed to transform the difference of adjacent RR intervals into proportion and distance with delay time to conduct time irreversibility analysis of HRV. The same calculating method was implemented on six items randomly picked out from the MIT-BIH normal sinus rhythm database as a control group. Results show that variation of time irreversibility of HRV in a hypoxic environment is different from that in a normoxic environment, time irreversibility indices of a hypoxic group decreases continually at a delay time of 1 and 2, and indices curves of time irreversibility gradually tend to be steady and gather with each other at a delay time of 3 or 4. The control group shows no consistent tendency no matter what the delay time is in the range of 1–4. Our study indicates that in short-time hypoxic exposure, as hypoxic time goes by, regulation of the cardiovascular autonomic nervous system weakens; regulation times and intensity of sympathetic and parasympathetic nerves tend to be equal.

Keywords: time irreversibility, hypoxia, heart rate variability, delay time, autonomic nervous system

OPEN ACCESS

Edited by:
Nianyin Zeng,
Xiamen University, China

Reviewed by:
Lei Yu,
Shanxi University, China
Guohun Zhu,
The University of Queensland, Australia

****Correspondence:***
Jianqing Li
ljq@seu.edu.cn
Zhengtao Cao
czhengtao@126.com
Chengyu Liu
chengyu@seu.edu.cn

Specialty section:
This article was submitted to Autonomic Neuroscience, a section of the journal Frontiers in Physiology

Received: *17 September 2020*
Accepted: *15 February 2021*
Published: *05 March 2021*

Citation:
Li Y, Li J, Liu J, Xue Y, Cao Z and Liu C (2021) Variations of Time Irreversibility of Heart Rate Variability Under Normobaric Hypoxic Exposure. Front. Physiol. 12:607356. doi: 10.3389/fphys.2021.607356

INTRODUCTION

Time irreversibility is one of the basic features of nonlinear systems. It describes the process of how the physical state and statistical characteristics of a system depend on the direction of time (Yao et al., 2019). If a sequence has the same joint probability distribution with its reversal sequence, it is said to be time invertible. Invertible processes include Gaussian linear processes and isentropic processes. On the contrary, time irreversibility of series indicates the existence of nonlinear processes in the underlying dynamics, including non-Gaussian stochastic processes and dissipative chaos (Lacasa et al., 2012). The heartbeat of normal individuals is jointly controlled by multiple factors of sympathetic and parasympathetic nerves, and thus the time series of cardiac intervals shows characteristics of chaos, nonlinearity, and complexity (Rajbhandari Panday and Panday, 2018). Costa et al. (2005) pointed out that time irreversibility of cardiac intervals is influenced by age and pathological factors; healthy young people have significantly higher time irreversibility of cardiac intervals than that of healthy elderly people, and healthy aged people

have distinctly higher time irreversibility of cardiac intervals than patients with congestive heart failure (CHF) or atrial fibrillation. This conclusion was also supported by Hou et al. (2011) with multi-scale time irreversibility method, and she also indicated that there are significant differences of time irreversibility in healthy young people between day and night (Hou et al., 2012).

Many methods have been proposed for the measurement of time irreversibility of cardiac interval sequences. Differentiate operations between adjacent RR intervals is used by Porta et al. (2006) who took the occurrence frequency of adjacent intervals' difference greater than or less than zero as the time irreversibility index. And Costa et al. (2008) introduced the coarse-grained multi-scale method on Porta's work. Guzik et al. (2006) performed scaling operations on the square of RR intervals. Lacasa et al. (2012) traced the amplitude of each point in the sequence with a graph, calculated the number of points that each point could reach without crossing with other points, and constructed a new sequence as the calculating basis of time irreversibility. Ordinal mode has been applied to calculate the emergence frequency of various patterns (Graff et al., 2013). Hou et al. (2011) defined high-dimensional irreversibility calculation index on the work of Porta (Graff et al., 2013) and Guzik (Guzik et al., 2006) by introducing different delay times for irreversibility analysis, and the mean value of all calculation results is the value of high-dimensional irreversibility. It reflects the comprehensive influence of different levels' time delay of nerves and fluid in the human body on time irreversibility of heart rate variability (HRV). And research shows that both age and exercise have a great effect on high-dimensional irreversibility (Hou et al., 2011, 2013).

Hypoxia is one of the stimulating factors that make HRV change. When entering a hypoxic environment, blood oxygen saturation (SpO_2) of the human body decreases, and regulation of the autonomic nervous system changes, which is usually manifested as an accelerated heart rate and shortened breath (Xue et al., 2010; Liu et al., 2019). Lower oxygen concentration may even result in altitude sickness, with symptoms such as headache and nausea (Zhang et al., 2014). Different methods have been applied to assess how HRV changes when people are exposed to a hypoxic environment. Most of the research results found that the standard deviation (SDNN) and root mean square (RMSSD) between adjacent heartbeat intervals and low frequency (LF) and high frequency (HF) in the spectrum decrease while the ratio of LF to HF increases (Krejčí et al., 2018) and sample entropy declines (Zhang et al., 2014) at the same time. However, other researchers gave different results (Iwasaki et al., 2006; Vigo et al., 2010). Normobaric hypoxic chambers are often used in laboratories to simulate the plateau hypoxic environment. It creates hypoxic effects by filling the chamber with excess nitrogen to dilute the air, which is safe, economic, and easy to use.

Existing studies mostly compare the physiological state under normoxia and hypoxia and find out how physiological indicators change after people are exposed to a hypoxic environment. For example, variation of electrocardiogram (ECG) in hypoxic exposure was studied in Coustet et al. (2015); variation tendencies of HRV and SPO_2 within 10 min of acute hypoxic exposure were revealed in Krejčí et al. (2018). Some other studies explored people's hypoxic acclimatization speed or physical training methods through experiments (Beidleman et al., 2008; Zeng et al., 2013, 2019; Sinex and Chapman, 2015). As one of the methods to evaluate the health of the physiological system, time irreversibility analysis method has been widely used in medical research. It can reflect the pathological information of the cardiovascular system and aging status, however, this method has never been used in the study of human physiological variation in a hypoxic environment. Some symptoms that occur in hypoxia are similar to other diseases and there must be connections between time irreversibility and those symptoms. We need to conduct research over a continuous period of time at the beginning of hypoxia to study the subtle variation trends of physiological indicators. This research is of great importance for the prevention of acute altitude sickness, the physical training of pilots and astronauts in response to accidents of loss of pressure, and the revelation of dynamics and autonomic rhythm changes of the human cardiovascular nervous system in the early stage of hypoxic exposure. In the process of cardiac feedback control, delay happens at all levels of the physiological system due to the reaction speed of chemical transmitters (Casali et al., 2008; Alvarez-Ramirez et al., 2009), and some useful information that the data can express may be lost if time irreversibility is calculated only based on the difference between adjacent RR intervals. In order to explore variation laws of time irreversibility of HRV in the early stage of hypoxic exposure, experiments will be carried out in a normobaric hypoxic chamber. By introducing different delay times, time irreversibility variation rules of HRV will be studied in an acute 2 h hypoxic exposure. Few reports have been posted on the research of time irreversibility of HRV in a hypoxic environment.

MATERIALS AND METHODS

Subjects

Six young and healthy male volunteers were recruited for our experiment. Over 24 h stay at an altitude >3,000 m in the previous 6 months, born at an altitude >1,000 m, being a smoker, or having a history of severe respiratory or cardiopulmonary diseases are the exclusion criteria. All subjects were forbidden to consume tea, coffee, alcohol, medicine that could get the neural system excited, or do intense physical exercise 24 h prior to the experiment. Characteristics of the subjects are presented in **Table 1**. The study was approved by the Ethics Committee of

TABLE 1 | Characteristics of subjects.

Body mass (kg)	65.8 ± 6.0
Body height (cm)	173.3 ± 5.1
Body mass index (kg/m^2)	21.9 ± 1.3
Age (years old)	20.4 ± 1.5
Smoker	None
Prior altitude >3,000 m experience	None
History of severe respiratory or cardiopulmonary disease	None

Data are listed as means ± SD or as number.

Southeast University, China. All the subjects were well informed of the aim and risks of the experiment and they all signed their informed consent before the study.

Experimental Design and Data Acquisition

This experiment was implemented in a normobaric hypoxic chamber. Before the experiment, nitrogen was filled into the chamber to reduce the oxygen concentration to 12.9% (equivalent to the altitude of 3,600 m), then subjects entered the chamber. During the experiment, the nitrogen flow was controlled by an automatic valve (Yu et al., 2011), so that the oxygen concentration in the chamber was maintained at 12.9% ± 0.2%. The experiment lasted for 2 h, and subjects were asked to keep quiet and lie in their bunks. They could watch cellphones, read books, and listen to gentle music. ECG signals were recorded by tiny equipment inlaid in a chest belt for each subject (SensEcho-5A, Health Regulation Co. Ltd., Beijing, China). The sampling rate was 200 Hz, and digitized ECG signals were stored in a built-in hard disk, which is capable of storing 24 h of data. The temperature and humidity in the chamber were maintained at 22 ± 1°C and 25 ± 1%, respectively.

Data Processing and Calculating Method

First, we identified each QRS complex automatically by P&T methods (Pan and Tompkins, 1985) for its extensively tested accuracy and efficiency. False and missing detection were calibrated artificially. Then we extracted R-R intervals series by corrected R wave positions.

Porta et al. (2006) counted the number of times that the difference of one RR interval with the next was less than zero and when the number was unequal to zero; the two numbers' ratio was defined as time irreversibility index $P\%$. Guzik et al. (2006) performed arithmetical operations on the square of the difference between adjacent RR intervals and defined time irreversibility index $G\%$. We denote any element in RR intervals sequence as $RR_i (1 \leq i \leq n)$, where n is length of RR intervals sequence. The difference between adjacent RR intervals is shown as $\Delta RR = RR_i - RR_{i-1}$. ΔRR is also expressed as ΔRR^+ when $\Delta RR > 0$, and ΔRR^- when $\Delta RR < 0$. $N(\Delta RR^+)$ is the number of ΔRR^+ in the entire ΔRR sequence. Delay time τ is introduced in this study. For $\tau = 1, 2, \ldots, n-1$, $\Delta RR_\tau = x_{i+\tau} - x_i$ $(1 \leq i \leq n-\tau)$, time irreversibility indices $P\%(\tau)$ and $G\%(\tau)$ for each τ are given by

$$P\%(\tau) = \frac{N(\Delta RR_\tau^-)}{N(\Delta RR_\tau^-) + N(\Delta RR_\tau^+)} \times 100, \quad (1)$$

$$G\%(\tau) = \frac{\sum_{i=1}^{N(\Delta RR_\tau^+)} \Delta RR_\tau^+(i)^2}{\sum_{i=1}^{N(\Delta RR_\tau^+)} \Delta RR_\tau^+(i)^2 + \sum_{i=1}^{N(\Delta RR_\tau^-)} \Delta RR_\tau^-(i)^2} \times 100. \quad (2)$$

Distance between $P\%(\tau)$, $G\%(\tau)$ and 50 are expressed as

$$QP_\tau = |50 - P\%(\tau)|, \quad (3)$$

$$QG_\tau = |50 - G\%(\tau)|. \quad (4)$$

In order to make comparisons with the hypoxic group, we randomly selected six items of 2 h from the MIT-BIH Normal Sinus Rhythm Database (Goldberger et al., 2000) to perform the same time irreversibility analysis on ECG data. We will take QP_τ and QG_τ as the measurement of time irreversibility.

RESULTS

Since QP_τ and QG_τ are all statistics independent of the sequence length, we calculated each index value from the beginning of the experiment and moved forward by 1 min each time. The first computation scope is [0,5 min]. That is, we calculated the indexes in the scope of [0,5 min], [0,6 min], [0,7 min], . . ., [0,120 min], respectively, and drew line graphs. In this way, data volume for calculation is gradually increased and the curve is relatively flat. Interference for observation and analysis because of a short or sharp change in a certain data section can be avoided.

Variation of QP_1, QG_1, QP_2, and QG_2

If delay time was set to be 1 or 2 (as shown in **Figure 1**), indexes of all subjects in the hypoxic group fluctuated greatly in the first 15 min after the experiment began, which is possibly due to the data volume for calculation being small meaning the autonomic nervous system of the human body must undergo the process of adjustment and adaptation when first entering the hypoxic environment. In the entire experiment, the trends of QP_1, QG_1, QP_2, and QG_2 are generally toward zero, which indicates time irreversibility of HRV of the hypoxic group tends to decline.

Different from the hypoxic group, which showed a significant downward trend of QP_1, QG_1, QP_2, and QG_2, indicators of the MIT normoxic group did not show an obvious consistent changing direction (as shown in **Figure 2**).

If we calculate indicators every 15 min and take both $\tau = 1$ and $\tau = 2$ into consideration, we can draw an error bar chart with the average value of QP_1 and QP_2 (as shown in **Figure 3**), which clearly shows the average time irreversibility of all subjects with a delay time of 1 and 2 in different groups. As in the charts above, indicators of the hypoxic group decreases and the MIT normoxic group shows no clear differences.

Variation of QP_3, QG_3, QP_4, and QG_4

If delay time was set to be 3 or 4 (as shown in **Figure 4**), a great difference in variation trend happened in QP and QG from that when delay time is 1 or 2. Indicators no longer show a slow downward trend toward zero with time going by, but gradually tend to be stable after the drastic transition period of the first 30 min. And most of the subjects' indicator curves gradually converge to a narrow range. In contrast, there was no significant trend change in the MIT normoxic group (as shown in **Figure 5**).

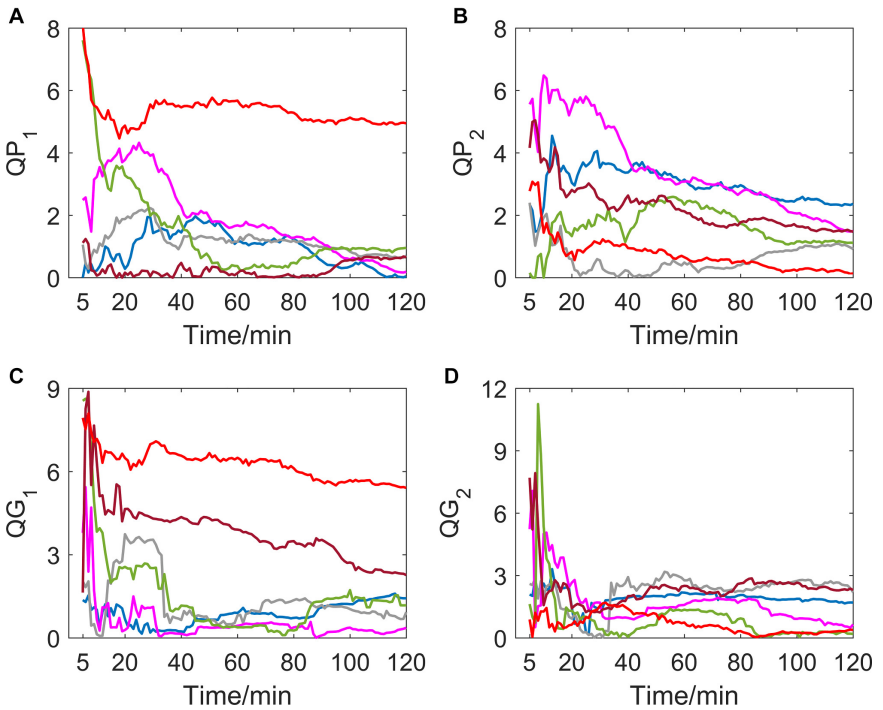


FIGURE 1 | (A–D) show the corresponding results of QP_1, QP_2, QG_1, and QG_2. Variation tendency of QP_1, QP_2, QG_1, and QG_2 within 2 h of hypoxic group (each color represents a subject, the same below).

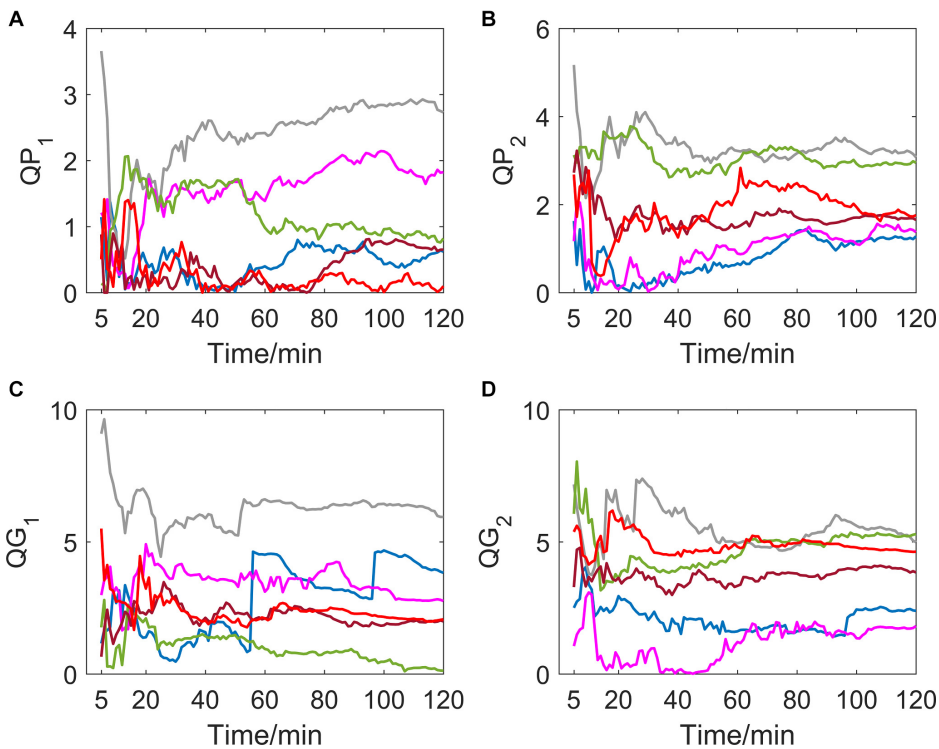


FIGURE 2 | (A–D) show the corresponding results of QP_1, QP_2, QG_1, and QG_2. Variation tendency of QP_1, QP_2, QG_1, and QG_2 within 2 h of MIT normoxic group.

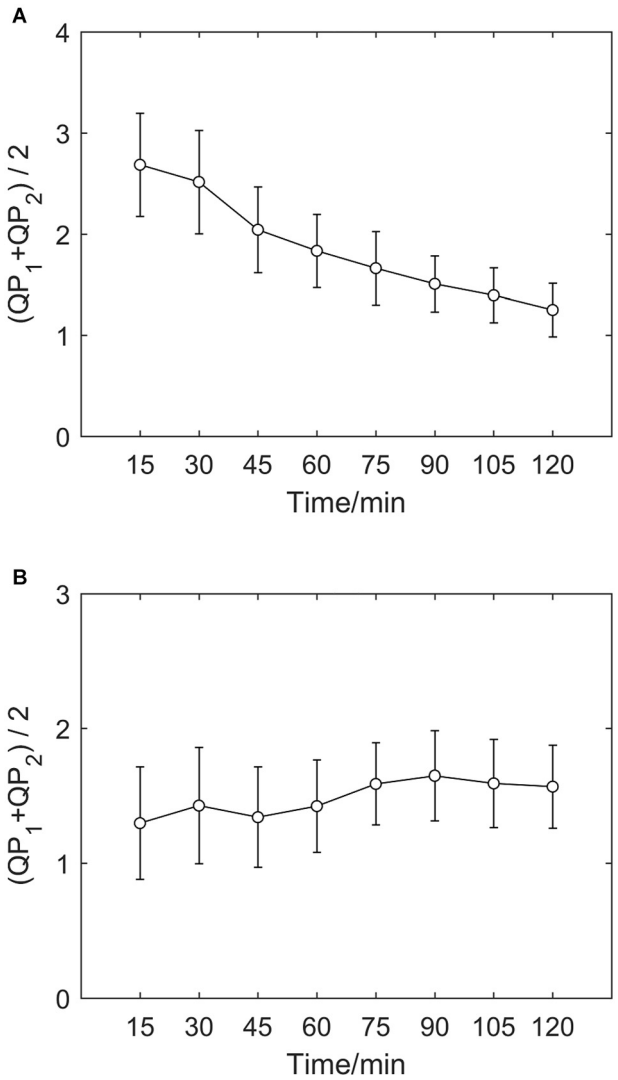


FIGURE 3 | Average value error bar chart of QP_1 and QP_2 of two groups. **(A)** Error bar chart of hypoxic group. **(B)** Error bar chart of MIT normoxic group.

DISCUSSION

Research on time irreversibility of HRV in a hypoxic environment has not been carried out before. Previous studies have shown that age and pathology will lead to a decrease of time irreversibility of HRV (Costa et al., 2005) and, compared to the daytime, time irreversibility of both healthy people and patients with congestive heart failure (CHF) declines at night (Porta et al., 2009). Our study in a hypoxic environment reveals a similar rule: prolonged hypoxic exposure leads to a sustained decline of time irreversible of HRV within 2 h when the delay time is 1 or 2.

Hypoxia is one of induction factors that can make regulatory mechanisms of cardiovascular nervous system change: parasympathetic activity was significantly reduced and autonomic nervous system regulation was greatly inhibited (Zhu et al., 2010). Since healthy physiological systems have the strongest nonlinearity and complexity, the decrease of time irreversibility in an anoxic state indicates a depression of nonlinearity and complexity.

In RR interval series, the positive difference between adjacent RR intervals represents a sudden increase of heart rate due to sympathetic modulation, whereas the negative difference represents a decrease in instantaneous heart rate due to parasympathetic modulation. The indicators variation of an MIT normoxic group reveals that there is always a certain difference of regulation times between sympathetic and parasympathetic nerves in normal individuals under normal oxygen concentration, while the difference tends to diminish in hypoxic environments. By definition, QP_1 reflects the absolute difference of regulation times between sympathetic and parasympathetic nerves, and QG_1 is a measurement of regulation intensity by calculating the square of RR intervals difference. The fact that both of them are approaching zero indicates that regulation times and intensity of sympathetic and parasympathetic nerves tend to be close to each other.

The LancetIn the present study, we observed a decrease in time irreversibility of the hypoxic group when the delay time was 1 and 2, but the variation trend changed when the delay time was increased to 3 or 4. It can be seen that the trend of QP_3, QG_3, QP_4, and QG_4 in the hypoxic group gradually becomes stable and converges with each other, while there is no significant difference whatever the delay time is of the MIT normoxic group, which indicates that time irreversibility of HRV is closely related to the delay time in a hypoxic environment. The results of mutual information analysis shows that the delay time of RR interval sequence in the elderly was significantly longer than that in the young (Ma and Zhang, 2010). Further research is needed to find out what generates the difference of delay time between hypoxic and normoxic environments. Hou et al. (2011) defined the average value at different delay times of QP and QG as the high dimensional time irreversibility index P_m, G_m, and the square root of their sum of squares was defined as their comprehensive measurement index D_m. Her study pointed out that the relationship between age, CHF, and time irreversibility of HRV was affected by the embedded dimension. When the value range of the embedded dimension changed, the variation rule revealed by time irreversibility indicators also changed. If the embedding dimension is m, the essential meaning of P_m and G_m is the average value of QP and QG when the delay time is taken from 1 to m. In the multi-level control loop and feedback network system of the human body, they can reflect the average effect of signals derived from different levels and different delays, but researchers cannot make observations on the specific value of time irreversibility index under a certain delay time. When the value of delay time increases continuously, variation of P_m and G_m must become gentler and have the characteristics of lag. It was found that time irreversibility of HRV during exercise was significantly increased when taking multiple dimensions into consideration (Hou et al., 2013). Time irreversibility of

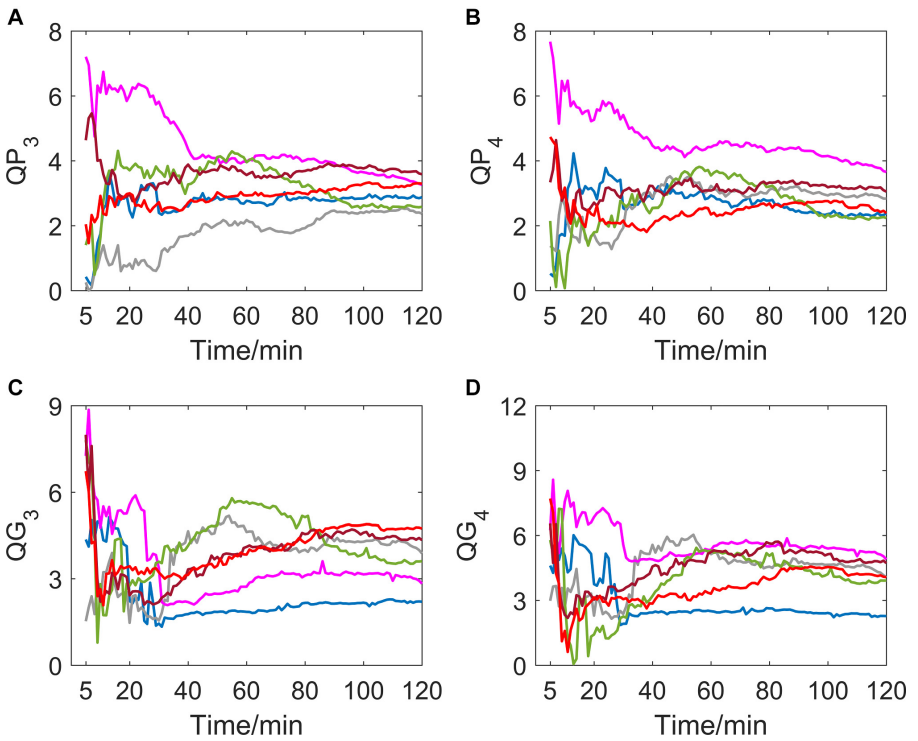


FIGURE 4 | (A–D) show the corresponding results of QP_3, QP_4, QG_3, and QG_4. Variation tendency of QP_3, QP_4, QG_3, and QG_4 within 2 h of hypoxic group.

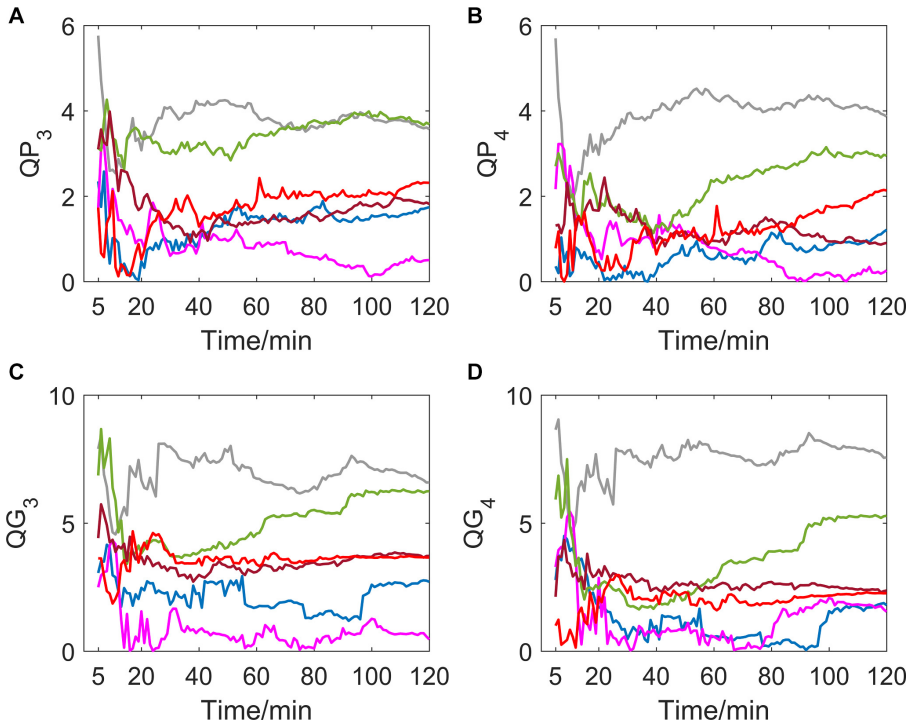


FIGURE 5 | (A–D) show the corresponding results of QP_3, QP_4, QG_3, and QG_4. Variation tendency of QP_3, QP_4, QG_3, and QG_4 within 2 h of MIT normoxic group.

HRV in a short period of time at the beginning of hypoxic exposure was investigated in this paper. Studies have shown that intermittent hypoxic training over several days significantly improves SDNN, RMSSD, LF HF (Xu et al., 2004), and SPO_2 (Fulco et al., 2013), however, the impact on time irreversibility needs further research.

CONCLUSION

Variation laws of time irreversibility of HRV in a normobaric hypoxic environment were researched in the present study. We found that in 2 h hypoxic exposure, time irreversibility indicators QP_1, QG_1, QP_2, and QG_2 decrease continuously, while QP_3, QG_3, QP_4, and QG_4 gradually tend to be stable and converge to a narrow range, which means there is a relationship between time irreversibility of HRV and delay time, and that proper selection of delay time is important for the observation of the relationship. In contrast, normoxic environments do not have any influence on time irreversibility of HRV whatever the delay time is. Results when delay time is 1 or 2 suggests a decline in automatic nervous system activities in a hypoxic environment, and regulation times and intensity of sympathetic and parasympathetic nerves get close to each other.

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 Southeast University, Nanjing, China.

AUTHOR CONTRIBUTIONS

YL did all the analysis work and wrote the article. JqL checked research protocol. JL performed the data calibration. YX conducted the experiment and collected data. ZC provided the experimental site and equipment. CL guided the research method, wrote the article and examined the research results. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program of China (2019YFE0113800), the National Natural Science Foundation of China (81871444 and 62071241), and the Distinguished Young Scholars of Jiangsu Province (BK20190014).

ACKNOWLEDGMENTS

We are grateful for the experimental site and equipment provided by the Aeromedicine Engineering Center in Air Force Medical Center.

REFERENCES

Alvarez-Ramirez, J., Rodriguez, E., and Echeverría, J. C. (2009). Delays in the human heartbeat dynamics. *Chaos* 19:028502. doi: 10.1063/1.3152005

Beidleman, B. A., Muza, S. R., Fulco, C. S., Cymerman, A., Sawka, M. N., Lewis, S. F., et al. (2008). Seven intermittent exposures to altitude improves exercise performance at 4300 m. *Med. Sci. Sports Exerc.* 40, 141–148. doi: 10.1249/mss.0b013e31815a519b

Casali, K. R., Casali, A. G., Montano, N., Irigoyen, M. C., Macagnan, F., Guzzetti, S., et al. (2008). Multiple testing strategy for the detection of temporal irreversibility in stationary time series. *Phys Rev. E Stat. Nonlin. Soft. Matter. Phys.* 77(6 Pt 2):66204. doi: 10.1103/PhysRevE.77.066204

Costa, M. D., Goldberger, A. L., and Peng, C. K. (2005). Broken asymmetry of the human heartbeat: loss of time irreversibility in aging and disease. *J. Electrocardiol.* 38(Suppl. 4):141. doi: 10.1016/j.jelectrocard.2005.06.076

Costa, M. D., Peng, C., and Goldberger, A. L. (2008). Multiscale analysis of heart rate dynamics: entropy and time irreversibility measures. *Cardiovasc. Eng.* 8, 88–93. doi: 10.1007/s10558-007-9049-1

Coustet, B., Lhuissier, F. J., Vincent, R., and Richalet, J. (2015). Electrocardiographic changes during exercise in acute hypoxia and susceptibility to severe high-altitude illnesses. *Circulation* 131, 786–794. doi: 10.1161/CIRCULATIONAHA.114.013144

Fulco, C. S., Beidleman, B. A., and Muza, S. R. (2013). Effectiveness of preacclimatization strategies for high-altitude exposure. *Exerc. Sports Sci. Rev.* 41, 55–63. doi: 10.1097/JES.0b013e31825eaa33

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., et al. (2000). PhysioBank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation [Online]* 101, 215–220.

Graff, G., Graff, B., Kaczkowska, A., Makowiec, D., Amigó, J. M., Piskorski, J., et al. (2013). Ordinal pattern statistics for the assessment of heart rate variability. *Eur. Phys. J. Spec. Top.* 222, 525–534. doi: 10.1140/epjst/e2013-01857-4

Guzik, P., Piskorski, J., Krauze, T., Wykretowicz, A., and Wysocki, H. (2006). Heart rate asymmetry by Poincaré plots of RR intervals. *Biomed. Tech. Biomed. Eng.* 51, 272–275. doi: 10.1515/BMT.2006.054

Hou, F. Z., Huang, X. L., Zhuang, J. J., Huo, C. Y., and Ning, X. B. (2012). Multi-scale strategy and data surrogating test: two elements for the detection of time irreversibility in heart rate variability. *Acta Phys. Sin.* 61:220507. doi: 10.7498/aps.61.220507

Hou, F. Z., Ning, X. B., and Huang, X. L. (2013). High-dimensional time irreversibility analysis of human heart rate variability before and after exercise. *J. Data Acquis. Process.* 28, 638–642.

Hou, F. Z., Ning, X. B., Zhuang, J. J., Huang, X. L., Fu, M. J., and Bian, C. H. (2011). High-dimensional time irreversibility analysis of human interbeat intervals. *Med. Eng. Phys.* 33, 633–637. doi: 10.1016/j.medengphy.2011.01.002

Iwasaki, K., Ogawa, Y., Aoki, K., Saitoh, T., Otsubo, A., and Shibata, S. (2006). Cardiovascular regulation response to hypoxia during stepwise decreases from 21% to 15% inhaled oxygen. *Aviat. Space Environ. Med.* 77, 1015–1019.

Krejčí, J., Botek, M., and McKune, A. J. (2018). Dynamics of the heart rate variability and oxygen saturation response to acute normobaric hypoxia within the first 10 min of exposure. *Clin. Physiol. Funct. Imaging* 38, 56–62. doi: 10.1111/cpf.12381

Lacasa, L., Nuñez, A., Roldán, É, Parrondo, J. M. R., and Luque, B. (2012). Time series irreversibility: a visibility graph approach. *Eur. Phys. J. B* 85, 1–11. doi: 10.1140/epjb/e2012-20809-8

Liu, C. Y., Yang, M. C., Di, J. N., Xing, Y. T., Li, Y. W., and Li, J. Q. (2019). Wearable ECG: history, key technologies and future challenges. *Chin. J. Biomed. Eng.* 38, 641–652.

Ma, Y. X., and Zhang, D. Z. (2010). The study on delay time of cardiac RR intervals series. *Chin. J. Med. Phys.* 27, 2135–2137.

Pan, J., and Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* 32, 230–236. doi: 10.1109/TBME.1985.325532

Porta, A., D'Addio, G., Bassani, T., Maestri, R., and Pinna, G. D. (2009). Assessment of cardiovascular regulation through irreversibility analysis of heart period variability: a 24 hours Holter study in healthy and chronic heart failure populations. *Philos. Trans. A Math. Phys. Eng. Sci.* 367, 1359–1375. doi: 10.1098/rsta.2008.0265

Porta, A., Guzzetti, S., Montano, N., Gnecchi-Ruscone, T., Furlan, R., and Malliani, A. (2006). "Time reversibility in short-term heart period variability," in *Proceedings of the 2006 Computers in Cardiology* (Valencia: IEEE), 77–80.

Rajbhandari Panday, K., and Panday, D. R. (2018). Heart rate variability (HRV). *J. Clin. Exp. Cardiol.* 09, 1–12. doi: 10.4172/2155-9880.1000583

Sinex, J. A., and Chapman, R. F. (2015). Hypoxic training methods for improving endurance exercise performance. *J. Sport Health Sci.* 4, 325–332. doi: 10.1016/j.jshs.2015.07.005

Vigo, D. E., Perez, L. S., Videla, A. J., Perez, C. D., Hunicken, H. M., Mercuri, J., et al. (2010). Heart rate nonlinear dynamics during sudden hypoxia at 8230 m simulated altitude. *Wilderness Environ. Med.* 21, 4–10. doi: 10.1016/j.wem.2009.12.022

Xu, X., Lu, L. L., Chen, Z. H., Liu, X. X., and Li, Y. Z. (2004). Analysis of heart rate variability during acute exposure to hypoxia. *Space Med. Med. Eng.* 17, 334–339.

Xue, Y., Yang, J., Feng, Y. T., Zhou, Y. B., Qin, Y. F., Li, Y., et al. (2010). Effects of mindful breathing on rapid hypoxia preacclimatization training. *J. Med. Imaging Health Inform.* 10, 718–723. doi: 10.1166/jmihi.2020.2923

Yao, W., Yao, W., and Wang, J. (2019). Equal heartbeat intervals and their effects on the nonlinearity of permutation-based time irreversibility in heart rate. *Phys. Lett. A* 383, 1764–1771. doi: 10.1016/j.physleta.2019.03.002

Yu, M. S., Yang, J., Wang, H. T., Miao, Y., Wang, N. Z., and Li, Y. (2011). *Normobaric Hypoxic Cabin for Intermittent Hypoxic Training and Method for Manipulate Oxygen Concentration.* CN Patent No 102198321 B. Beijing: State Intellectual Property Office of the People's Republic of China.

Zeng, N., Wang, Z., Li, Y., and Du, M. (2013). Time series modeling of nano-gold immunochromatographic assay via expectation maximization algorithm. *IEEE Trans. Biomed. Eng.* 60, 3418–3424. doi: 10.1109/TBME.2013.2260160

Zeng, N., Wang, Z., Zhang, H., Kim, K.-E., Li, Y., and Liu, X. (2019). An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips. *IEEE Trans. Nanotechnol.* 18, 819–829. doi: 10.1109/TNANO.2019.2932271

Zhang, D., She, J., Zhang, Z., and Yu, M. (2014). Effects of acute hypoxia on heart rate variability, sample entropy and cardiorespiratory phase synchronization. *Biomed. Eng. Online* 13:73. doi: 10.1186/1475-925X-13-73

Zhu, S., Ding, N. W., and Liu, L. (2010). Research progress on the influence of altitude and hypoxic environment on heart rate variability. *Chin. J. Sports Med.* 29, 359–362.

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 Medical Technology

TECHNOLOGY AND CODE
published: 09 March 2021
doi: 10.3389/fmedt.2021.619280

A Psychometric Platform to Collect Somatosensory Sensations for Neuroprosthetic Use

Giacomo Valle[1,2†], Francesco Iberite[1†], Ivo Strauss[1], Edoardo D'Anna[3], Giuseppe Granata[4], Riccardo Di Iorio[4], Thomas Stieglitz[5], Stanisa Raspopovic[2], Francesco M. Petrini[2], Paolo M. Rossini[4] and Silvestro Micera[1,3*]*

[1] The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy, [2] Department of Health Sciences and Technology, Institute for Robotics and Intelligent Systems, ETH Zürich, Zurich, Switzerland, [3] Bertarelli Foundation Chair in Translational Neuroengineering, Centre for Neuroprosthetics, School of Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Institute of Bioengineering, Lausanne, Switzerland, [4] Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS)-Policlinic A. Gemelli Foundation, Institute of Neurology, Catholic University of the Sacred Heart, Rome, Italy, [5] Laboratory for Biomedical Microtechnology, Department of Microsystems Engineering–IMTEK, Bernstein Center, BrainLinks-BrainTools Cluster of Excellence, University of Freiburg, Freiburg, Germany

OPEN ACCESS

Edited by:
Yonghong Peng,
Manchester Metropolitan University,
United Kingdom

Reviewed by:
Ning Lan,
Shanghai Jiao Tong University, China
Xia-an Bi,
Hunan Normal University, China

***Correspondence:**
Giacomo Valle
giacomo.valle@hest.ethz.ch
Silvestro Micera
silvestro.micera@epfl.ch

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Medtech Data Analytics, a section of the journal Frontiers in Medical Technology

Received: 19 October 2020
Accepted: 08 February 2021
Published: 09 March 2021

Citation:
Valle G, Iberite F, Strauss I, D'Anna E, Granata G, Di Iorio R, Stieglitz T, Raspopovic S, Petrini FM, Rossini PM and Micera S (2021) A Psychometric Platform to Collect Somatosensory Sensations for Neuroprosthetic Use. Front. Med. Technol. 3:619280. doi: 10.3389/fmedt.2021.619280

Somatosensory neuroprostheses exploit invasive and non-invasive feedback technologies to restore sensorimotor functions lost to disease or trauma. These devices use electrical stimulation to communicate sensory information to the brain. A sensation characterization procedure is thus necessary to determine the appropriate stimulation parameters and to establish a clear personalized map of the sensations that can be restored. Several questionnaires have been described in the literature to collect the quality, type, location, and intensity of the evoked sensations, but there is still no standard psychometric platform. Here, we propose a new psychometric system containing previously validated questionnaires on evoked sensations, which can be applied to any kind of somatosensory neuroprosthesis. The platform collects stimulation parameters used to elicit sensations and records subjects' percepts in terms of sensation location, type, quality, perceptual threshold, and intensity. It further collects data using standardized assessment questionnaires and scales, performs measurements over time, and collects phantom limb pain syndrome data. The psychometric platform is user-friendly and provides clinicians with all the information needed to assess the sensory feedback. The psychometric platform was validated with three trans-radial amputees. The platform was used to assess intraneural sensory feedback provided through implanted peripheral nerve interfaces. The proposed platform could act as a new standardized assessment toolbox to homogenize the reporting of results obtained with different technologies in the field of somatosensory neuroprosthetics.

Keywords: neuroprosthesis, neurostimulation, electrodes, sensory feedback, amputees, psychophysics, somatosensations, platform

INTRODUCTION

Somatosensory neuroprostheses are highly innovative devices (1). Several research groups have investigated the ability to restore sensory feedback in patients with upper or lower limb amputation, tetraplegia, or paraplegia using invasive (2–14) and non-invasive (15–19) interfaces with the peripheral (PNS) and the central nervous systems (CNS) (**Figure 1**). The main aim of these

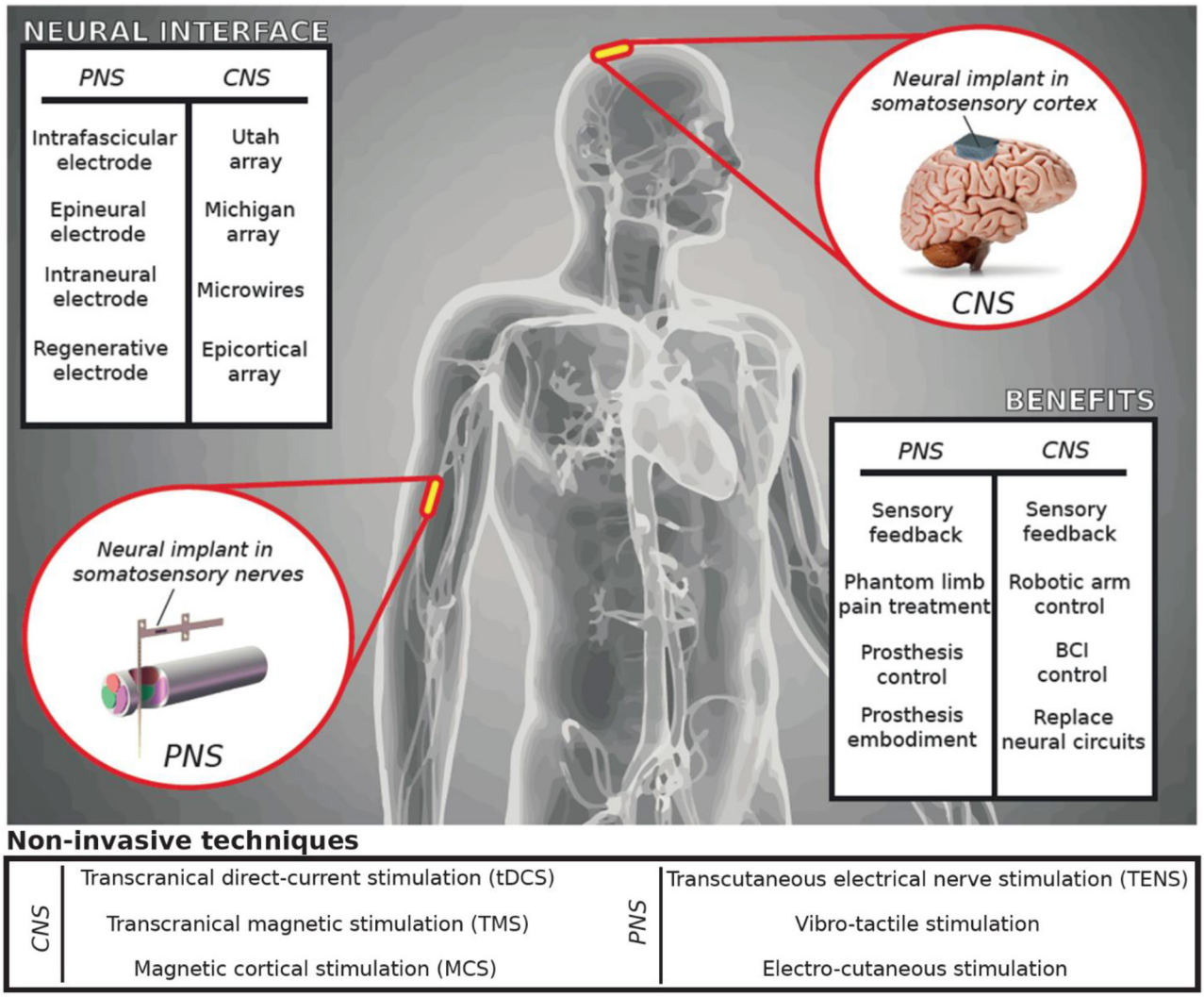


FIGURE 1 | Neuroprosthetic applications. Neurotechnologies for restoring somatosensations have been developed for peripheral (PNS) or central nervous systems (CNS). The stimulation technique used to restore sensory feedback can be invasive (surgically implanted and in intimate contact with the nervous tissue) or non-invasive (applied on the skin surface). Delivering a stimulation to the brain or peripheral nerves provides benefits such as the control of robotics, smart prosthetics, or other assistive technologies.

technologies is to elicit somatotopic-referred sensations emanating from the affected limb, creating a personalized map of the these sensations which could be used as a sensory feedback aimed at improving the patients' quality of life (20, 21). All these approaches use neural stimulation to evoke sensations stemming from contact with sensory peripheral nerves or the neural interfaces are placed directly on the somatosensory cortex.

The external stimulation of the residual (still functional proximal to the lesion) nervous structures guarantees to evoke an artificial sensation that can also be controlled by modulating the stimulation parameters (22–24).

Since there is intersubject variability due to the different nerve structures, implantation levels, and innervation (25–27), together with the subjective perception of the elicited sensations, a "sensation characterization" procedure is necessary to obtain a uniform sensation mapping (**Figure 2**). The goal of this procedure is to collect all the stimulation parameters corresponding to the evoked sensations characterized by the intensity, quality, location, and type in order to have a personalized sensation map. The mapping phase is crucial to implement an effective real-time assistive system, e.g., bidirectional hand or leg prostheses, eliciting homologous referred sensations emanating from the phantom limb (somatotopic) for therapeutic or functional purposes. In fact, the personalized sensation map is often translated into a robotic arm or hand in order to elicit sensations during object manipulation tasks aimed at increasing patient motor control performance (4, 6, 28, 29). When the patient is controlling a robotic arm, and touches a surface with the second robotic digit, the sensation perceived should be in the same location

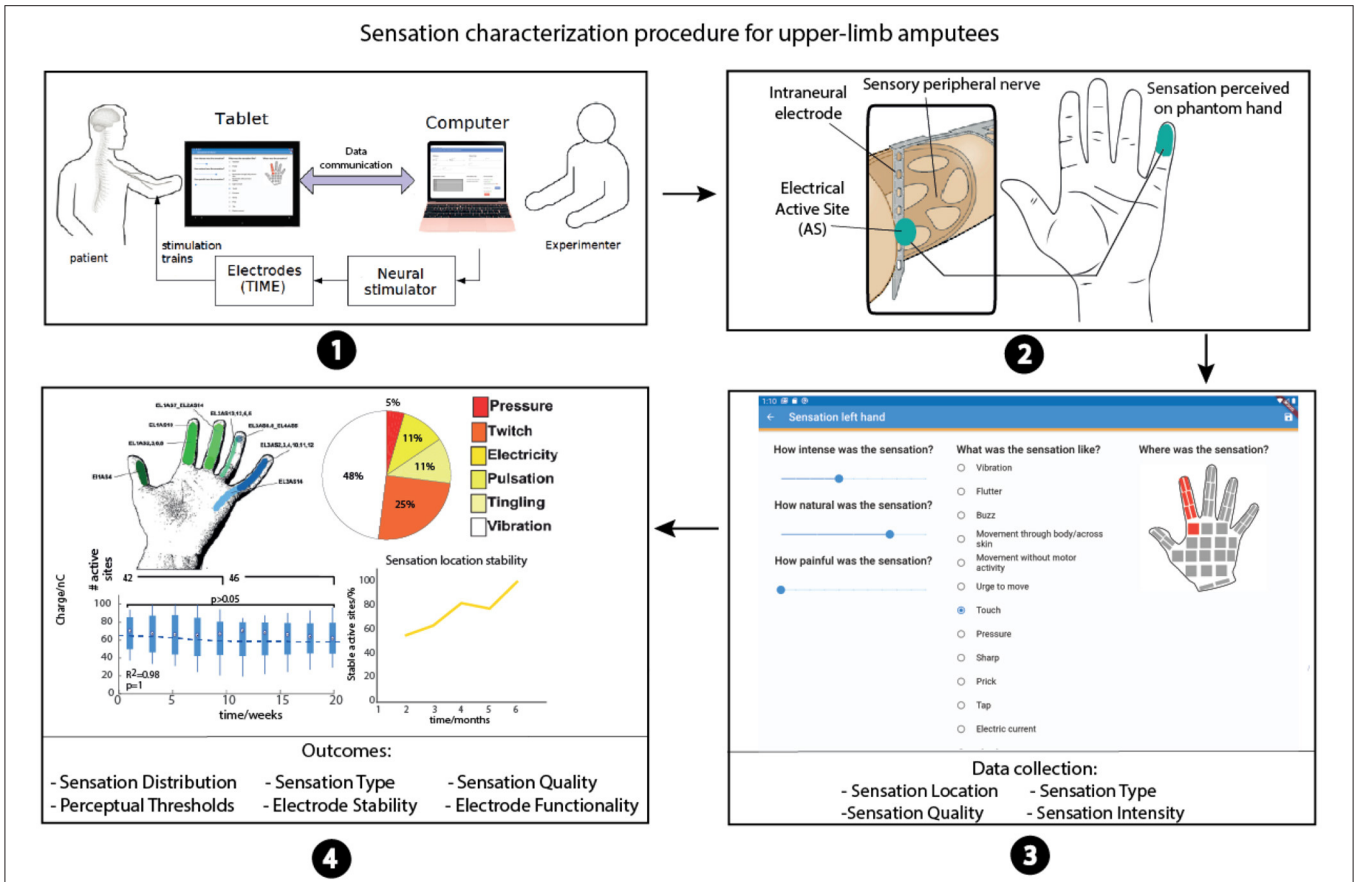


FIGURE 2 | Sensation characterization procedure. (1) Stimulation parameters are selected. The stimulation trains are delivered using the neurostimulator; the software also sends control commands to the Easy Quest app. (2) Patient perceives a stimulation-evoked sensation on the phantom hand thanks to the neural implant. (3) Easy Quest app in ODF mode is used to report the sensations. (4) Experimenters collect all sensation characterization outcomes and import them in MATLAB or Excel to plot the results.

(index), with the safety and exact intensity (mapped with the pressure force of the robotic finger) and the type should be in line with finger pressure (i.e., no electricity or warmness). The personalized sensation map should thus be as detailed as possible.

In addition, to provide a rich and reliable artificial sensory feedback to the patient, it is necessary to accurately test multiple sets of stimulation parameters and re-test them even over multiple days during long-term applications (14, 30–33). This characterization procedure is fundamental to re-create an artificial sensory feedback that could be effectively exploitable in neuroprosthetic applications by the users. To this aim, it is crucial to collect the data in an effective and efficient way using a platform in which the information are exhaustive and standardized. This would guarantee an easy comparison of the results with different technology or algorithm.

Several psychometric questionnaires exist regarding the quality and type of the sensations evoked (15, 34–36). However, they do not appear to be easy-to-use or fast for recording and integrating all the properties of the elicited sensations with detailed standard questionnaires and which could be used for several types of sensory feedback.

The psychometric platform presented in this study provides a uniform way of characterizing and quantifying the artificial sensory feedback systems used for invasive and non-invasive, peripheral, and central sensory feedback, in order to efficiently compare, optimize, and evaluate all the different approaches even over time. Our platform records the stimulation parameters, quality, type, intensity, and location of the evoked sensations. All the sensation data are collected from questionnaires already presented in the literature.

The platform also provides a user-friendly graphical user interface with a touch screen for the patient's answers that not only enables the patient to describe the percept in detail but also provides clinicians with all the main information on the evoked sensation. The platform accepts new questionnaire definitions as text and is easy to understand and implement. This means that researchers can add new questionnaires, such as phantom limb pain (PLP) (37, 38), in order to collect information on new treatments.

This psychometric platform was tested on three trans-radial amputees who had four intrafascicular electrodes (39) implanted in their median and ulnar nerves for 6 months each. The patients responded using the psychometric platform when they received electrical stimulation by the electrical contacts of the neural interfaces. The software was used by clinicians and engineers to collect the data. This has proven to be more convenient than writing down hundreds of answers in weekly trials over 18 months.

In this study, we describe the usability of this new platform. We believe that our new psychometric platform will facilitate and unify the characterization of percepts and the comparison of the effects when applying different neural stimulation techniques or using different devices.

METHODS

Software Platform

The psychometric platform is made up of a mobile application for compiling questionnaires (which we have called Easy Quest), two desktop tools (Easy Quest Create and Easy Quest Evaluate), and a desktop application to control the neurostimulator.

The Easy Quest mobile app is described in depth in the following sections.

Easy Quest Create shows a simple graphical user interface in which the experimenter can create a list of questions from a set of predefined types. The content can be customized.

Easy Quest Evaluate is devised for the rapid evaluation of a set of answers; the software reads the archive file exported by Easy Quest and exports a comma-separated values (CSV) file. The choice of CSV format of the results makes further analyses easier, as it is compatible with MATLAB (MathWorks, Inc., Natick, Massachusetts, United States) and Microsoft Excel (Microsoft Corporation, Redmond, Washington, United States).

The desktop application for the actual neurostimulation is not described here, because its design is strongly dependent on the type of experiment and neurostimulation device (communication protocols, stimulator commands, and architecture); however it is mentioned as part of the experimental setup.

Somatosensory Questionnaires

Somatosensory descriptors were selected from the literature and clinical settings also including questionnaires that have already been used in neuroprosthetic studies. Several options describing the type, quality, intensity, and the location are presented in order to characterize the somatosensory percepts being evoked during the stimulation. To describe the quality of sensations, we used a scale presented by Lenz et al. (34) and used also by Valle et al. (9). For the sensation type, we adapted the questionnaire proposed by Kim et al. (36) based on our experience with several upper limb patients stimulated with invasive (4, 9, 23, 24, 29, 31, 40–47) and non-invasive technologies (17, 48). We also considered other studies on sensations elicited using peripheral (5, 10, 14, 22, 49–55) or central (6, 8, 13) neural stimulation approaches. For the intensity, we used a Visual Analog Scale (VAS) (37) already presented by Tan et al. (56). Lastly, the perceived sensation locations were shown directly on a schematic representation of the human hand. It is further possible to select the feet, arms, or legs (11, 12) with several possible spots (**Figure 2**). In this way, the patient can accurately indicate the affected areas.

We added several questionnaires in order to collect information on PLP: VAS (37) and neuropathic pain symptom inventory (NPSI) (38). It is also possible to add or modify the existing questionnaires in order to adapt the platform to the needs and specifications of the clinical trial.

Use Cases

Two main use cases for the app were identified (**Figure 3**). In the first, the user fills in a questionnaire and saves the results on the device, defined as the "local fill-in" (LF). In the second, an external software prompts the app to show a questionnaire and to send back the results, defined as "on demand fill-in" (ODF). The two cases (**Figures 3A,B**) involve the same procedure in the part where the user is asked to fill in the answers.

The main difference, besides the location where the results are stored, is how the procedure starts: in the first case, the user choses a questionnaire by selecting it from the main menu, in the second, the app waits for an external command, usually from the network, instructing the software to show a specific questionnaire.

The application can set recurrent reminders for specific questionnaires, enabling the experimenter to plan the follow-up for home use by the patient, and these reminders prompt the user to fill in the questions in LF mode.

A third use case (**Figure 3C**) explains the workflow from the perspective of the experimenter, who uses the companion software to define new questionnaires at the beginning of the experimentation and to display the results at the end.

An explanatory example of the platform is displayed in **Supplementary Video 1**.

Software Architecture

The software was developed in Dart, an object-oriented programming language developed by Google in order to address server-side, web, and mobile platforms. The mobile SDK, Flutter, compiles the code in fast native apps for Android and iOS devices.

The app is developed following the Model View Controller (MVC) pattern, and a simple Object-Relational Mapping (ORM) is implemented to store the models in an SQLite database in the device's memory. The ORM is accessed through classes which show APIs where serialized objects can be stored and retrieved.

To implement the ODF, a simple Hypertext Transfer Protocol (HTTP) server runs in the background thus the app can, when requested, wait for remote commands from the network. While doing so, the app shows a numerical code, which must be notified to the experimenter to secure the remote connection.

An interface with the mail app is used to send the completed questionnaires as a CSV file by e-mail.

Another provider class parses the questionnaires defined in JavaScript Object Notation (JSON) format, making it possible to create and add new questionnaires to working devices,

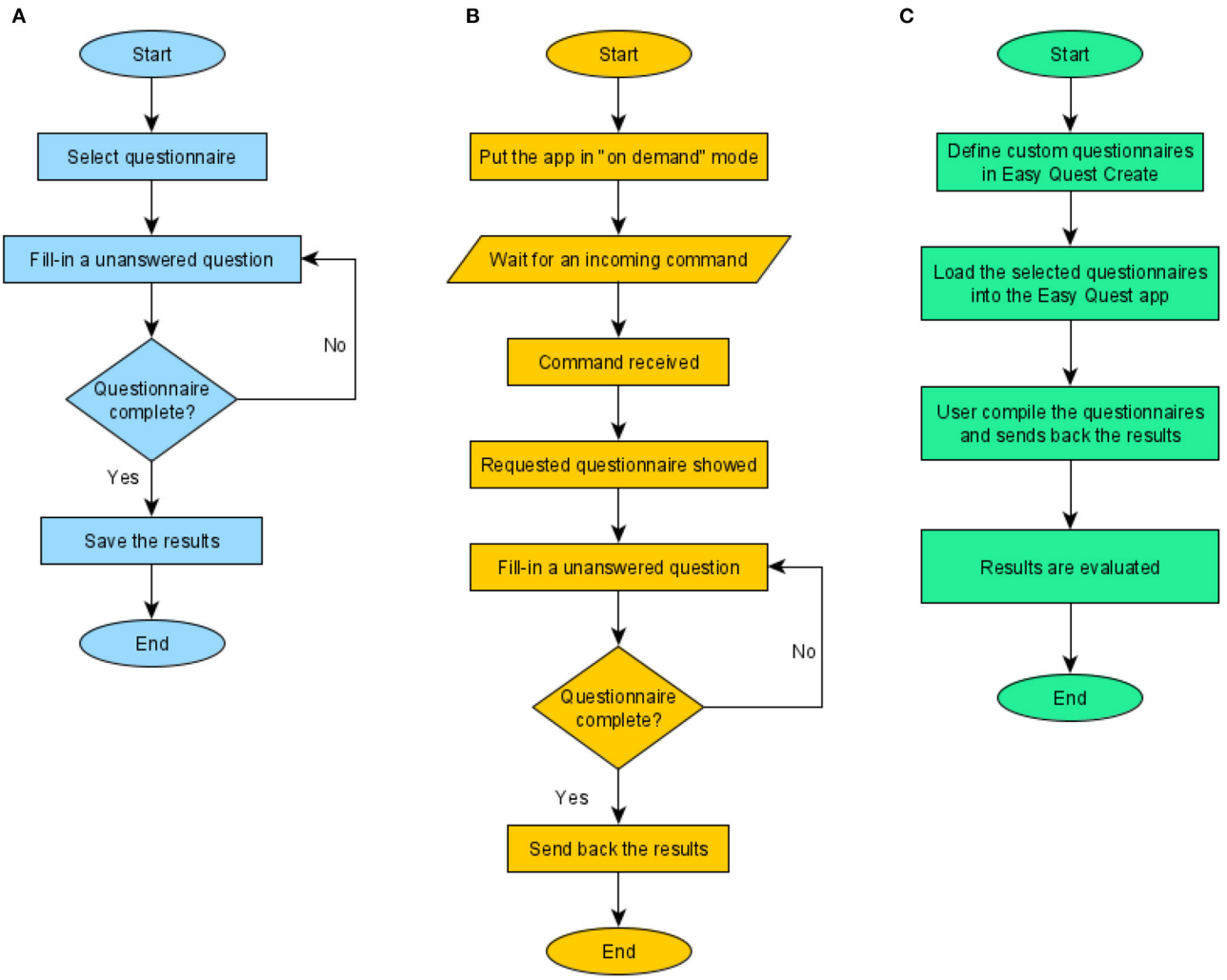


FIGURE 3 | Use cases. The three main features of the psychometric platform: the first two are implemented by the mobile app and the last by the whole system. **(A)** Defined as Local Fill-in (LF), where the users compile a questionnaire and the answers are stored in the device. **(B)** On demand fill-in (ODF), in this case, the app waits for an external command from a controller app containing information on the questionnaire to be shown; the fill-in procedure is the same but nothing is stored within the device, instead results are sent back to the controller. **(C)** The procedure seen from the experimenter's point of view, here, the role of the other software programs of the platform (Easy Quest Create, Easy Quest Evaluate) is explained.

without code interventions and recompiling the whole app. The import service can parse a compressed file containing a set of questionnaires and also a collection of images referred to in the questions. There are five questions accepted by the parser: (1) open, which prompts the user for a string (2); radio, which asks the user to choose one option from a set (3); multiple choice; (4) slider, where users have to select a number or a label with a slider; and (5) image touch, where the user selects a set of touchable areas displayed on top of a given background image.

The app enables multiple users to access the same device while keeping the results separate.

The system architecture is shown in **Figure 4**, along with the external software highlighting its relationship with the app modules.

The app UI/UX is designed in accordance with Material, an open source system of guidelines developed by Google. The view layer written for the app exploits all the available space, presenting the questionnaire as a list of questions on small devices and as a grid on larger screens.

Quality and Usability Assessment

During the clinical trial, we collected feedback information from patients, clinicians, and engineers who used the platform presented in this study in three clinical trials ($N = 12$). The investigations regarded the development and assessment of bidirectional hand prostheses for upper limb amputees with a neural sensory feedback delivered by implantable electrodes (9, 23, 31, 41, 42). After 6 months of use, we asked participants to answer different quality and usability questions using: questionnaires for user interface satisfaction (QUIS) (57), system usability scales (SUS) (58, 59), Nielsen's attributes of usability (NAU) (60), and after-scenario questionnaires (ASQ) (61). We collected and analyzed all the information using validated and standardized questionnaires (**Figure 5**).

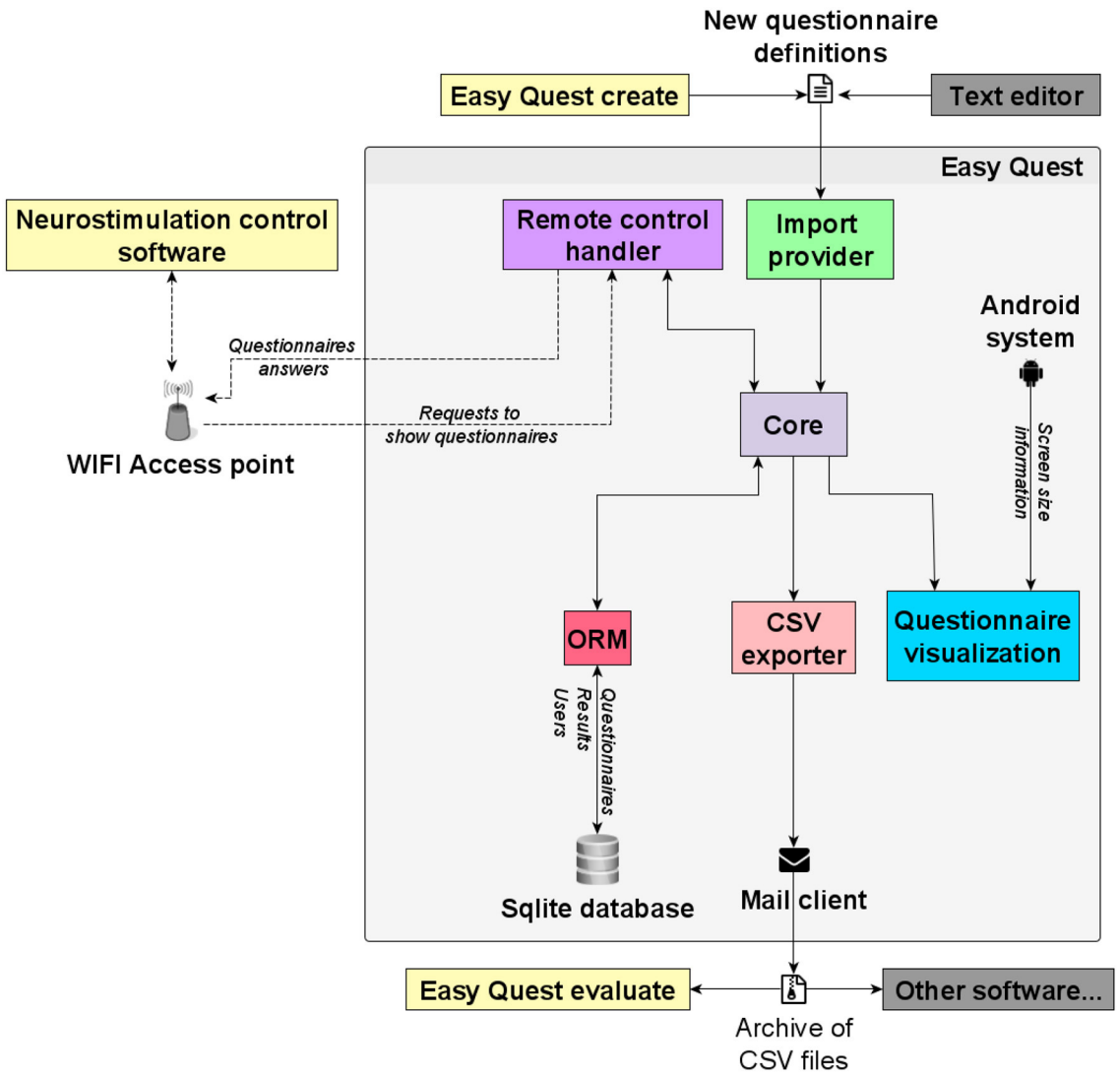


FIGURE 4 | Software architecture. The main components of the platform depicted as squares, external services are shown with an icon and communication with arrows, some show a label with examples of the information flowing through. A gray shadow surrounds the software modules of the mobile app (Easy Quest).

Data Storage

Alongside the answers to the psychometric questionnaires, a MATLAB script running in experimenter's computer stored date and time for further analysis about the use of the platform. All the relevant measurements are stored as.mat files, as all the data processing have been performed inside the MATLAB software.

RESULTS

Somatosensory Questionnaires for Sensation Characterization

To efficiently characterize the sensations emanating from (invasive or non-invasive) electrical (central or peripheral) stimulation, a user-friendly platform is needed with a set of somatosensory-related questionnaires. This helps to reduce the long time required to collect all the electrically evoked sensation data.

To assess the properties of the sensations being evoked by stimulating peripheral nerves using a neural interface in trans-radial amputees, we used the psychometric platform presented here. We performed a procedure called "sensation characterization" with all the patients involved in the clinical investigation (**Figure 2**). For each electrically active site used to stimulate the nerve, the neural stimulation was delivered, and the patient was asked to report the sensations he/she felt. This mapping phase enabled us to identify the sensation properties for all the stimulation channels of the implanted electrodes by varying the stimulation parameters and building a personalized map of the sensations. The stimulation parameters varied in terms of frequency (1–1,000 Hz), pulse-width (1–120 μs), and amplitude (1–1,000 μA), as well as stimulation

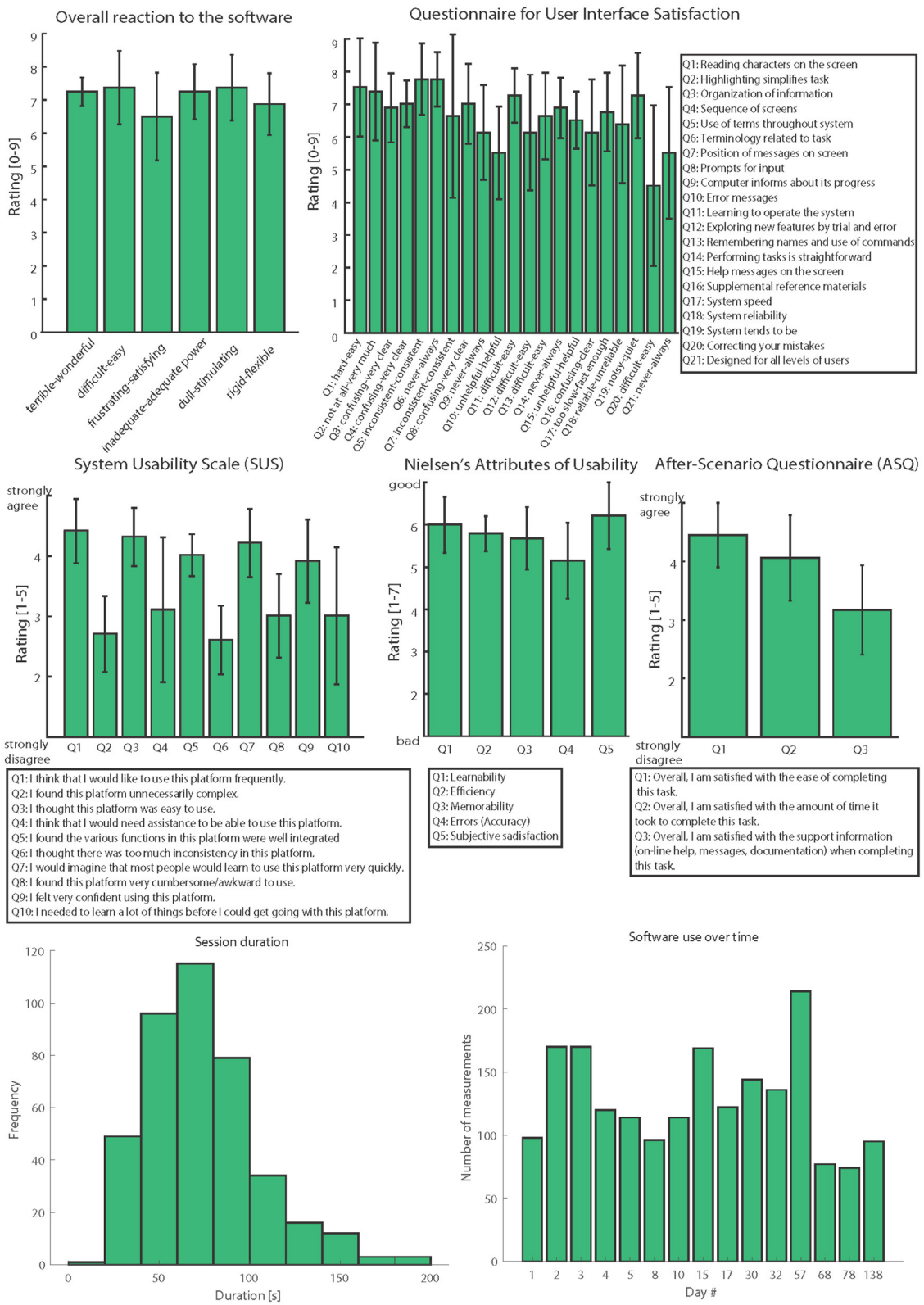


FIGURE 5 | Usability assessment. All the usability scales are reported: Overall reaction to the software, QUIS, SUS, NAU, and ASQ. Three clinicians, six engineers, and three patients evaluated the psychophysical platform ($N = 12$). The data in the figure are represented as means ± standard deviations. The last row resumes usage metrics.

TABLE 1 | List of sensory descriptors.

Vibration
Flutter
Buzz
Movement through body/across skin
Movement without motor activity
Urge to move
Touch
Pressure
Sharp
Prick
Tap
Electric current
Shock
Pulsing
Tickle
Itch
Tingle
Numb
Warm
Cool

Here, the chosen descriptors are shown, the user can also enter free text when the other options are insufficient.

train duration (discrete or continuous). We collected the sensation intensity, quality, type, and location of the patient's perceived sensations.

The intensity was used to find the perceptual thresholds for each stimulation channel (4, 23, 24, 31), together with the range of stimulation (between threshold and below pain level). Using a VAS scale in the range from 0 to 10 also enables us to identify perceptual magnitude levels (5, 23, 32, 52, 62).

The quality of the sensory feedback was assessed in order to test different stimulation strategies and approaches (9, 23), since this quality is considered to be an important factor for prosthesis acceptance (53, 63). To quantify the perception quality and naturalness, we used a scale (34) from 1 (totally unnatural) to 5 (totally natural).

The type of sensation was collected in order to understand the type of fibers being recruited during the stimulation and to identify the best channels for restoring homologous sensations while using the bidirectional prosthesis. We used 20 descriptors (**Table 1**) considering all the important aspects. In this platform, the patient could also report a new sensation or add comments in an empty text box when a correct descriptor for the elicited sensation was lacking.

The sensation location was reported using a picture of the limb of interest (foot, arm, leg, or hand) with several highlighted spots (20 for foot, 24 for leg, 48 for arm, and 45 for hand) (**Figure 2**). The zones with a higher density of receptors had more selectable spots. This information is useful to understand the electrode stimulation selectivity (analyzing the spreading of the zone) and the layout of the fibers inside the nerve. In addition while the bidirectional prosthesis was being used, the location map was needed to stimulate the correct active sites eliciting the somatotopic sensation during the prosthesis hand/finger contact with objects (4).

Finally, several questions can be used to assess phantom limb pain levels before and after a pain treatment with electrical stimulation (31). We decided to use two different questionnaires (VAS and NPSI) to characterize the location, quality, and intensity of the pain (11, 31).

Software Usability

The usability testing of the app was performed on an Android phone (a Nexus 6p), designed by Huawei and running Android 8.

The app loading time is <2 s, needing only the time to open the local database, and after the login screen, the user can access all the main functions in no more than two taps.

The home page shows a list of all the available questionnaires, the user can tap on each one to see the questions and fill in the answers, which are stored in the internal database.

From the lateral menu (drawer), the ODF mode can be accessed in only one tap, after which the app will wait for a network command containing the identifier of the questionnaire to be shown.

Minimal user interaction is needed to complete a questionnaire, usually all the questions need just one tap, except for the multiple choice and clickable area ones. The average time to fill in a sensation characterization questionnaire is 10 s.

The export page lets the user write all the stored data in a CSV archive file and opens the default mail to send to the experimenters for further analysis, facilitating and speeding up the data-gathering phase.

A specific section of the app lets the user choose which questionnaire should be visible in the home page, personalizing the user interface for a specific use.

Other pages are designed for secondary tasks, such as previewing stored answers and editing settings.

Psychometric System Validation

In order to assess the usability and quality of this novel psychometric platform to collect somatosensory percepts, several questionnaires were filled in by different kinds of users. Three patients, six engineers, and three clinicians evaluated the system by answering four questionnaires after using the platform in clinical applications (**Figure 5**). Analyzing the results, the overall reactions to the system were very positive. The average score was 7.1 ± 0.3. Considering the user interface satisfaction (QUIS), the rating achieved was 6.6 ± 0.8. In both these questionnaires, the maximum achievable score was 9.

In the SUS (range 1–5), Q1-Q3-Q5-Q7-Q9 scored 4 ± 0.2, while Q2-Q4-Q6-Q8 scored 2.3 ± 0.2. These results indicate that the users agreed more with the positive sentences and disagreed more with negative ones. The NAU (range 1–7) showed high ratings of 5.6 ± 0.5, and the ASQ (range 1–5) showed an average value of 3.7 ± 0.8.

During the clinical trial, the psychophysical platform stored 1,913 measurements.

Usage Metrics

The time passed between a measurement and the subsequent have been calculated and plotted in **Figure 5**. Records were aggregated by day and outliers due to technical problems and breaks for the subject were ignored, the mean time between two measurements is 74 ± 31.8 s. The number of measurements per day have also being considered an indicator of the overall use of the platform, the mean use overall was 128 ± 39.6 measurements per day. Each measurement consisted of a.mat file of roughly 1.14 kB, the whole dataset totaled then for almost 2.2 MB.

DISCUSSION

Electrical stimulation has been proposed as a way of restoring somatosensations (15, 63–67) in cases where they have been lost due to injury or disease in both the CNS or the PNS. In fact, sensory feedback is crucial to improve the motor control of robotic limbs or prostheses, enabling the patient to be more efficient in manipulating objects (4, 28, 55, 68). The sensations evoked thus had to be characterized in detail in patients receiving stimulation in order to restore the sensory information. The psychometric questionnaires were able to register all the aspects of the sensations being restored in a reliable and efficient way, considering more descriptors than in previous studies (34) and using a user-friendly platform.

Currently, there are many important sensation properties which need to be collected in order to obtain an intuitive and rich sensory feedback. In particular, the sensation location, type, quality, and intensity are valid and extendable for all the approaches in different neurological conditions. Considering the previously presented interface to collect stimulation-evoked somatosensory percepts, Geng et al. (35) showed a platform used to evaluate electrical stimulation to relieve PLP. Their platform was interfaceable with one type of neural stimulator and contained three questions to characterize the evoked sensation considering 12 sensation descriptors. The psychometric platform presented here reports somatosensory percepts based on five questionnaires containing 20 standard sensory descriptors (**Table 1**). The platform exploits a customizable, fast, and easy-to-use GUI which can be efficiently connected to several neural stimulators (31, 69, 70).

Since several groups are currently using electrical stimulation to restore sensory feedback, a standard somatosensory platform could facilitate their comparison, assessment, and optimization. Our findings support the conclusion that this psychometric platform could help and accelerate the development of sensorimotor neuroprostheses.

Given the simple software architecture, this platform is flexible in terms of modifications and upgrades. It is possible to add new questionnaires regarding other aspects of sensory feedback restoration. For example, two important features to be considered for the development of the next generation of somatosensory neuroprostheses are embodiment (42, 71, 72) and psychological/affective aspects (73).

The psychometric platform is simple to interface with other devices and also with existing software, thanks to its open and platform-agnostic interfaces: in ODF mode the HTTP interface accepts commands regardless of the device and the programming language of the sender application (all major languages can implement HTTP communication effortlessly). Answers to the questionnaires are exported in a CSV format, making it easy for any other software program to import and analyze them.

Considering the results of the usability assessments (**Figure 5**), users highlighted various positive and negative aspects which will then help us to improve the platform. The most positive aspect in terms of the "overall reaction to the software" was that the software is easy to use, which is crucial both for patients and experimenters.

The QUIS answers revealed that this system is consistent and very clear; however, we still need to improve error and warning messages. These aspects mainly regard the experimenters' side. The SUS again indicated that the system is easy to use and intuitive, but additional material and instructions should be included as support. Moreover, the NAU showed a high user satisfaction along with a request for more error messages. Finally, the ASQ revealed "the ease of completing this task," thus highlighting the need for more support, information, and documentation. We thus intend to improve the platform using these usability results.

Usage metrics confirmed the ease of the fill-in process, with a mean time needed for a complete session of 74 s. Thanks to that stimulation spots were probed with a suitable rate (128 ± 39.6 measurements per day), for the intended objectives. The small dimension of the measurements made easy the subsequent steps of data revision, analysis, and storage, anyway authors are planning further reduction in the memory needed.

STUDY LIMITATIONS

There are several limitations connected to the patient attention at the time of testing. To solve this issue, it is important to repeat the test multiple times over multiple days in order to increase its reliability. The test is also highly subjective, and the mapping results could strongly depend on the sensation of the patient and his/her personal experience (74). The individual subjective differences remain a big challenge for interpreting the somatosensory results and also the semantic differences. The usability and the utility of the platform need to be demonstrated with a larger population of patients with sensory deficits and with other technologies for the restoration of sensory feedback (e.g., non-invasive stimulation, brain stimulation).

Sham (placebo) and blind stimulations could also be delivered to test individual response bias and identify possible unreliable self-reports.

The software design, particularly the GUI, was inspired by the principles of the ISO 9241 standard. In fact, the users' opinions of the platform were taken into consideration during the design phase and the assessment.

The software will be actively used during experiments, and the user experience will be monitored to improve new versions, ensuring an iterative development driven by user feedback, as also stated in ISO 9241. In particular, in addition to the online availability of the software, the platform needs to be used in future clinical trials for both upper- and lower-limb amputees

provided with a fully implantable sensory feedback system in long-term studies.

CONCLUSIONS

This study has presented a psychometric platform used to record a complete somatosensory percept description, which can be evoked by several different methods of electrical stimulation in humans. The subjective somatosensory sensation type, location, quality, and intensity are collected and used to develop a somatosensory questionnaire, which can be used for neuroprosthesis calibration and optimization. The psychometric toolbox is implemented in a user-friendly software program. The platform was validated in patients with electrodes implanted in the PNS.

We believe that this new somatosensory psychometric system will help to establish a standard and uniform methodology of subjective sensory reports, which is a pivotal step to uniformly develop, adapt, and improve somatosensory neuroprostheses.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The Easy Quest Android application is available from the Google Play Store platform: https://play.google.com/store/apps/details?id=me.francescoiberite.research.easy_quest.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Ethics Committees of Policlinic A. Gemelli at the Catholic University and Italian Ministry of Health. 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

GV designed the study, developed the software, analyzed the data, and wrote the paper. FI developed the software, analyzed the data, and reviewed the manuscript. IS designed the study, developed the software, and reviewed the manuscript. ED'A developed the software and reviewed the manuscript. GG, RD, and PR tested the platform with patients. TS developed the TIME electrodes. TS and PR discussed the results and reviewed the manuscript. SR designed the platform. FP designed the platform and discussed the results. SM designed the study, discussed the results, and reviewed the manuscript. All authors read, commented, and approved the manuscript.

FUNDING

This work was supported by the EU Grant FET 611687 NEBIAS Project (NEurocontrolled BIdirectional Artificial upper limb and hand prosthesiS) and Bertarelli Foundation.

ACKNOWLEDGMENTS

This manuscript has been released as a preprint at Biorxiv (75).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmedt.2021.619280/full#supplementary-material

Supplementary Video 1 | Example of collection of somatosensory sensation data. The video shows the ODF procedure, where the Easy Quest app and the software controlling the stimulator interact. After the login procedure, the user sets the android application in ODF, opening the communication port. As soon as the controller software, after a verification procedure, detects the running app, there is a visual confirmation in the dashboard on the right-hand side of the screen. When stimulation parameters are correctly set, the software can deliver the stimulation and automatically send a sensation questionnaire to the subject's device. Instead of being stored in the device, the answers are sent back to the control software and are shown in the lower part of the interface.

REFERENCES

1. Borton D, Micera S, Millan J del R, Courtine G. Personalized neuroprosthetics. *Sci Transl Med.* (2013) 5:210rv2. doi: 10.1126/scitranslmed.3005968
2. Klaes C, Shi Y, Kellis S, Minxha J, Revechkis B, Andersen RA. A cognitive neuroprosthetic that uses cortical stimulation for somatosensory feedback. *J Neural Eng.* (2014) 11:056024. doi: 10.1088/1741-2560/11/5/056024
3. Ono T, Shindo K, Kawashima K, Ota N, Ito M, Ota T, et al. Brain-computer interface with somatosensory feedback improves functional recovery from severe hemiplegia due to chronic stroke. *Front Neuroeng.* (2014) 7:19. doi: 10.3389/fneng.2014.00019
4. Raspopovic S, Capogrosso M, Petrini FM, Bonizzato M, Rigosa J, Di Pino G, et al. Restoring natural sensory feedback in real-time bidirectional hand prostheses. *Sci Transl Med.* (2014) 6:222ra19. doi: 10.1126/scitranslmed.3006820
5. Tan DW, Schiefer MA, Keith MW, Anderson JR, Tyler J, Tyler DJ. A neural interface provides long-term stable natural touch perception. *Sci Transl Med.* (2014) 6:257ra138. doi: 10.1126/scitranslmed.3008669
6. Flesher SN, Collinger JL, Foldes ST, Weiss JM, Downey JE, Tyler-Kabara EC, et al. Intracortical microstimulation of human somatosensory cortex. *Sci Transl Med.* (2016) 8:361ra141. doi: 10.1126/scitranslmed.aaf8083
7. Hiremath SV, Tyler-Kabara EC, Wheeler JJ, Moran DW, Gaunt RA, Collinger JL, et al. Human perception of electrical stimulation on the surface of somatosensory cortex. *PLoS ONE.* (2017) 12:e0176020. doi: 10.1371/journal.pone.0176020
8. Salas MA, Bashford L, Kellis S, Jafari M, Jo H, Kramer D, et al. Proprioceptive and cutaneous sensations in humans elicited by intracortical microstimulation. *eLife Sci.* (2018) 7:e32904. doi: 10.7554/eLife.32904
9. Valle G, Mazzoni A, Iberite F, D'Anna E, Strauss I, Granata G, et al. Biomimetic intraneural sensory feedback enhances sensation naturalness, tactile sensitivity, and manual dexterity in a bidirectional prosthesis. *Neuron.* (2018). doi: 10.1016/j.neuron.2018.08.033
10. Overstreet CK, Cheng J, Keefer E. Fascicle specific targeting for selective peripheral nerve stimulation. *J. Neural Eng.* (2019) 16:066040. doi: 10.1088/1741-2552/ab4370

11. Petrini FM, Bumbasirevic M, Valle G, Ilic V, Mijović P, Cvančara P, et al. Sensory feedback restoration in leg amputees improves walking speed, metabolic cost and phantom pain. *Nat Med.* (2019) 25:1356–63. doi: 10.1038/s41591-019-0567-3
12. Petrini FM, Valle G, Bumbasirevic M, Barberi F, Bortolotti D, Cvancara P, et al. Enhancing functional abilities and cognitive integration of the lower limb prosthesis. *Sci Transl Med.* (2019) 11:eaav8939. doi: 10.1126/scitranslmed.aav8939
13. Chandrasekaran S, Nanivadekar AC, McKernan G, Helm ER, Boninger ML, Collinger JL, et al. Sensory restoration by epidural stimulation of the lateral spinal cord in upper-limb amputees. *eLife.* (2020) 9:e54349. doi: 10.7554/eLife.54349.sa2
14. Ortiz-Catalan M, Mastinu E, Sassu P, Aszmann O, Brånemark R. Self-contained neuromusculoskeletal arm prostheses. *N Engl J Med.* (2020) 382:1732–8. doi: 10.1056/NEJMoa1917537
15. Heming EA, Choo R, Davies JN, Kiss ZHT. Designing a thalamic somatosensory neural prosthesis: consistency and persistence of percepts evoked by electrical stimulation. *IEEE Trans Neural Syst Rehabil Eng.* (2011) 19:477–82. doi: 10.1109/TNSRE.2011.2152858
16. Chai G, Sui X, Li S, He L, Lan N. Characterization of evoked tactile sensation in forearm amputees with transcutaneous electrical nerve stimulation. *J Neural Eng.* (2015) 12:066002. doi: 10.1088/1741-2560/12/6/066002
17. D'Anna E, Petrini FM, Artoni F, Popovic I, Simanić I, Raspopovic S, et al. A somatotopic bidirectional hand prosthesis with transcutaneous electrical nerve stimulation based sensory feedback. *Sci Rep.* (2017) 7:10930. doi: 10.1038/s41598-017-11306-w
18. Osborn LE, Dragomir A, Betthauser JL, Hunt CL, Nguyen HH, Kaliki RR, et al. Prosthesis with neuromorphic multilayered e-dermis perceives touch and pain. *Sci Robotics.* (2018) 3:1–11. doi: 10.1126/scirobotics.aat3818
19. Slopsema JP, Boss JM, Heyboer LA, Tobias CM, Draggoo BP, Finn KE, et al. Natural sensations evoked in distal extremities using surface electrical stimulation. *Open Biomed Eng J.* (2018) 12:1–15. doi: 10.2174/1874120701812010001
20. Rothwell JC, Traub MM, Day BL, Obeso JA, Thomas PK, Marsden CD. Manual motor performance in a deafferented man. *Brain.* (1982) 105:515–42. doi: 10.1093/brain/105.3.515
21. Jarrassé N, Kühne M, Roach N, Hussain A, Balasubramanian S, Burdet E, et al. Analysis of grasping strategies and function in hemiparetic patients using an instrumented object. In: *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*. Seattle, WA (2013). p. 1–8.
22. Graczyk EL, Schiefer MA, Saal HP, Delhaye BP, Bensmaia SJ, Tyler DJ. The neural basis of perceived intensity in natural and artificial touch. *Sci Transl Med.* (2016) 8:362ra142. doi: 10.1126/scitranslmed.aaf5187
23. Valle G, Petrini FM, Strauss I, Iberite F, D'Anna E, Granata G, et al. Comparison of linear frequency and amplitude modulation for intraneural sensory feedback in bidirectional hand prostheses. *Sci Rep.* (2018) 8:16666. doi: 10.1038/s41598-018-34910-w
24. Valle G, Strauss I, D'Anna E, Granata G, Di Iorio R, Stieglitz T, et al. Sensitivity to temporal parameters of intraneural tactile sensory feedback. *J NeuroEngineering Rehabil.* (2020) 17:110. doi: 10.1186/s12984-020-00737-8
25. Raspopovic S, Petrini FM, Zelechowski M, Valle G. Framework for the development of neuroprostheses: from basic understanding by sciatic and median nerves models to bionic legs and hands. *Proc IEEE.* (2017) 105:34–49. doi: 10.1109/JPROC.2016.2600560
26. Romeni S, Valle G, Mazzoni A, Micera S. Tutorial: a computational framework for the design and optimization of peripheral neural interfaces. *Nat Protoc.* (2020) 15:3129–53. doi: 10.1038/s41596-020-0377-6
27. Zelechowski M, Valle G, Raspopovic S. A computational model to design neural interfaces for lower-limb sensory neuroprostheses. *J NeuroEng Rehabil.* (2020) 17:24. doi: 10.1186/s12984-020-00657-7
28. Schiefer M, Tan D, Sidek SM, Tyler DJ. Sensory feedback by peripheral nerve stimulation improves task performance in individuals with upper limb loss using a myoelectric prosthesis. *J Neural Eng.* (2016) 13:016001. doi: 10.1088/1741-2560/13/1/016001
29. Clemente F, Valle G, Controzzi M, Strauss I, Iberite F, Stieglitz T, et al. Intraneural sensory feedback restores grip force control and motor coordination while using a prosthetic hand. *J Neural Eng.* (2019) 16:026034. doi: 10.1088/1741-2552/ab059b
30. Tan DW, Schiefer MA, Keith MW, Anderson JR, Tyler DJ. Stability and selectivity of a chronic, multi-contact cuff electrode for sensory stimulation in human amputees. *J Neural Eng.* (2015) 12:026002. doi: 10.1088/1741-2560/12/2/026002
31. Petrini FM, Valle G, Strauss I, Granata G, Di Iorio R, D'Anna E, et al. Six-months assessment of a hand prosthesis with intraneural tactile feedback. *Ann Neurol.* (2018) 85:137–54. doi: 10.1002/ana.25384
32. George JA, Page DM, Davis TS, Duncan CC, Hutchinson DT, Rieth LW, et al. Long-term performance of Utah slanted electrode arrays and intramuscular electromyographic leads implanted chronically in human arm nerves and muscles. *J Neural Eng.* (2020) 17:056042. doi: 10.1088/1741-2552/abc025
33. Hughes CL, Flesher SN, Weiss JM, Downey JE, Collinger JL, Gaunt RA. Neural stimulation and recording performance in human somatosensory cortex over 1500 days. *medRxiv [preprint].* (2020). doi: 10.1101/2020.01.21.20018341
34. Lenz FA, Seike M, Richardson RT, Lin YC, Baker FH, Khoja I, et al. Thermal and pain sensations evoked by microstimulation in the area of human ventrocaudal nucleus. *J Neurophysiol.* (1993) 70:200–12. doi: 10.1152/jn.1993.70.1.200
35. Geng B, Harreby KR, Kundu A, Yoshida K, Boretius T, Stieglitz T, et al. Developments towards a psychophysical testing platform - a computerized tool to control, deliver and evaluate electrical stimulation to relieve phantom limb pain. In: Dremstrup K, Rees S, Jensen ØM, editors. *15th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC 2011) IFMBE Proceedings*. Berlin: Springer (2011). p. 137–40.
36. Kim LH, McLeod RS, Kiss ZHT. A new psychometric questionnaire for reporting of somatosensory percepts. *J Neural Eng.* (2018) 15:013002. doi: 10.1088/1741-2552/aa966a
37. Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. *Res Nurs Health.* (1990) 13:227–36. doi: 10.1002/nur.4770130405
38. Bouhassira D, Attal N, Fermanian J, Alchaar H, Gautron M, Masquelier E, et al. Development and validation of the neuropathic pain symptom inventory. *Pain.* (2004) 108:248–57. doi: 10.1016/j.pain.2003.12.024
39. Boretius T, Badia J, Pascual-Font A, Schuettler M, Navarro X, Yoshida K, et al. A transverse intrafascicular multichannel electrode (TIME) to interface with the peripheral nerve. *Biosens Bioelectr.* (2010) 26:62–9. doi: 10.1016/j.bios.2010.05.010
40. Rossini PM, Micera S, Benvenuto A, Carpaneto J, Cavallo G, Citi L, et al. Double nerve intraneural interface implant on a human amputee for robotic hand control. *Clin Neurophysiol.* (2010) 121:777–83. doi: 10.1016/j.clinph.2010.01.001
41. Granata G, Di Iorio R, Romanello R, Iodice F, Raspopovic S, Petrini F, et al. Phantom somatosensory evoked potentials following selective intraneural electrical stimulation in two amputees. *Clin Neurophysiol.* (2018) 129:1117–20. doi: 10.1016/j.clinph.2018.02.138
42. Rognini G, Petrini FM, Raspopovic S, Valle G, Granata G, Stauss I, et al. Multisensory bionic limb to achieve prosthesis embodiment and reduce distorted phantom limb perceptions. *J Neurol Neurosurg Psychiatry.* (2018) 1–3. doi: 10.1136/jnnp-2018-318570
43. Risso G, Valle G, Iberite F, Strauss I, Stieglitz T, Controzzi M, et al. Optimal integration of intraneural somatosensory feedback with visual information: a single-case study. *Sci Rep.* (2019) 9:7916. doi: 10.1038/s41598-019-43815-1
44. Strauss I, Valle G, Artoni F, D'Anna E, Granata G, Iorio RD, et al. Characterization of multi-channel intraneural stimulation in transradial amputees. *Sci Rep.* (2019) 9:1–11. doi: 10.1038/s41598-019-55591-z
45. Granata G, Valle G, Di Iorio R, Iodice F, Petrini FM, Strauss I, et al. Cortical plasticity after hand prostheses use: is the hypothesis of deafferented cortex "invasion" always true? *Clin Neurophysiol.* (2020) 131:2341–8. doi: 10.1016/j.clinph.2020.06.023
46. Mazzoni A, Oddo CM, Valle G, Camboni D, Strauss I, Barbaro M, et al. Morphological neural computation restores discrimination of naturalistic textures in trans-radial amputees. *Sci Rep.* (2020) 10:1–14. doi: 10.1038/s41598-020-57454-4
47. Valle G, D'Anna E, Strauss I, Clemente F, Granata G, Di Iorio R, et al. Hand control with invasive feedback is not impaired by increased cognitive load. *Front Bioeng Biotechnol.* (2020) 8:287. doi: 10.3389/fbioe.2020.00287

48. D'Anna E, Valle G, Mazzoni A, Strauss I, Iberite F, Patton J, et al. A closed-loop hand prosthesis with simultaneous intraneural tactile and position feedback. *Sci Robot.* (2019) 4:eaau8892. doi: 10.1126/scirobotics.aau8892
49. Dhillon GS, Horch KW. Direct neural sensory feedback and control of a prosthetic arm. *IEEE Trans Neural Syst Rehabil Eng.* (2005) 13:468–72. doi: 10.1109/TNSRE.2005.856072
50. Davis TS, Wark HAC, Hutchinson DT, Warren DJ, O'Neill K, Scheinblum T, et al. Restoring motor control and sensory feedback in people with upper extremity amputations using arrays of 96 microelectrodes implanted in the median and ulnar nerves. *J Neural Eng.* (2016) 13:036001. doi: 10.1088/1741-2560/13/3/036001
51. Wendelken S, Page DM, Davis T, Wark HAC, Kluger DT, Duncan C, et al. Restoration of motor control and proprioceptive and cutaneous sensation in humans with prior upper-limb amputation via multiple Utah Slanted Electrode Arrays (USEAs) implanted in residual peripheral arm nerves. *J Neuroeng Rehabil.* (2017) 14:121. doi: 10.1186/s12984-017-0320-4
52. Graczyk EL, Resnik L, Schiefer MA, Schmitt MS, Tyler DJ. Home use of a neural-connected sensory prosthesis provides the functional and psychosocial experience of having a hand again. *Sci Rep.* (2018) 8:9866. doi: 10.1038/s41598-018-26952-x
53. George JA, Kluger DT, Davis TS, Wendelken SM, Okorokova EV, He Q, et al. Biomimetic sensory feedback through peripheral nerve stimulation improves dexterous use of a bionic hand. *Sci Robot.* (2019) 4:eaax2352. doi: 10.1126/scirobotics.aax2352
54. Zollo L, Pino GD, Ciancio AL, Ranieri F, Cordella F, Gentile C, et al. Restoring tactile sensations via neural interfaces for real-time force-and-slippage closed-loop control of bionic hands. *Sci Robot.* (2019) 4:eaau9924. doi: 10.1126/scirobotics.aau9924
55. Mastinu E, Engels LF, Clemente F, Dione M, Sassu P, Aszmann O, et al. Neural feedback strategies to improve grasping coordination in neuromusculoskeletal prostheses. *Sci Rep.* (2020) 10:11793. doi: 10.1038/s41598-020-67985-5
56. Tan D, Tyler D, Sweet J, Miller J. Intensity modulation: a novel approach to percept control in spinal cord stimulation. *Neuromodulation.* (2016) 19:254–9. doi: 10.1111/ner.12358
57. Chin JP, Diehl VA, Norman KL. Development of an instrument measuring user satisfaction of the human-computer interface. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'88.* New York, NY: ACM (1988). p. 213–8. doi: 10.1145/57167.57203
58. Brooke J. SUS - A quick and dirty usability scale. Usability evaluation in industry. (1996). p. 189.
59. Tullis TS, Stetson JN. A comparison of questionnaires for assessing website usability. In: *Usability professional association conference.* (2004). p. 1–12.
60. Usability Engineering: Book by Jakob Nielsen Nielsen Norman Group. (1994). Available online at: https://www.nngroup.com/books/usability-engineering/ (accessed September 13, 2018).
61. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and Instructions for use. *Int J Human Comput Interaction.* (1995) 7:57–78. doi: 10.1080/10447319509526110
62. Graczyk EL, Delhaye BP, Schiefer MA, Bensmaia SJ, Tyler DJ. Sensory adaptation to electrical stimulation of the somatosensory nerves. *J Neural Eng.* (2018) 15:046002. doi: 10.1088/1741-2552/aab790
63. Bensmaia SJ. Biological and bionic hands: natural neural coding and artificial perception. *Philos Trans R Soc B Biol Sci.* (2015) 370:20140209. doi: 10.1098/rstb.2014.0209
64. Heming E, Sanden A, Kiss ZHT. Designing a somatosensory neural prosthesis: percepts evoked by different patterns of thalamic stimulation. *J Neural Eng.* (2010) 7:064001. doi: 10.1088/1741-2560/7/6/064001
65. Weber DJ, Friesen R, Miller LE. Interfacing the somatosensory system to restore touch and proprioception: essential considerations. *J Motor Behav.* (2012) 44:403–18. doi: 10.1080/00222895.2012.735283
66. Bensmaia SJ, Miller LE. Restoring sensorimotor function through intracortical interfaces: progress and looming challenges. *Nat Rev Neurosci.* (2014) 15:313–25. doi: 10.1038/nrn3724
67. Bensmaia SJ, Tyler DJ, Micera S. Restoration of sensory information via bionic hands. *Nat Biomed Eng.* (2020) 4:1–13. doi: 10.1038/s41551-020-00630-8
68. Ortiz-Catalan M, Hakansson B, Branemark R. An osseointegrated human-machine gateway for long-term sensory feedback and motor control of artificial limbs. *Sci Transl Med.* (2014) 6:257re6. doi: 10.1126/scitranslmed.3008933
69. Andreu D, Guiraud D, Souquet G. A distributed architecture for activating the peripheral nervous system. *J Neural Eng.* (2009) 6:026001. doi: 10.1088/1741-2560/6/2/026001
70. Bisoni L, Carboni C, Raffo L, Carta N, Barbaro M. An HV-CMOS integrated circuit for neural stimulation in prosthetic applications. *IEEE Trans Circuits Syst II Express Briefs.* (2015) 62:184–8. doi: 10.1109/TCSII.2014.2387679
71. Giummarra MJ, Georgiou-Karistianis, Nellie., Nicholls MER, Gibson SJ, Chou, et al., Bradshaw JL. Corporeal awareness and proprioceptive sense of the phantom. *Br J Psychol.* (2010) 101:791–808. doi: 10.1348/000712610X492558
72. Preatoni G, Valle G, Petrini FM, Raspopovic S. Lightening the perceived weight of a prosthesis with cognitively integrated neural sensory feedback. *Curr Biol.* (2021) 31:1–7. doi: 10.1016/j.cub.2020.11.069
73. McGlone F, Wessberg J, Olausson H. Discriminative and affective touch: sensing and feeling. *Neuron.* (2014) 82:737–55. doi: 10.1016/j.neuron.2014.05.001
74. Meisenberg G, Williams A. Are acquiescent and extreme response styles related to low intelligence and education? *Personal Individual Differ.* (2008) 44:1539–50. doi: 10.1016/j.paid.2008.01.010
75. Valle G, Iberite F, Strauss I, D'Anna E, Granata G, Stieglitz T, et al. A psychometric platform to collect somatosensory sensations for neuroprosthetic use. *bioRxiv.* (2020). doi: 10.1101/2020.07.23.218222

Conflict of Interest: FP, SR, and SM hold shares of Sensars Neuroprosthetics Sarl, a start-up company dealing with potential commercialization of neurocontrolled artificial limbs.

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 Public Health

BRIEF RESEARCH REPORT
published: 12 April 2021
doi: 10.3389/fpubh.2021.668351

Fatty Liver Disease Prediction Model Based on Big Data of Electronic Physical Examination Records

Mingqi Zhao[1], *Changjun Song*[2], *Tao Luo*[1], *Tianyue Huang*[3] *and Shiming Lin*[2,3]*

[1] School of Mathematical Sciences Xiamen University, Xiamen, China, [2] Department of Computer Engineering, Changji University, Changji, China, [3] School of Informatics Xiamen University (National Demonstrative Software School), Xiamen, China

Fatty liver disease (FLD) is a common liver disease, which poses a great threat to people's health, but there is still no optimal method that can be used on a large-scale screening. This research is based on machine learning algorithms, using electronic physical examination records in the health database as data support, to a predictive model for FLD. The model has shown good predictive ability on the test set, with its AUC reaching 0.89. Since there are a large number of electronic physical examination records in most of health database, this model might be used as a non-invasive diagnostic tool for FLD for large-scale screening.

Keywords: fatty liver disease, electronic medical records, genetic algorithm, machine learning, XGBoost, chi-square binning algorithm

OPEN ACCESS

Edited by:
Yonghong Peng,
Manchester Metropolitan University, United Kingdom

Reviewed by:
Hua Su,
Fuzhou University, China
By C,
Hebei University of Economics and Business, China

***Correspondence:**
Shiming Lin
xmulsm@xmu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *16 February 2021*
Accepted: *11 March 2021*
Published: *12 April 2021*

Citation:
Zhao M, Song C, Luo T, Huang T and Lin S (2021) Fatty Liver Disease Prediction Model Based on Big Data of Electronic Physical Examination Records. Front. Public Health 9:668351. doi: 10.3389/fpubh.2021.668351

1. INTRODUCTION

Fatty liver disease (FLD) is a lesion with excessive accumulation of fat in liver cells, which is divided into non-alcoholic fatty liver disease (NAFLD) and alcoholic fatty liver disease (AFLD) (1). In recent years, with the improvement of living standards, changes in lifestyle and diet, and the wide use of ultrasound and other imaging technology, the prevalence of FLD is growing rapidly (2). In fact, it has become the most common cause of chronic liver disease in developed and developing countries (3). According to research, about 25% of people worldwide and 21% of people in China catch NAFLD (4, 5).

At present, the pathogenesis of NAFLD is not completely clear, and there is no ideal and effective treatment drug, but it is reversible in the early stages. Research shows that effective lifestyle interventions such as energy restriction, dietary changes, and increased physical activity are particularly effective in the early stages of NAFLD (6). Therefore, early detection and treatment is the key. At present, the main clinical diagnostic methods are ultrasound, CT, and liver biopsy (7). For their invasiveness and complexity, they are not suitable for large-scale epidemiological screening (8–10).

Based on the above situation, many scientists try to use machine learning algorithm to build the prediction model of FLD. In recent years, several machine learning models based on medical data have been proposed (11–13). Italian scholar Giorgio Bedogni collected data by gender, age, alcohol intake, alanine aminotransferase, aspartate aminotransferase, body mass index (BMI), waist circumference, the sum of four skinfolds, etc., and established a prediction model for NAFLD (13). However, most of the models are carried out through questionnaire surveys and medical experiments and use some features that are not easy to obtain in large quantities. The limitation of data quantity and the complexity of features make these models difficult to generalize.

The purpose of this study is to establish an efficient and convenient FLD prediction model using machine learning algorithm which can help doctors to screen out the patients that need further liver examination and can be applied to large-scale epidemiologic screening. To facilitate the generalization of the model, the features we use will be as convenient as possible, and the amount of data we use will be as much as possible.

2. MATERIALS AND METHODS

2.1. Dataset

The development of the medical system, the popularity of electronic physical examination records, and the establishment of health databases provide data support for large-scale epidemiological research. The data set used in this study is from the health database of a hospital in China, which contains the electronic physical examination records of 44,854 patients. And in this data set, no patient's privacy information is included, only routine physical examination data and age are included. To simplify and generalize the model, we only extracted 129 routine physical examination items of all patients, including blood routine, biochemistry, urine routine, etc.

In this study, patients diagnosed with FLD by ultrasound were marked as 1, and the remaining patients were marked as 0. The prevalence of FLD in the data set is 23%, which is close to the previous research (5).

2.2. Data Preprocessing

Firstly, for the accuracy of the model, we deleted individuals who had not undergone ultrasound examination because we did not know if they had FLD. Then, we delete the items with more than $\frac{2}{3}$ missing values that most people have not been examined. Finally, we randomly select 70% of the data set as the training set of the model and 30% as the test set.

Figure 1 shows the process of data preprocessing. **Figure 2** shows the mean (standard deviation) of the different features of FLD patients and normal people and whether these features have passed the chi-square test with significance level of 0.05. It can be seen that there are significant differences in Male gender percentage, Uric acid (UA), Triglycerides (TG), Alanine

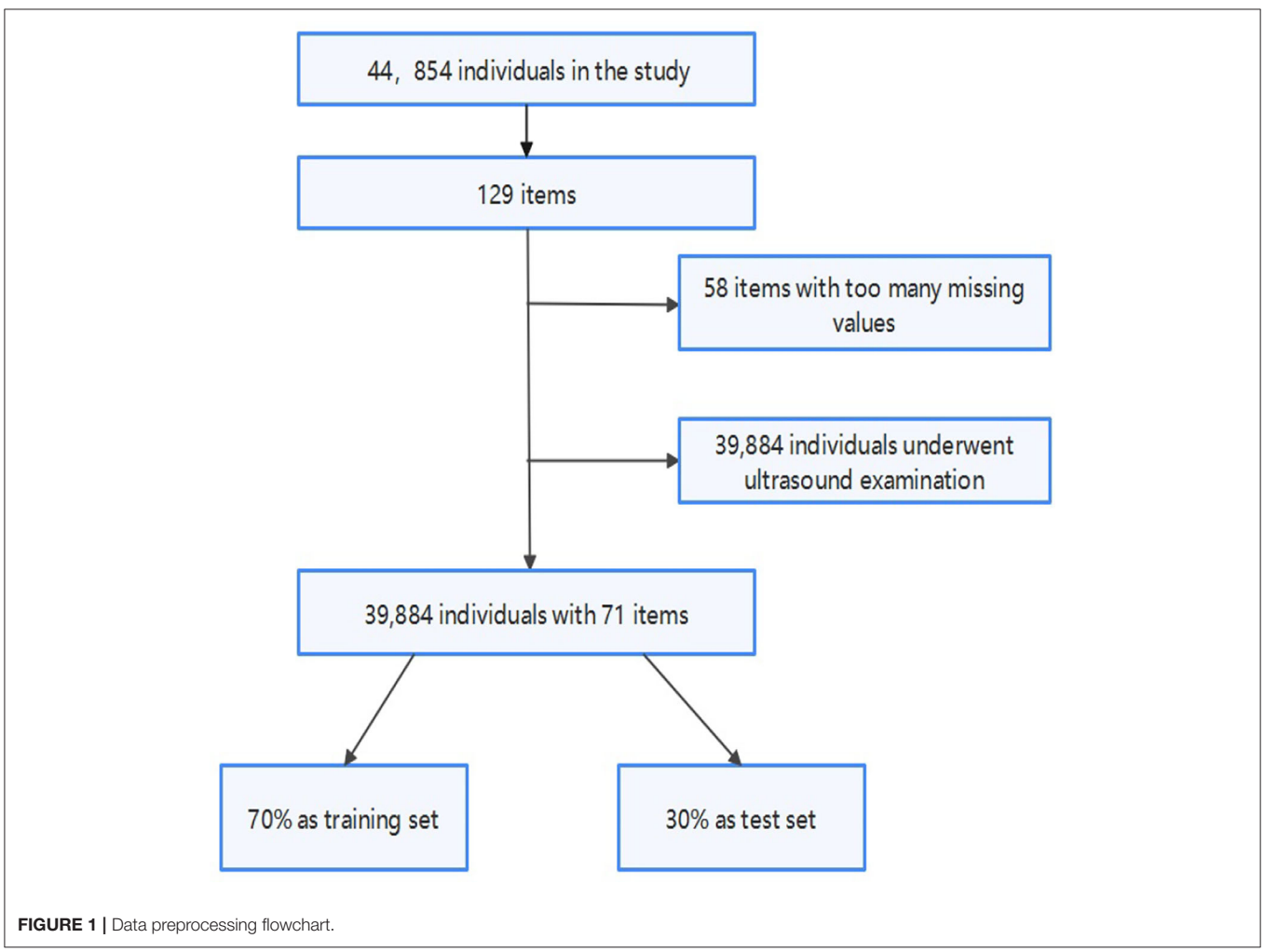


FIGURE 1 | Data preprocessing flowchart.

Feature	No FLD	FLD	Pass the chi-square test
Male gender percentage	56.8	82.2	NO
Age	36.1(13.9)	43.6(13.7)	NO
Gamma glutamine transpeptidase (GGT)	22.7(22.3)	40.6(33.3)	NO
Triglycerides (TG)	1.08(0.908)	2.07(1.65)	NO
Alanine aminotransferase (ALT)	22.8(19.8)	39.6(25.4)	NO
Uric acid(UA)	330(79.7)	392(79.5)	NO
AST/ALT	1.11(0.333)	0.835(0.332)	NO
Aspartate aminotransferase (AST)	22.8(12.6)	28.1(17.2)	NO
...	...	...	...
Carbon dioxide (CO2)	26.2(2.26)	26.2(2.10)	YES
Total bilirubin (TBIL)	18.5(6.73)	18.9(6.24)	YES
Total protein (TP)	72.8(4.81)	73.7(4.93)	YES
Anion gap	14.0(2.56)	14.8(2.48)	YES

FIGURE 2 | Statistical information and chi-square test results of different features in different groups.

aminotransferase (ALT), Aspartate aminotransferase (AST), Gamma glutamine transpeptidase (GGT), Age and AST/ALT between FLD patients and normal people, while Carbon dioxide (CO_2), Total bilirubin (TBIL), Total protein (TP), and Anion gap do not.

2.3. Missing Value Processing

Compared with conducting medical experiments and questionnaire surveys, the advantage of using electronic physical examination records in the health database for modeling is that the amount of data is large and the model is easy to be generalized, but the disadvantage is that there are lots of missing values. Therefore, how to fill in missing values is critical to modeling. The usual practice is to fill in the mean or median for missing values. In fact, the distribution of medical indicators varies with gender and age, and the range is large. So it's a good choice to fill in the median according to age and gender.

For age grouping method, standard age grouping can be used, but the result is not ideal. So we use the chi-square binning algorithm to group age. Chi-square binning algorithm is a binning algorithm based on the chi-square test, which is specifically implemented by the independence test in the chi-square test. The theoretical basis for binning is: the lower the chi-square value between two bins, the more likely they are to have similar distributions (14). If two adjacent bins have very similar distributions, then the two bins should be merged, otherwise, they should be separated. Therefore, in each step of the algorithm, the two bins with the smallest chi-square value must be combined until the number of bins meets the stopping condition.

In the present study, a bin refers to an age group and distribution refers to the prevalence of FLD. And we set the expected number of bins to 5, and the result after calculation on the training set is: [0, 17], (17, 29], (29, 35], (35, 47], (47, 197]. According to the results of age grouping, **Figure 3** shows the distribution of several important features that need to be filled with missing values under different age and gender groups. It can be seen that the difference in distribution is obvious, so our strategy of filling in missing values is necessary and effective.

2.4. Feature Engineering

In machine learning modeling, the quality of features often determines the upper bound of model performance. Therefore, we need to do feature engineering on the existing routine features to maximize the usage of them. In clinical diagnosis, the combination of multiple characteristics often plays an important role in the judgment of diseases. For example, AST/ALT (Aspartate aminotransferase/Alanine aminotransferase) is of great significance in the diagnosis of liver diseases (1). So we want to generate new features through a combination of features.

In the present study, we use Spearman's correlation coefficient as a standard to measure the quality of features and use the genetic algorithm to find the optimal solution. Spearman's correlation coefficient, also known as rank correlation coefficient, can measure the rank correlation between two variables. If the machine learning model used is based on a decision tree, the Spearman correlation coefficient can measure the correlation between a feature and the target. The genetic algorithm is a method of searching for the optimal solution by simulating the natural evolution (15, 16). The algorithm transforms the problem-solving process into a process similar to the crossover and mutation of chromosomal genes in biological evolution. When solving more complex combinatorial optimization problems, Compared with some conventional optimization algorithms, it can usually obtain better optimization results faster (16).

Figure 4 shows the process of feature engineering using genetic algorithm. In the algorithm, an individual in the population is defined as a binary tree. Each leaf node of the binary tree is a certain feature in the data set, and each inner node of the binary tree is an operator in {+, -, *, /, log, sqrt}. Each individual represents an expression composed of features and operators. Fitness is the Spearman correlation coefficient between the new feature and the target. In each generation, individuals with high fitness will be retained, and individuals with low fitness will be eliminated. The upper left of **Figure 5** shows an individual example, which represents $TG * AST + GLU$. The upper right and lower parts of **Figure 5** respectively show crossover operations and mutation operations, both of which generate new individuals by changing subtrees in the way that simulates biological variation.

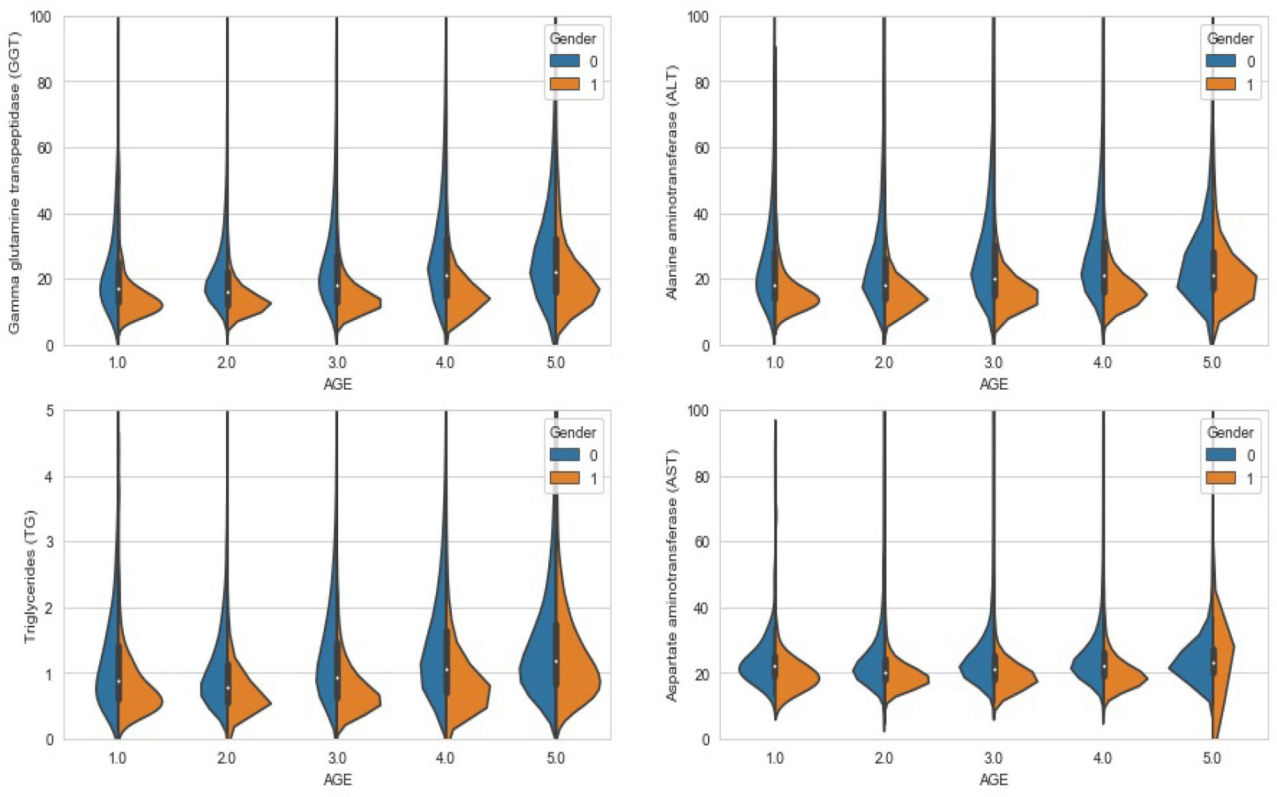


FIGURE 3 | Violin chart: the distribution of different features under different age groups and genders.

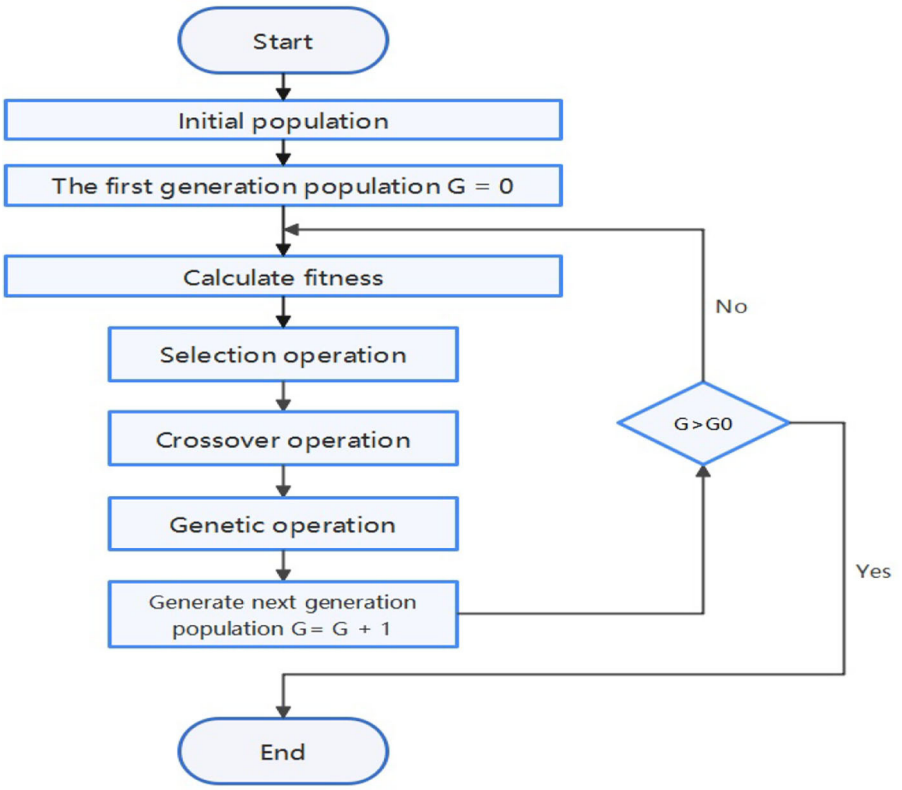


FIGURE 4 | Genetic algorithm flowchart.

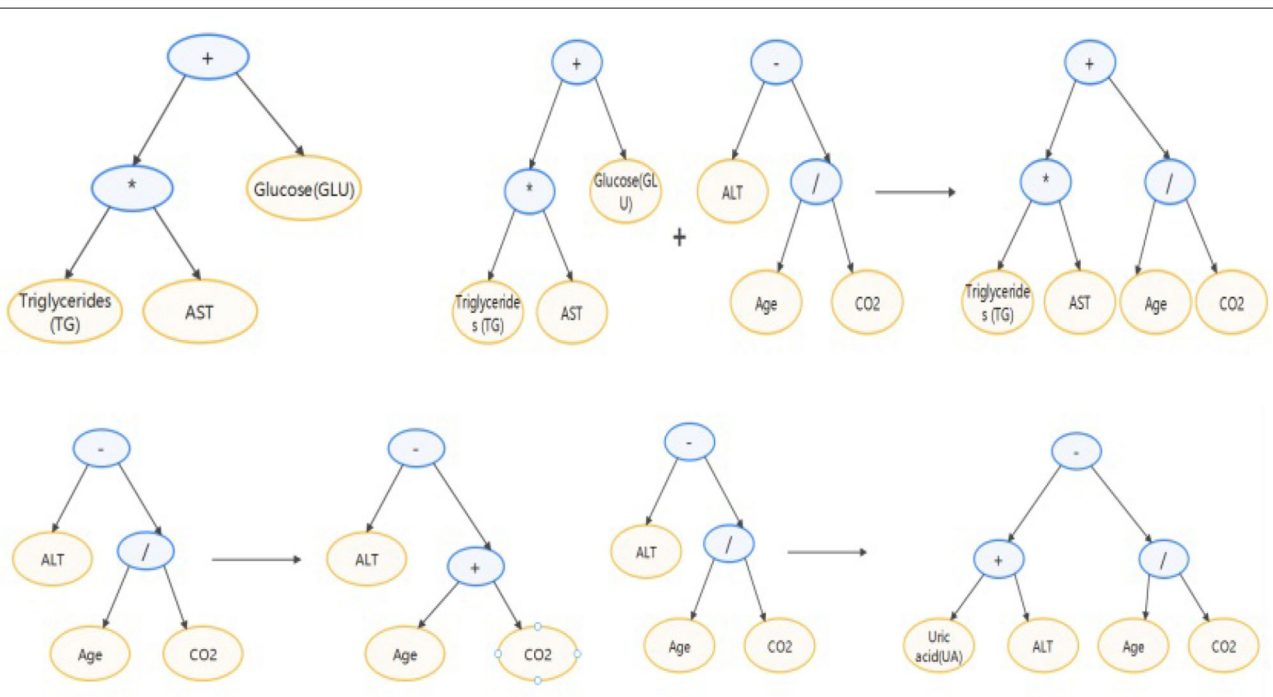


FIGURE 5 | Demonstration of individual and individual variation.

We set the number of individuals in each generation to 1,000 and set the maximum depth of the binary tree to three. Use the normalized features and iterating ten generations, the individuals with the first three fitness levels are added to the data set as new features. The result is:$GA\_fea1 = TG + log(ALT)$ *with fitness* 0.89, $GA\_fea2 = TG * GGT$ *with fitness* 0.87, and $GA\_fea3 = (UA + AST) * log(ALT)$ *with fitness* 0.79.

3. EXPERIMENTS AND RESULTS

XGBoost (eXtreme Gradient Boosting) is an engineering implementation of gradient boosting decision tree (GBDT) (17). Its core idea is to perform a second-order Taylor expansion of the loss function, and gradually train the decision tree based on the objective function, and greatly improve the training model speed (18, 19). XGboost has many advantages. For example, traditional GBDT only uses first-order derivative information in optimization, while XGboost performs a second-order Taylor expansion on the cost function to make the result more accurate. Xgboost adds a regular term to the cost function to control the complexity of the model, which reduces the variance of the model and makes the learned model simpler and prevents overfitting. XGboost supports parallel computing on feature granularity, which greatly reduces the amount of calculation and improves the training speed. In addition, XGBoost is a model based on the decision tree model, it is more explanatory than neural networks and other algorithms, which can enable us to better understand how a physical examination data plays a role in the model (20). Therefore, the present study uses the XGBoost model for modeling.

The error of a machine learning model includes two aspects: variance and bias (21). High bias models usually have relatively simple parameter settings and tend to underfit, that is, there is little difference in performance between the training set and test set, but both are relatively low. High variance models usually have complex parameter settings and tend to overfit. They perform well on the training set, but the performance on the test set drops seriously. The usual practice is to make a trade-off between variance and bias to get a reasonable model. AUC (Area Under Curve) is defined as the area under the ROC curve (Receiver Operating Characteristic curve), which is a commonly used indicator to measure the quality of a machine learning model (22). AUC has nothing to do with the ratio of positive and negative samples, it represents the model's ability to sort samples to a certain extent (23). In present study, we use AUC as the evaluation criterion of the XGBoost model. On the training set, Bayesian optimization of hyperparameters is performed using triple cross-validation, and then the obtained results are fine-tuned to prevent over-fitting and ensure the rationality of the parameters. The main results are as follows: *max_depth* : 3, *learning_rate* : 0.07, *n_estimators* : 150, *scale_pos_weight* : 2, *min_child_weight* : 6, *gamma* : 0.2, *reg_alpha* : 0.1.

The upper left and upper right of **Figure 6** respectively show the performance of the high variance model and the high bias model. The lower left shows the effect of the hyperparameter *iterations* on the model performance. It can be seen that with the increase of *iterations*, the over fitting phenomenon of the model appears, and the variance of the model becomes larger. The lower right shows the performance of the model with the optimal hyperparameter combination set. It can be seen that the AUC of the model reached 0.89, which shows that the model has a strong predictive ability for FLD, and the performance of the model in the test set and training set is basically the same, without over fitting phenomenon.

4. DISCUSSION

Using the number of times the feature is used as the basis for splitting in the decision tree splitting as the importance of the feature, we can sort all the features by importance. Left of **Figure 7** shows the model performance obtained by gradually adding the top 60 features of importance to the model. It can be seen that the top 10 features are the most important, and the features after the 20th place are dispensable. This shows that even if we only use the first ten features to train the model, its AUC can still reach the level of 0.87–0.88, but the model is greatly simplified at this time.

Right of **Figure 7** shows the importance of the top 10 features. According to research, the degree of fat accumulation in the liver is directly proportional to body weight. The prevalence of obesity in NAFLD patients is estimated to be 51.34% (95% CI: 41.38–61.20) (1), so many FLD patients have a significant increase in TG. At the same time, and when liver disease occurs, ALT and GGT will increase significantly. Right of **Figure 7** shows that TG, ALT, GGT, *GA_fea*1, and *GA_fea*2 play a vital role in the model, which is in line with the facts. Studies have also shown that the prevalence of diabetes in NAFLD patients is estimated to be 22.51% (95%CI: 17.92–27.89) (1), and with the increase of age, people's metabolism slows down and people are more likely to suffer from metabolic diseases. So the importance of GLU and Age is also well-understood.

We analyzed the patients with FLD who were mispredicted in the test set and found that their indicators were basically normal. We think that these people may be patients with AFLD or patients with mild FLD, they often do not have obvious symptoms and indicators change (1). Our data set does not include the alcohol intake and body condition of patients, which limits our prediction ability, because we can not exclude the interference of AFLD and we can not use the waist circumference of patients to judge whether they are obese(Even so, the AUC of our model is still high). But because of this, our model can be directly applied to the electronic physical examination records of the current health database for large-scale epidemics screening.

5. CONCLUSION

In the present study, we use the electronic physical examination records in the health database as data support, use the chi-square binning algorithm to help fill in the missing values, and use the genetic algorithm as the optimization algorithm for feature engineering, which tentatively solves the two disadvantages of the large-scale electronic medical record–missing values and

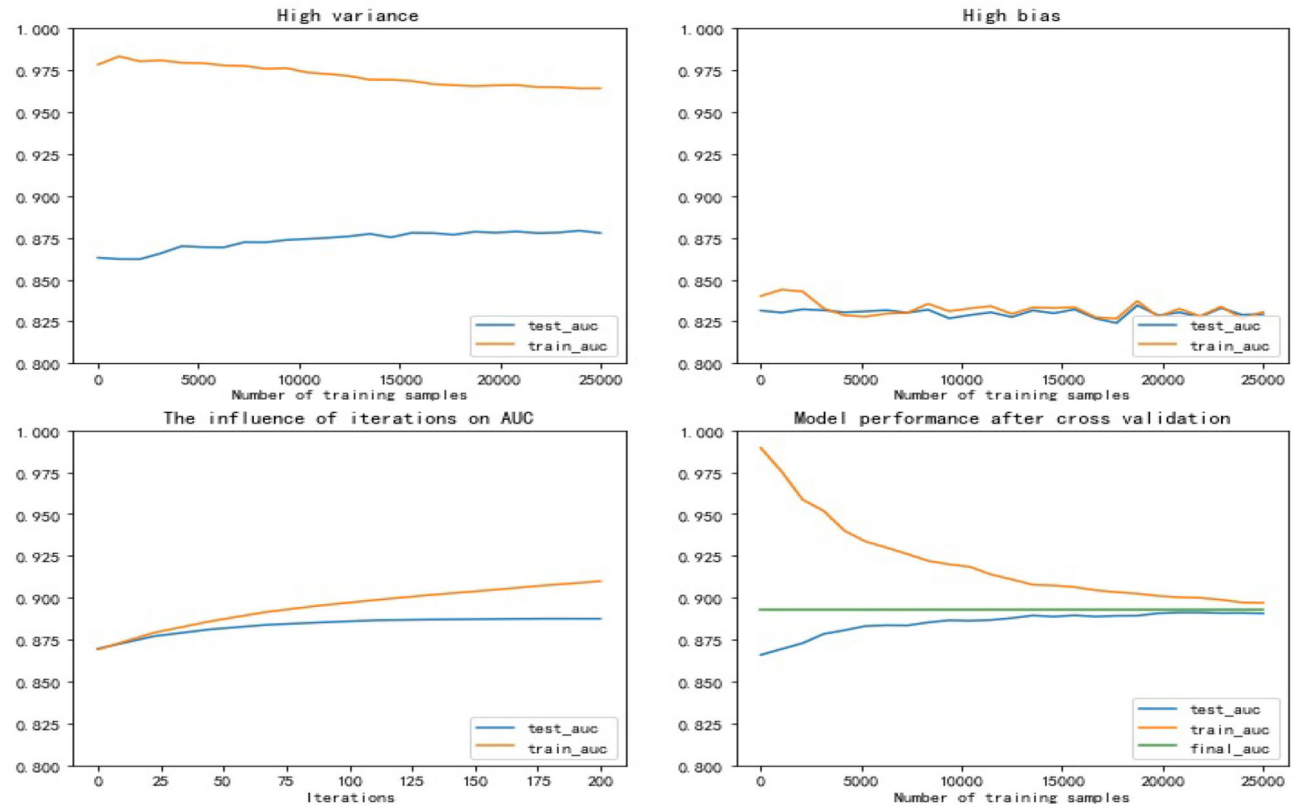


FIGURE 6 | Trade-off between variance and bias.

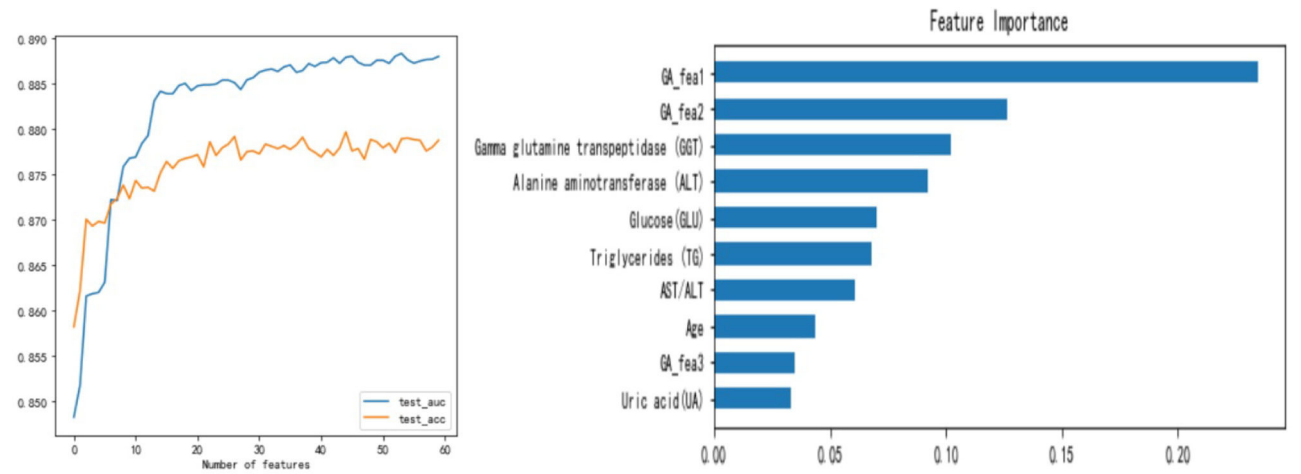


FIGURE 7 | The influence of the number of features used on the model and feature importance.

lack of features. In the end, this study established an FLD prediction model based on the XGBoost algorithm with an AUC of 0.89. The satisfactory performance of the model makes large-scale screening of FLD possible, but due to the limited data breadth, more data is needed for external verification before applications.

DATA AVAILABILITY STATEMENT

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

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 to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was not obtained from the individual(s), nor the minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

MZ puts forward the idea and realizes it, completes the visualization and the writing of the paper. CS is responsible for data sorting and data analysis. TL and TH assisted in writing the first draft of the paper. SL leads research and conduct thesis writing guidance. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the study of liver diseases. *Hepatology*. (2018) 67:328–57. doi: 10.1002/hep.29367
2. Brunt EM, Wong VW, Nobili V, Day CP, Sookoian S, Maher JJ, et al. Nonalcoholic fatty liver disease. *Nat Rev Dis Primers*. (2015) 1:15080. doi: 10.1038/nrdp.2015.80
3. Bellentani S. The epidemiology of non-alcoholic fatty liver disease. *Liver Int*. (2017) 37:81–4. doi: 10.1111/liv.13299
4. Younossi Z, Koenig A, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*. (2016) 64:73–84. doi: 10.1002/hep.28431
5. Li Z, Xue J, Chen P, Chen L, Yan S, Liu L. Prevalence of nonalcoholic fatty liver disease in mainland of China: a meta-analysis of published studies. *J Gastroenterol Hepatol*. (2014) 29:42–51. doi: 10.1111/jgh.12428
6. El-Agroudy N, Kurzbach A, Rodionov R, O'Sullivan J, Roden M, Birkenfeld A, et al. Are lifestyle therapies effective for NAFLD treatment? *Trends Endocrinol Metabol*. (2019) 30:701–9. doi: 10.1016/j.tem.2019.07.013
7. Mishra P, Younossi ZM. Abdominal ultrasound for diagnosis of non alcoholic fatty liver disease (NAFLD). *Am J Gastroenterol*. (2007) 102:2716–7. doi: 10.1111/j.1572-0241.2007.01520.x
8. Noureddin M, Lam J, Peterson MR, Middleton M, Hamilton G, Le TA, et al. Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. *Hepatology*. (2013) 58:1930–40. doi: 10.1002/hep.26455
9. Sumida Y, Nakajima A, Itoh Y. Limitations of liver biopsy and non invasive diagnostic tests for the diagnosis of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World J Gastroenterol*. (2014) 20:475–85. doi: 10.3748/wjg.v20.i2.475
10. Zeng N, Li H, Wang Z, Liu W, Liu S, Alsaadi F, et al. Deep-reinforcement-learning-based images segmentation for quantitative analysis of gold immunochromatographic strip. *Neurocomputing*. (2021) 425:173–80. doi: 10.1016/j.neucom.2020.04.001
11. Yip T, Ma A, Wong V, Tse Y, Chan H, Yuen P, et al. Laboratory parameter based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. *Aliment Pharmacol Ther*. (2017) 46:447–56. doi: 10.1111/apt.14172
12. Poynard T, Ratziu V, Naveau S, Thabut D, Charlotte F, Messous D, et al. The diagnostic value of biomarkers (SteatoTest) for the prediction of liver steatosis. *Comp Hepatol*. (2005) 4:1–14. doi: 10.1186/1476-5926-4-10
13. Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol*. (2006) 6:33. doi: 10.1186/1471-230X-6-33
14. Franke TM, Ho T, Christie CA. The chi-square test: often used and more often misinterpreted. *Am J Eval*. (2012) 33:448–58. doi: 10.1177/1098214011426594
15. Zeng N, Song D, Li H, You Y, Liu Y, Alsaadic F. A competitive mechanism integrated multi-objective whale optimization algorithm with differential devolution. *Neurocomputing*. (2021) 432:170–82. doi: 10.1016/j.neucom.2020.12.065
16. Mitchell M. *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press (1998).
17. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD'16 - XGBoost*. San Francisco, CA: ACM Press (2016). p. 785–94. doi: 10.1145/2939672.2939785
18. Dietterich TG. Ensemble methods in machine learning. In: Kittler J, Roli F, editors. *International Workshop on Multiple Classifier Systems*. Berlin; Heidelberg: Springer (2000). p. 1–15.
19. Zopluoglu C. Detecting examinees with item preknowledge in large-scale testing using extreme gradient boosting (XGBoost). *Educ Psychol. Meas.* (2019) 79:13164419839439. doi: 10.1177/0013164419839439
20. Zeng N, Wang Z, Zineddin B, Li Y, Du M, Xiao L, et al. Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach. *IEEE Trans Med Imaging*. (2014) 33:1129–36. doi: 10.1109/TMI.2014.2305394
21. Mehta P, Bukov M, Wang C-H, Day AG, Richardson C, Fisher CK, et al. *A High-Bias, Low-Variance Introduction to Machine Learning for Physicists*. (2018). Available online at: https://arxiv.org/abs/1803.08823
22. McClish DK. Analyzing a portion of the ROC curve. *Med Decis Making*. (1989) 9:190–5.
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. (1988) 44:837–45.

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 Public Health

BRIEF RESEARCH REPORT
published: 07 May 2021
doi: 10.3389/fpubh.2021.669190

Adaptive Sparse Detector for Suppressing Powerline Component in EEG Measurements

Bin-qiang Chen[1], *Bai-xun Zheng*[1], *Chu-qiao Wang*[1] *and Wei-fang Sun*[2]*

[1] School of Aerospace Engineering, Xiamen University, Xiamen, China, [2] College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou, China

Powerline interference (PLI) is a major source of interference in the acquisition of electroencephalogram (EEG) signal. Digital notch filters (DNFs) have been widely used to remove the PLI such that actual features, which are weak in energy and strongly connected to brain states, can be extracted explicitly. However, DNFs are mathematically implemented via discrete Fourier analysis, the problem of overlapping between spectral counterparts of PLI and those of EEG features is inevitable. In spite of their effectiveness, DNFs usually cause distortions on the extracted EEG features, which may lead to incorrect diagnostic results. To address this problem, we investigate an adaptive sparse detector for reducing PLI. This novel approach is proposed based on sparse representation inspired by self-adaptive machine learning. In the coding phase, an overcomplete dictionary, which consists of redundant harmonic waves with equally spaced frequencies, is employed to represent the corrupted EEG signal. A strategy based on the split augmented Lagrangian shrinkage algorithm is employed to optimize the associated representation coefficients. It is verified that spectral components related to PLI are compressed into a narrow area in the frequency domain, thus reducing overlapping with features of interest. In the decoding phase, eliminating of coefficients within the narrow band area can remove the PLI from the reconstructed signal. The sparsity of the signal in the dictionary domain is determined by the redundancy factor. A selection criteria of the redundancy factor is suggested via numerical simulations. Experiments have shown the proposed approach can ensure less distortions on actual EEG features.

Keywords: EEG, spare representation, fourier transform, powerline interference, basis pursuit

OPEN ACCESS

Edited by:
Yonghong Peng,
Manchester Metropolitan University,
United Kingdom

Reviewed by:
Wei Cao,
Huaqiao University, China
Wenrong Xiao,
China Three Gorges University, China

***Correspondence:**
Wei-fang Sun
swf@wzu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *18 February 2021*
Accepted: *06 April 2021*
Published: *07 May 2021*

Citation:
Chen B-q, Zheng B-x, Wang C-q and Sun W-f (2021) Adaptive Sparse Detector for Suppressing Powerline Component in EEG Measurements. Front. Public Health 9:669190. doi: 10.3389/fpubh.2021.669190

INTRODUCTION

Electroencephalography (EEG) aims at measuring potentials that reflect the electrical activity of the human brain (1). It has been recognized as a powerful tool in psychophysiology due to its high temporal resolution and sensitivity to index different functional brain states (2). However, because of imperfect measurement conditions, noises are likely to be incorporated in the records of EEG. For instance, EEG signals in actual measurements can often be exposed to strong powerline interferences (PLIs) at 50 or 60 Hz, which is originated from AC power (3). In laboratory environments, good shielding measures, such as shielded rooms, can be helpful to reduce the influence of PLI. But shielding measures are usually impractical for healthcare practices of EEG monitoring via mobile instrument such as wearable devices (4, 5).

In practical applications, solutions to remove the PLI can be roughly categorized into two types, the hardware ways and the software ways (6). The hardware ways are referred to active electrodes with integrated analog filters, but a significant residual interference still remains. In contrast, the software based solutions are more flexible and they can be combined with hardware based techniques. To achieve a more robust performance of the PLI removal, novel approaches based on digital signal processing techniques are more popular and have been extensively studied (7–9). The digital notch filter (DNF) at the powerline frequency, which is designed based on the theory of traditional Fourier analysis, has been widely employed to reduce PLI in biomedical measurements (10). In spite of its effectiveness, side effects of the DNF are also reported in related researches. Firstly, the stop band of the DNF is difficult to be selected. Although the main frequency of PLI stays at 50 or 60 Hz nominally, there is always a variation of ±2Hz in the actual power system. Secondly, the sampling parameters also affect the filtering results. A phenomenon of energy leakage is likely to be produced by sinusoidal waves whose sampling does not satisfy the full period sampling condition (11). Thirdly, the actual EEG features cover a comparatively broad range in the frequency domain. Thus, overlapping between the spectral counterparts of PLI and those of the EEG features is inevitable. As a result, DNF is not a perfect tool for removing PLI.

Multiresolution analysis (MRA) is a powerful tool to characterize non-stationary and transient components. The past three decades have witnessed the rapid development of MRA. As a concrete example of MRA, discrete wavelet transform (DWT) is implemented using a scaling function and a wavelet function (12). DWT has been widely utilized to separate signal components in many scientific researches. On the other hand, flexible DWTs, based on parameter optimization schemes, are also employed to extract transient features (13). In the field of biomedical signal processing, wavelet transform can reveal weak features related to the transient nature of biomedical measurements which are not so obvious by using spectrum analysis. In the literature, it has been reported that DWT can be used as an effective tool in dealing noises such as PLI and baseline wandering from a corrupted biomedical measurement (14). DWT can be also combined with artificial intelligence methods (15, 16), such as deep learning and support vector machine, to realize signal classification applications. However, in suppressing the PLI components, which are more similar to stationary contents in waveform, DWT may be not perfect. Recently, compressed sensing and sparse representation, which still rely on the idea of signal representation, emerge as enhancements to the conventional DWT.

In this paper, to achieve a better PLI suppressing performance, we proposed a novel adaptive detector (ASD). It can be regarded as an improvement to fast Fourier transform, which is a classical spectral analysis tool. Within the proposed method, a redundant Fourier dictionary, developed from the orthogonal Fourier basis, is employed as an over-complete dictionary. The linear combination coefficients with respect to the redundant dictionary are optimized to achieve a sparse representation. Therefore, a sinusoidal wave can be expressed as a combination of a very limited number of sinusoidal atoms. Due to the narrow band property of PLI in the spare representation, they can be more easily isolated from other contents. The performance of the proposed method is verified by numerical simulations and a case study of actual EEGs.

MATERIALS AND METHODS

Signal Modeling and Notations

For the convenience of discussion, we use the following notations for mathematical argument. Let $\{x(n)|,\ n \in \mathbb{Z}^+ \bigcup \{0\}\}$ be a corrupted EEG signal containing both the actual brain potential waveform $s(n)$ and the PLI component $p(n)$. That is to say, $x(n) = s(n) + p(n)$. Let the sampling length and the sampling frequency of $x(n)$ be denoted as N and f_s. The estimated signal after PLI cancellation using some specific algorithm is expressed as $\tilde{s}(n)$. In this article, $p(n)$ is modeled as a simple harmonic wave characterized by the harmonic parameters of amplitude (Amp_{pli}), frequency (f_{pli}), and initial phase (φ_{pli}).

$$p(n) = Amp_{pli} \cdot \cos(2\pi f_{pli}(n-1)\Delta t) + \varphi_{pli}) \tag{1}$$

where the time interval $\Delta t = 1/f_{pli}$. To remove PLI, an important task lies in accurate estimation of these harmonic parameters such that a compensation signal can be reconstructed. The actual EEG waveform $s(n)$ is by nature non-stationary and its spectrum covers a comparatively wide range of area in the spectral domain, which is essentially different from that of $p(n)$.

In optimization theory, the norms of $x(n)$ are indispensable. Two types of commonly used norms are the ℓ_1 and ℓ_2 norms. They are computed using the following formulae.

$$\begin{aligned} \ell_1\ norm:\ ||x||_1 &:= \sum_{i=0}^{N-1} \left|x(i)\right| \\ \ell_2\ norm:\ ||x||_2 &:= \left(\sum_{i=0}^{N-1} \left|x(i)\right|^2\right)^{1/2} \end{aligned} \tag{2}$$

The ℓ_1 is essential to ensure sparse representation using redundant dictionaries. The ℓ_2 is strongly connected to least squares approximation, and it is also known as the Euclidean distance.

Notch Filters Based on Fourier Transform Theory

Fast Fourier transform (FFT) is a fundamental tool for discrete signal analysis. It is an orthogonal decomposition of input signal $x(n)$ of length N via a orthonormal basis $\Phi_{Fourier}$.

$$\Phi_{Fourier} = \frac{1}{\sqrt{N}} [\phi_0\ \phi_1\ \ldots \phi_{N-1}] \tag{3}$$

The k-th sinusoidal atom in the Fourier dictionary can be defined as

$$\phi_k = \left(e^{\mathrm{j}\cdot\frac{2\pi}{N}\cdot k\cdot 0},\ e^{\mathrm{j}\cdot\frac{2\pi}{N}\cdot k\cdot 1},\ e^{\mathrm{j}\cdot\frac{2\pi}{N}\cdot k\cdot 2},\ \ldots,\ e^{\mathrm{j}\cdot\frac{2\pi}{N}\cdot k\cdot (N-1)}\right)^{\mathrm{T}} \tag{4}$$

where $\mathrm{j} = \sqrt{-1}$ is defined as the imaginary number and T means the transpose operation in matrix algebra. That is to say, the frequency of the sinusoidal atom ϕ_k is $2\pi(k-1)/N$. A uniform step of $2\pi/N$ is assumed for angular frequencies of two adjacent atoms in the Fourier dictionary. The transform coefficient is computed using the following formula

$$c_k = \phi_k^H x, \tag{5}$$

where H means conjugate transpose in complex analysis. The representation coefficients of a input signal x with respect to the dictionary $\Phi_{Fourier}$ is expressed using the matrix operation of $\Phi_{Fourier}^H x$. By using the above transformation, the input signal is expressed as the sum of a few sinusoidal waves with different frequencies, which is expressed as

$$x = \sum_{k=0}^{N-1} c_k \phi_k. \tag{6}$$

The above Equation is also named as the inverse fast Fourier transform (IFFT). For a sinusoidal wave whose sampling parameters satisfy the full period sampling condition, it is mapped to a single spectral line (11). Otherwise the phenomenon of energy leakage occurs and causes dense representation of the sinusoidal wave in the frequency domain.

The method of digital notch filter is theoretically based on the Fourier transform theory. Either finite impulse response (FIR) filters or infinite impulse response (IIR) filters can be designed for conducting PLI removal applications. Some mature development toolkit for designing notch filters can be found in commercial software of numerical computation. By specifying a few parameters, notch filters can be designed conveniently. Despite the effectiveness reported in the literature, these filters also suffer a few drawbacks and may cause distortions on the extracted EEG features. Owing to frequency variations existed in the PLI, it is not possible to design a uniform DNF with pre-determined parameters. Hence, during the implementation of DNFs, strategies allowing self-adaptivity of parameters should be considered.

Redundant Fourier Dictionary for Spectral Analysis

In the FFT spectrum, the spectral lines are sampled at the frequency of $k \cdot f_s/N$, in which $k = 0, 1, \ldots, N-1$. The interval between adjacent spectral lines are uniform. Just as the digital signal can be regarded as a sampling of the analog signal, the FFT spectrum can be also interpreted as a sampling of the continuous Fourier spectrum. For an arbitrary frequency of f_a, the associated Fourier coefficient can be calculated using the following formula

$$X(f_a) = \sum_{k=0}^{N-1} x(k) e^{-\mathrm{j}\cdot 2\pi f_a k \Delta t}, \tag{7}$$

in which $\Delta t = 1/f_s$. Besides the FFT sampling grids mentioned above, additional Fourier coefficients can be computed in order to allow insightful investigations. Because redundancy is introduced in signal representation, the relevant dictionary is called as redundant Fourier Dictionary (RFD). A few efficient implementation using RFD have been developed, such as Goertzel algorithm (17), chirp-Z transform (18) and zero padded FFT (19). Among these implementations, the combination of zero padding and FFT algorithm is usually adopted for computing redundant Fourier spectra, in which the RFD is composed of uniformly spaced spectral bins. Although redundancy is beneficial in revealing information in the frequency domain, the corresponding representations are likely to be dense. Therefore, post-processing steps are usually required to estimate harmonic parameters of PLI.

In this paper, in order to sparsely represent a signal, a tight dictionary composed of redundant Fourier atoms is employed. For a N-point digital signal, a tight Fourier dictionary A containing K atoms ($K > N$) is defined as

$$A = \left(\tilde{\phi}_0 \ \tilde{\phi}_1 \ \ldots \ \tilde{\phi}_{K-1}\right), \tag{8}$$

in which the angular frequency associated with the sinusoidal atom $\tilde{\phi}_k$ is $2\pi k/N$. It can be regarded as a representation matrix of mapping: $\mathbb{C}^N \mapsto \mathbb{C}^K$. Different from ϕ_k, the definition of $\tilde{\phi}_k$ is given as

$$\phi_k = \left(e^{j\cdot\frac{2\pi}{K}\cdot k\cdot 0}, \ e^{j\cdot\frac{2\pi}{K}\cdot k\cdot 1}, \ e^{j\cdot\frac{2\pi}{N}\cdot k\cdot 2}, \ \ldots, \ e^{j\cdot\frac{2\pi}{K}\cdot k\cdot(N-1)}\right)^T. \tag{9}$$

The vectors x, ϕ_k, $\tilde{\phi}_k$ are of the same dimension. The redundancy factor of the dictionary of A can be defined as $Q = K/N$. The forward transform of x with respect to A can be written as $A^H x$. However, according to the Fourier theory, this process is somewhat time consuming. By using the algorithm of FFT, a fast implementation can be shown as blow. The first step is to enlarge the length of x to the dimension of K by zero-padding,

$$x \xrightarrow{Zero\ Padding} \begin{bmatrix} x \\ 0_{1\times(N-k)} \end{bmatrix}. \tag{10}$$

The second step is to perform FFT on the augmented series. However, it can be seen that this forward transform is generally a dense representation. Therefore, a sparse coding algorithm is required.

Sparse Fourier Spectrum Analysis via SALSA

In optimization theory, a constraint optimization problem can be summarized as

$$\arg\min_z E(z) \text{ such that } Cz - b = 0, \tag{11}$$

where $E(\cdot)$ is the cost function. The associated augmented Lagrangian is defined as

$$L_A(z, \lambda, \mu) = E(z) + \lambda^T(cz - d) + \mu||cz - b||_2^2. \quad (12)$$

where the vector-valued variable λ is the Lagrange multiplier. Based on these methodologies, we can find numerical solutions for sparse representation problems.

Sparse representation (SR) is a novel idea of signal expansion by using redundant dictionaries, which can be expressed as

$$y = Ax. \quad (13)$$

In the above equality, the matrix A represents a dictionary of the dimension $N \times K$, in which $K \gg N$. The variable x is the linear combination coefficient vector of the dimension $K \times 1$. The theory of SR requires that most of the entries in x are zero. The solution to the above optimization problem can by numerically implemented via either matching pursuit (MP) or basis pursuit (BP). In this article, we present the idea of sparse Fourier spectrum based on BP algorithms. In engineering applications, in order to handle the noises in the measurement, an improved problem of P_1^ε can be formulated as

$$\min ||x||_1 \text{such that } ||Ax - y||_2 \leqslant \varepsilon, \quad (14)$$

where ε stands for the admissible error. That is to say, the ℓ_1 norm is utilized as the measure of the sparsity. In the literature, various techniques have been developed to solve the above P_1^ε problem. In this paper, the strategy of split augmented Lagrangian shrinkage algorithm (SALSA) is employed. The strategy of SALSA (20) is celebrated due to its flexibility and fast convergence. By introducing the ideals of variable splitting and augmented Lagrangian (21) into this algorithm, it can address the constraint optimization problem with robust performance. As such, let the observed signal be y and the dictionary matrix be A, the P_1^ε problem to obtain the optimized solution $\hat{x}$ can be written as

$$\hat{x} = \arg\min_x \frac{1}{2}||y - Ax||_2^2 + ||\lambda \odot x||_1, \quad (15)$$

where the vector-valued variable λ is the Lagrange multiplier and the operator $\odot$ means element-wise product of two vectors of equal size. The i-th element of $\lambda \odot x$ is defined as

$$[\lambda \odot x]_i = \lambda_i x_i. \quad (16)$$

Applying the strategy of variable splitting, we can have the following problem.

$$\hat{x} = \arg\min_{x,u} \frac{1}{2}||y - Ax||_2^2 + ||\lambda \odot u||_1 \text{ such that } u - x = 0. \quad (17)$$

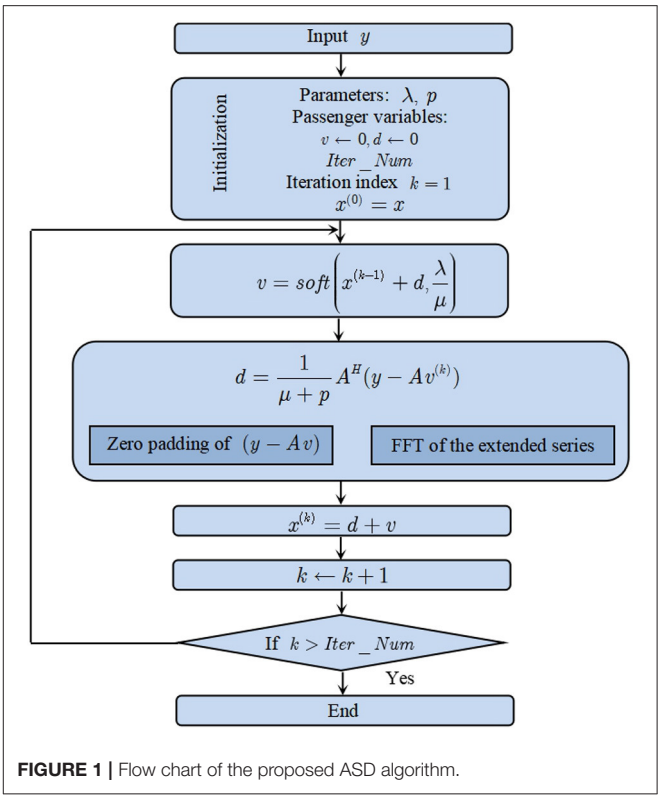


FIGURE 1 | Flow chart of the proposed ASD algorithm.

According to the augmented Lagrangian theory, the problem in Equation (17) can be prepared in a matrix form, which is shown as below.

$$z_1 = x, z_2 = u, z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, C = \begin{bmatrix} I & -I \end{bmatrix}, b = 0. \quad (18)$$

and

$$E(z) = \frac{1}{2}||y - Az_1||_2^2 + ||\lambda \odot z_2||_1 \quad (19)$$

The Proposed Adaptive Sparse Detector (ASD) for PLI

It can be inferred form the definition of the redundant dictionary A that it is a Parserval tight frame because $A^H A = pI$, where p is a constant. According to the theory of SALSA, the following algorithm based on iterations can be employed. The series y is the measured EGG signal, which is used as the input of the algorithm. The series x is the linear combination coefficient series. The variables λ and μ are necessary parameters required by the algorithm. The procedure of the algorithm is summarized as below.

Step 1. Initialize the parameters and passenger variables: $k = 1$ $\mu > 0, d$

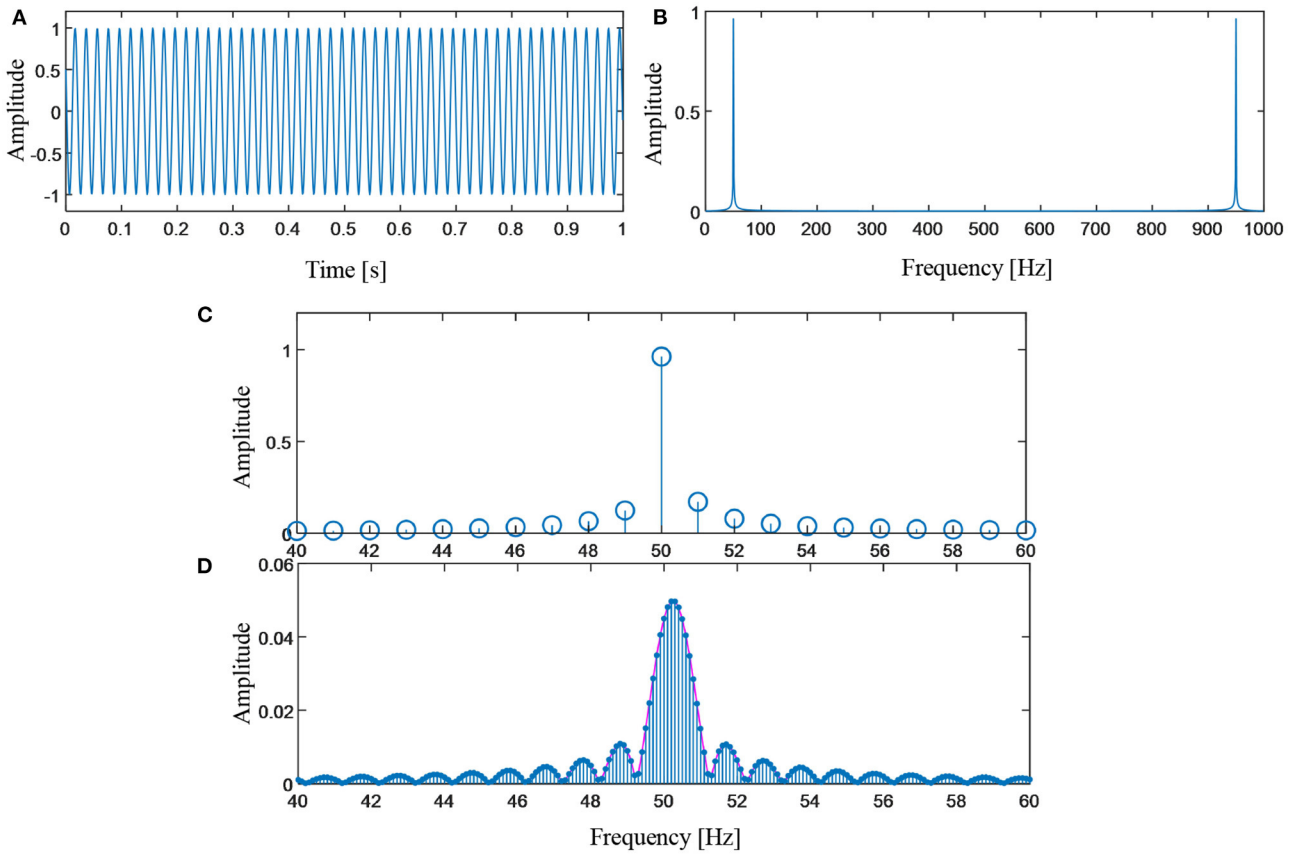


FIGURE 2 | (A) Time domain waveform of the simulated signal; **(B)** FFT spectrum of the simulated signal; **(C)** zoom-in plot of the FFT spectrum; and **(D)** linear combination coefficients of the redundant Fourier dictionary.

Step 2. Repeat the following routine

$$v = soft\left(x + d, \frac{\lambda}{\mu}\right) \tag{20}$$

$$d = \frac{1}{\mu + p} A^H (y - Av) \tag{21}$$

$$x = d + v \tag{22}$$

Step 3. $k = k + 1$. If $k \leqslant Inter\_num$, repeat Step 2. Otherwise, the iteration ends.

In the above procedure, the function $soft(\cdot)$ indicates the soft threshoding function defined by

$$soft(x, T) = \max\left(T./x, 0\right), \tag{23}$$

in which the symbol $'-/'$ means division by element and T is the threshoding value. For the ease of argument, we require that K, the number of atoms in the redundant dictionary A, to be multiples of N. The relationship of K and N is expressed as $K = MN$, in which M is a positive integer greater than one. The flowchart of the algorithm is depicted in **Figure 1**.

RESULTS

Numerical Simulation

In this subsection, numerical simulations are utilized to validate the performance of the proposed ASD. Following the definition of PLI in Equation (1), a digital signal is set as

$$p(t) = \cos\left(2\pi \cdot (50 + \Delta f) \cdot t + \frac{\pi}{3}\right), \tag{24}$$

in which the frequency shift Δf=0.15 Hz. The number of samplings and the sampling frequency are set as 1,000 and 1,000 Hz respectively. The time domain waveform of the $p(t)$ is shown in **Figure 2A**. In the spectral analysis, the frequency resolution is calculated as.

$$\Delta f = \frac{fs}{N} = 1Hz/Spectral\ Line \tag{25}$$

Because the sampling of the sinusoidal component does not meet the full period sampling condition, the energy leakage phenomenon occurs in the FFT spectrum. **Figure 2B** shows the energy of this sinusoidal component spreads across the entire frequency domain with slow decaying rate.

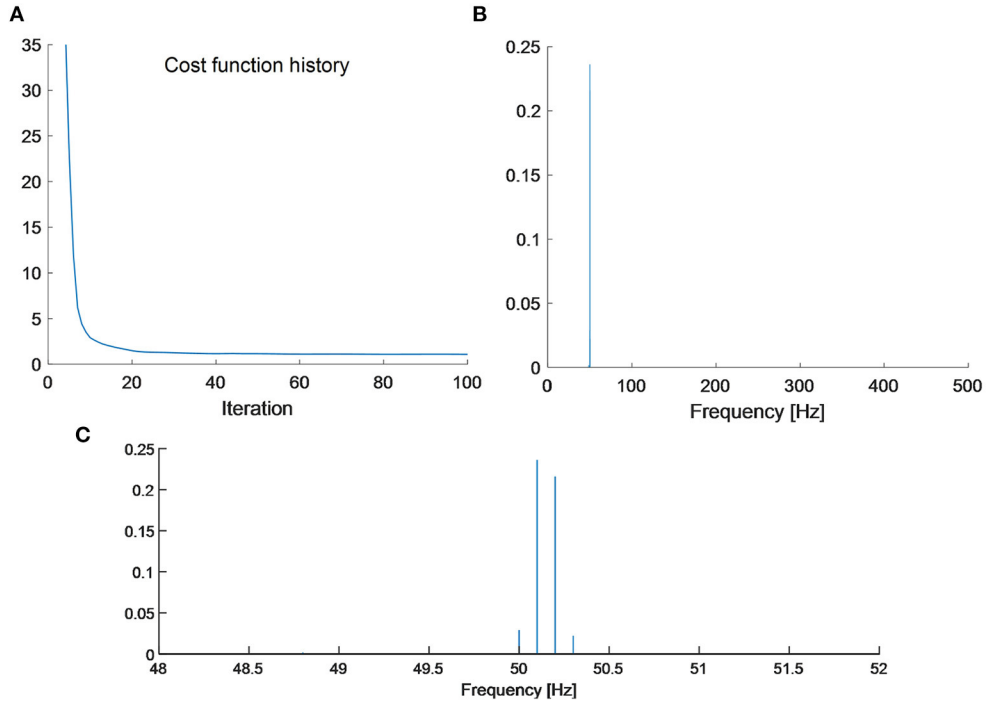


FIGURE 3 | (A) The cost function history; **(B)** sparse Fourier spectrum by the proposed ASD methodology; and **(C)** zoom-in plot of sparse Fourier spectrum.

A redundant Fourier dictionary with the definition in Equation (8) is used to represent the signal. The redundancy factor of the dictionary is set as 10. That is to say, ten thousand sinusoidal atoms are employed. A direct transform generates the redundant Fourier spectrum (**Figure 2D**). Compared with the waveform in **Figure 2C**, the space between adjacent spectral lines is reduced to one-tenth of that in the FFT spectrum. However, the energy leakage problem remains unchanged, which can be observed from the envelope of the redundant spectrum.

To sparsely represent the signal, the ASD algorithm is performed by using the redundant Fourier basis. The iteration number is set as 100. Let $x^{(k)}$ be the linear combination coefficients in the k-th iteration, an indicator of loss function can be defined as

$$\text{Loss Fun}(k) = \left\| \lambda \odot x^{(k)} \right\|_1. \tag{26}$$

The curve of the cost function is shown in **Figure 3A**. It can be seen that the cost function converges quickly. About tem times of iterations are enough to guarantee an approximate sparse representation with respect to the dictionary A. The final linear combination coefficient vector $\hat{x}$ of the ASD methodology is shown in **Figure 3B**. It can be observed that only three spectral lines are large in amplitude. The associated frequencies of them are 50, 50.1, 50.2, and 50.3 Hz. The amplitudes of the other spectral lines are very small in value, so they are negligible.

Processing Results of Actual EEG Measurements

In this subsection, the analyzed datasets are provided by the Department of Epileptology at University of Bonn (22). The EEGs were recorded at the sampling frequency of 173.61 Hz. According to the Shannon sampling theorem, the spectral bandwidth of the EEG recordings is 0.5–85 Hz. Digital filters, with the passing band of 0.53–40 Hz, were utilized as pre-processings of the EEGs. **Figure 4A** illustrates a EEG segment collected from healthy volunteers in an awake state with eyes open. The FFT spectrum of the signal segment is shown in **Figure 4B**. Due to the band pass filtering step, the spectrum components in the frequency range 40–85 Hz are relatively small. However, by visual inspection, an energy concentration area can be still found near the power line frequency (**Figure 4C**).

To remove the PLI, the ASD methodology is performed on the EEG segment. In order to guarantee the convergence of the algorithm, two thousand iterations are employed. The convergence of the loss function can be found in **Figure 4G**. The sparse Fourier spectrum is depicted in **Figure 4D**. The contents of the signal in the frequency domain are quite complicated. Many non-stationary components can be detected. Due to the sparse representation algorithm, the amplitudes of many spectral lines are almost zero. As such, sinusoidal waves, whose sparse spectrum consists of a very limited number of spectral lines, can be effectively isolated from other non-stationary components.

Two spectral lines, whose frequencies are 49.95 and 49.96 Hz, can be used for retrieve the PLI component. Besides, another

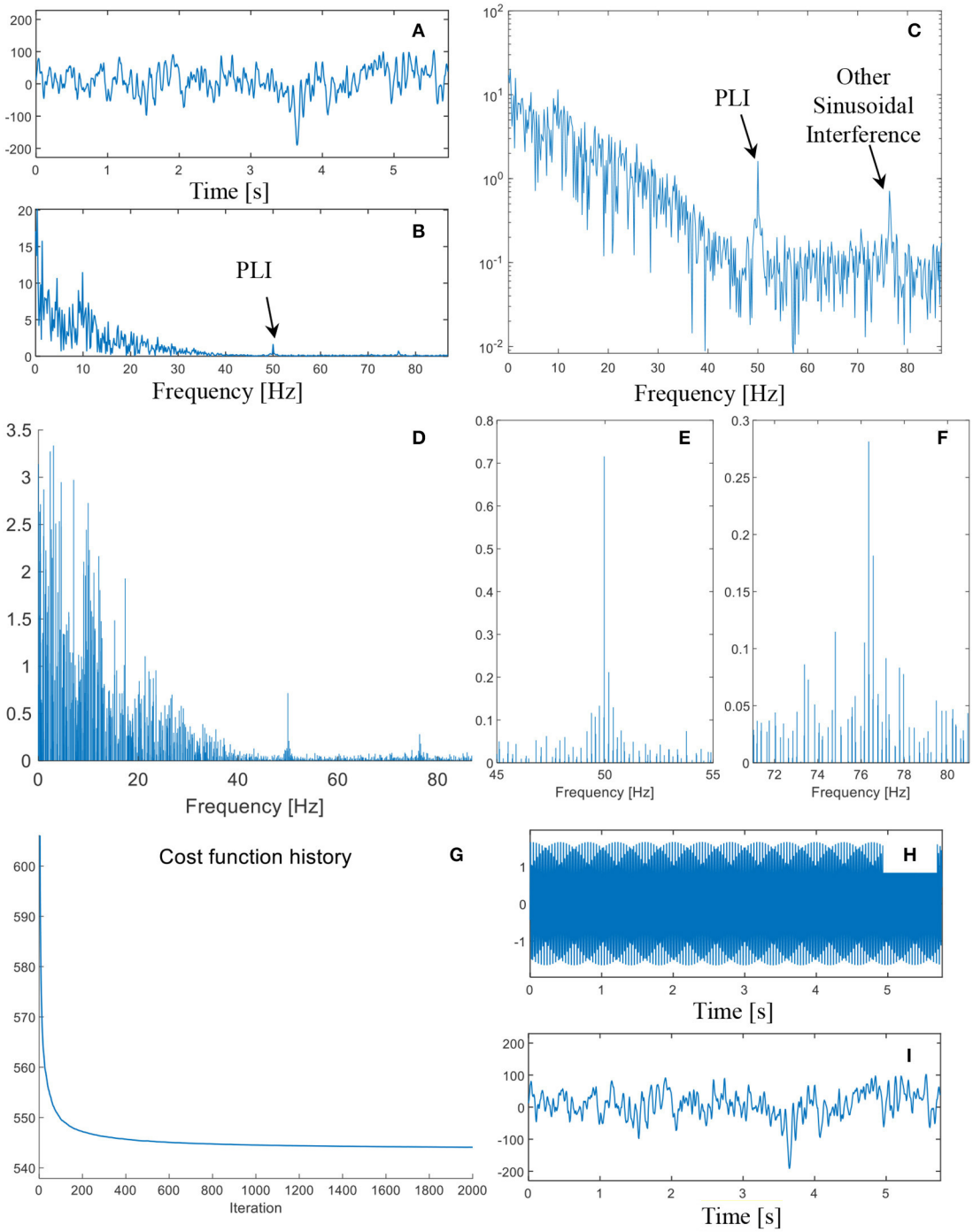


FIGURE 4 | (A) The time domain waveform of the EEG measurement; **(B)** FFT spectrum in linear scale; **(C)** FFT spectrum in logarithmic scale; **(D)** sparse FFT spectrum by the proposed method; **(E)** zoom-in plot of the sparse spectrum in the neighbor of 50 Hz; **(F)** zoom-in plot of the sparse spectrum in the neighbor of 76 Hz; **(G)** the cost function history of the iterated algorithm; **(H)** the synthesized compensation signal; and **(I)** the denoised signal.

strong sinusoidal wave, whose frequency is ∼76 Hz, is also found in the spectrum. zoom-in plots of two sinusoidal waves are illustrated in **Figures 4E,F**. The spectral lines of 49.95 and 49.96 Hz are picked out to reconstruct the PLI component, whose time domain waveform is shown in **Figure 4H**. By checking the shape, it is confirmed that the sinusoidal wave is perfectly

reconstructed. By subtracting the retrieved PLI component from the EEG segment, the denoised signal is shown in **Figure 4I**. The PLI component is relatively weak in energy, and therefore the denoised signal is very similar to the original EEG segment.

DISCUSSION

According to the above arguments, it can be concluded that the proposed ASD, which is based on the SALSA, can be utilized as an effective algorithm to retrieve sinusoidal waves.

In spectral analysis of sinusoidal waves of finite digital samples, the ASD can achieve sparse representation. For a sinusoidal wave, the phenomenon of energy leakage occurs if its sampling does not meet the full period sampling condition. In such circumstances, the spectral counterpart of a sinusoidal wave is composed of a main lobe and a few side lobes. The side lobes caused by energy leakage spread across the entire frequency domain. For signal analysis with a segmentation window of the rectangular shape, the side lobes decay very slowly. Reconstruction error by using the main lobe will cause a big error. However, by using a redundant Fourier dictionary containing evenly spaced sinusoidal atoms, it is possible to alleviate this problem. In the numerical simulation, the simple harmonic wave can be sparsely represented by four spectral lines. The amplitudes of two adjacent side lobes are 1.63×10^{-4} and 1.39×10^{-4}respectively. The reconstructed signal can be obtained using the four spectral lines shown in **Figure 3C**. The relative error between the reconstruction error and the original signal is calculated to be 0.2%. While in the EEG recorded from actual measurements, only two spectral lines are sufficient to allow a reconstruction of PLI with high accuracy. The amplitudes of two adjacent side lobes are 3.88×10^{-7} and 3.78×10^{-7} respectively. The side lobes also exist in the ASD, but they very small in energy. Therefore, the side lobes can be ignored in the reconstruction process.

As shown in the flow chart (**Figure 1**), there are many parameters in the SALSA algorithm. The redundancy factor (Q) of the employed dictionary and the iteration number can directly affect the sparse Fourier spectrum. In the numerical simulation, there are 4 strong spectral lines when $Q = 10$ and $Iter\_Num = 100$. Keeping the parameter Q unchanged, there can be only 2 strong spectral lines when the iteration number is increased to 1,000. On the other hand, a large value of Q is beneficial in prompting the sparsity of the resultant spectrum. However, larger values of Q and $Iter\_Num$ requires are more time consuming. To suppress the PLI component as well as other sinusoidal waves, we can set the amplitudes of spectral lines as zero relevant to them and keep other spectral contents unchanged. It is unnecessary to identify their harmonic parameters of amplitude, frequency and phase. However, these parameters can be calculated by using FFT on their reconstruction signals.

Regarding how to determine the number of atoms in the Fourier dictionary, as the number of atoms increases, it will be beneficial to the compression of the main lobe width of the PLI, but it will also affect the efficiency of the whole algorithm, generally ten times the number of atoms of the original is enough.

In addition to the method proposed in this paper, the using of adaptive notch filters for PLI removal has been investigated by the authors (23). The core of the study is ratio-based spectral correction which can extract the spectral information of the PLI components. The difference between the study and ours is that the preceding method does not change the spectral resolution. The information of the PLI components is based on the ratio computation, while the energy leakage of the spectrum faces no mitigation. Our method reduces the phenomenon of overlapping by increasing the spectral resolution. And due to the narrow band property of PLI in the spare representation, they can be more easily isolated from other contents, making its signal information extraction efficient and accurate.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://epileptologie-bonn.de/cms/.

AUTHOR CONTRIBUTIONS

B-qC and W-fS conceived and designed the classification method and reviewed and edited the manuscript. C-qW and B-xZ performed the experiment. C-qW and B-xZ preprocess and analyzed the data and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported financially by National Natural Science Foundation of China (No. 51605403), Natural Science Foundation of Guangdong Province, China (No. 2015A030310010), Natural Science Foundation of Fujian Province, China (No. 2016J01012), the Aeronautical Science Foundation of China (No. 20183368004), and the Fundamental Research Funds for the Central Universities under Grant (No. 20720190009).

REFERENCES

1. Birbaumer N, Weber C, Neuper C, Buch E, Haagen K, Cohen K. Brain–computer interface research: coming of age. *Clin Neurophysiol.* (2006) 117:479–83. doi: 10.1016/j.clinph.2005.11.002
2. Lehmann D. EEG assessment of brain activity: spatial aspects, segmentation and imaging. *Int J Psychophysiol.* (1984) 1:267–76. doi: 10.1016/0167-8760(84)90046-1
3. Leske S, Dalal SS. Reducing power line noise in EEG and MEG data via spectrum interpolation. *Neurolmage.* (2019) 189:763–76. doi: 10.1016/j.neuroimage.2019.01.026
4. Gramann K, Ferris DP, Gwin J, Makeig S. Imaging natural cognition in action. *Int J Psychophysiol.* (2014) 91:22–9. doi: 10.1016/j.ijpsycho.2013.09.003
5. Makeig S, Gramann K, Jung TP, Sejnowski TJ, Poizner H. Linking brain, mind and behavior. *Int J Psychophysiol.* (2009) 73:95–100. doi: 10.1016/j.ijpsycho.2008.11.008

6. Benatti S, Casamassima F, Milosevic B, Farella E, Schönle P, Fateh S, et al. A versatile embedded platform for EMG acquisition and gesture recognition. *IEEE Trans Biomed Circuits Syst*. (2015) 9:620–30. doi: 10.1109/TBCAS.2015.2476555
7. Qin L, He B. A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications. *J Neural Eng*. (2005) 2:65–72. doi: 10.1088/1741-2560/2/4/001
8. Siuly S, Li Y. Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain computer interface. *IEEE Transact Neural Syst Rehabil Eng*. (2012) 2.0:526–38. doi: 10.1109/TNSRE.2012.2184838
9. Tomasini M, Benatti S, Milosevic B, Farella E, Benini L. Power line interference removal for high-quality continuous biosignal monitoring with low-power wearable devices. *IEEE Sens J*. (2016) 16:3887–95. doi: 10.1109/JSEN.2016.2536363
10. Verma AR, Singh Y. Adaptive tunable notch filter for ECG signal enhancement. *Proc Comp Sci*. (2015) 57:332–7. doi: 10.1016/j.procs.2015.07.347
11. Chen BQ, Lan QX, Li Y, Zhuang SQ, Cao XC. Enhancement of fault feature extraction from displacement signals by suppressing severe end distortions via sinusoidal wave reduction. *Energies*. (2019) 12:3536. doi: 10.3390/en12183536
12. Peng ZK, Tse PW, Chu FL. A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing. *Mech Syst Signal Process*. (2005) 19:974–88. doi: 10.1016/j.ymssp.2004.01.006
13. He WP, Zi YY, Chen BQ, Wu F, He ZJ. Automatic fault feature extraction of mechanical anomaly on induction motor bearing using ensemble super-wavelet transform. *Mech Syst Signal Proces*. (2015) 54–55:457–80. doi: 10.1016/j.ymssp.2014.09.007
14. Poornachandra S. Wavelet-based denoising using subband dependent threshold for ECG signals. *Digital Signal Process*. (2008) 18:49–55. doi: 10.1016/j.dsp.2007.09.006
15. Wang SH, Govindaraj VV, Górriz JM, Zhang X, Zhang YD. Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. *Inform Fusion*. (2021) 67:208–29. doi: 10.1016/j.inffus.2020.10.004
16. Wang SH, Nayak DR, Guttery DS, Zhang X, Zhang YD. COVID-19 classification by CCSHNet with deep fusion using transfer learning and discriminant correlation analysis. *Inform Fusion*. (2020) 68:131–48. doi: 10.1016/j.inffus.2020.11.005
17. Varadarajan KK, Suhasini PR, Manikantan K, Ramachandran S. Face recognition using block based feature extraction with CZT and Goertzel-algorithm as a preprocessing technique. *Proc Comp Sci*. (2014) 46:1458–67. doi: 10.1016/j.procs.2015.02.065
18. Li F, Gao Y, Cao Y, Iravani R. Improved teager energy operator and improved chirp-Z transform for parameter estimation of voltage flicker. *IEEE Trans Power Del*. (2015) 31:245–53. doi: 10.1109/TPWRD.2015.2448943
19. Zhang B, Xiao TY, Zhong JZ. A simple determination approach for zero-padding of FFT method in focal spot calculation. *Opt Commun*. (2019) 451:260–4. doi: 10.1016/j.optcom.2019.06.065
20. Afonso MV, Bioucas-Dias JM, Figueiredo MAT. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans Image Process*. (2010) 19:2345–56. doi: 10.1109/TIP.2010.2047910
21. Afonso MV, Bioucas-Dias JM, Figueiredo MAT. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans Image Process*. (2011) 20:681–95. doi: 10.1109/TIP.2010.2076294
22. Andrzejak RG, Lehnertz K, Rieke C, Mormann F, David P, Elger CE. Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Phys Rev Ed*. (2001) 64:061907. doi: 10.1103/PhysRevE.64.061907
23. Chen BQ, Li Y, Cao XC, Sun WF, He WP. Removal of power line interference from ECG signals using adaptive notch filters of sharp resolution. *IEEE Access*. (2019) 7:150667–76. doi: 10.1109/ACCESS.2019.2944027

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 Public Health

OPINION
published: 14 May 2021
doi: 10.3389/fpubh.2021.682377

Ultra-Resolution Spectral Correction Based on Adaptive Linear Neuron for Biomedical Signal Processing

Binqiang Chen[1], *Baixun Zheng*[1] *and Weifang Sun*[2*]

[1] School of Aerospace Engineering, Xiamen University, Xiamen, China, [2] College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou, China

Keywords: fast fourier transform, biomedical signal, adaptive linear neuron, sinusoidal wave, spectrum correction

INTRODUCTION

Sinusoidal waves (SWs) are often presented in engineering dynamic measurements. They also appear in biomedical signals (1–3), such as electrocardiogram (ECG), electroencephalogram (EEG) and electromyography (EMG). The power line interference (4–6) is among the major types of interferences in these biomedical signals and it should be removed using hardware or software based techniques (7–9). A simple harmonic wave can be characterized using a group of harmonic parameters. By identifying these parameters, a compensation signal can be constructed and subtracted from the original measurement (10). In the literature, current techniques to estimate harmonic parameters mainly rely on discrete Fourier analysis. According to the theory of discrete Fourier analysis, the spectral resolution of a spectrum is the reciprocal of the signal sample length. Therefore, greater values of the sampling length are beneficial in improving the performance of harmonic parameter identification.

During the past decades, various spectrum correction methods have been developed to estimate harmonic information with high precision (11–15). With the help of spectrum correction methods, the harmonic information of a signal SW can be retrieved effectively. In the practices of biomedical signal processing, multiple SWs may be presented in the signal simultaneously. In ideal cases, the frequency distances of adjacent SWs should be sufficiently large such that they can be separated individually. Generally, at least a distance of five spectral resolutions is required. Otherwise, as the distance between their frequencies gets closer, the PFE even deteriorates and the overlapping of their energies cannot be neglected. To separate MSWs with relative small differences in frequency, there are still no effective method in the literature.

In order to address this problem, an effective solution is to predict additional data samples beyond the recorded datasets. As the number of samples increases, including the original samples and the predicted samples, the spectral resolution in the FFT spectrum can be improved such that the coupled MSWs can be separated. This solution has been adopted in the theory of modern spectrum analysis (16). However, current techniques can only ensure effective time series prediction in a short time. The prediction errors increases dramatically in a long term situation. As such, only a limited number of prediction samples can be obtained by post-processing of the original dataset. Hence, a limited improvement of spectral resolution can be derived.

In this paper, a novel separation method, based on the adaptive linear neuron, is proposed for MSWs whose frequencies are closed spaced in the spectral domain. The autoregressive model is employed to describe intrinsic structure of the MSW data. The obtained samples are utilized to train the AR model. In order to reduce the training errors, a norm function is formulated and the pseudo inverse is used to derive the closed form solution. Additional samples can be predicted via the established AR model from the recorded samples. Simulation results show the auto-regression algorithm can guarantee precise predictions in a long time interval beyond the original sampling

OPEN ACCESS

Edited by:
Yonghong Peng,
Manchester Metropolitan University,
United Kingdom

Reviewed by:
Shuai Wang,
Hefei University of Technology, China
Zhe Yang,
Dongguan University of Technology, China

***Correspondence:**
Weifang Sun
swf@wzu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *18 March 2021*
Accepted: *19 April 2021*
Published: *14 May 2021*

Citation:
Chen B, Zheng B and Sun W (2021) Ultra-Resolution Spectral Correction Based on Adaptive Linear Neuron for Biomedical Signal Processing. Front. Public Health 9:682377. doi: 10.3389/fpubh.2021.682377

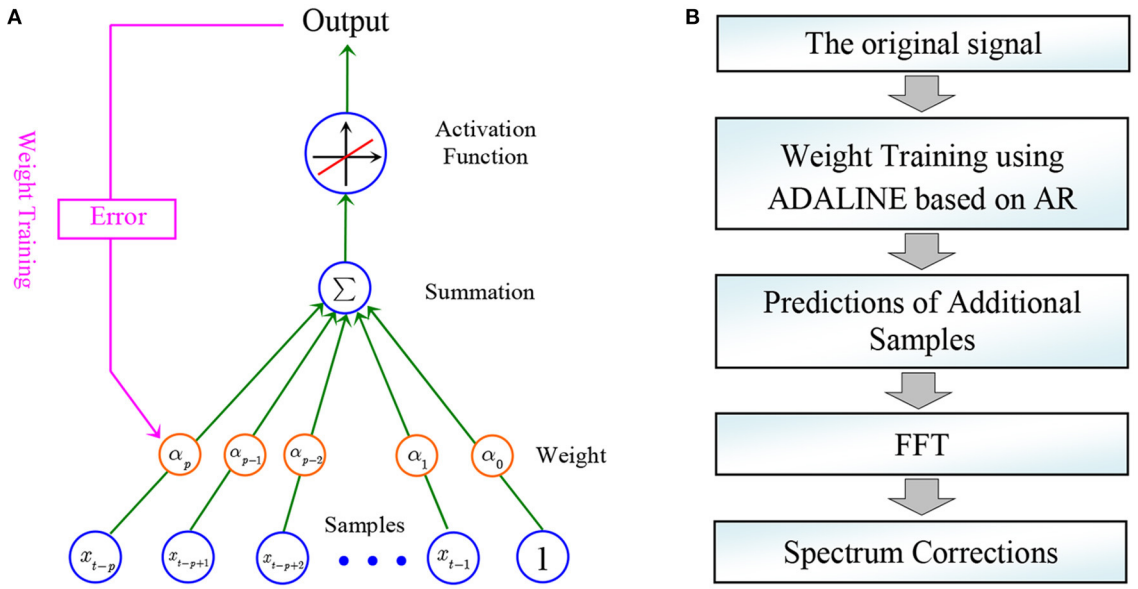


FIGURE 1 | (A) The structure of a typical adaptive linear neuron; **(B)** The procedure of the proposed method.

length. An extremely high spectral resolution can be achieved from the extended data length, therefore the coupled MSWs can be distinguished in the spectral domain. The proposed method can be regarded as a theoretical model to separate the coupled MSWs in the frequency domain. Numerical simulations are presented to show the effectiveness of the proposed method.

PERSPECTIVE METHOD FOR SEPARATION OF CLOSED SPACED MSWS

Picket Fence Effect and Aliasing of MSW

In mathematics, FFT generates a series of spectral lines evenly spaced at a specific frequency resolution, which is the reciprocal of the entire sampling interval. For a sinusoidal wave with a specific frequency, If the signal sampling period does not contain an integer number of harmonic periods (full period sampling condition), the energy leakage problem of the harmonic components occurs. This phenomenon is also known as the picket fence effect (PFE). Side effects of PFE are also reported in wavelet decompositions (17, 18).

Adaptive Linear Neuron

The adaptive linear neuron is a naive and useful tool in artificial neural networks (19, 20). As a single layer artificial neural network, the structure of a typical ADALINE is illustrated in **Figure 1A**. A classical proportional function $y = x$ is employed as the activation function. On the other hand, the indicator of mean squared error can be employed as the loss function. As a supervised learning method, the strategy of back propagation (BP) can be used to train the weights. The research of artificial neural network is a hot topic in the scientific and engineering community. In the past decade, by introducing new activation functions and loss functions, deep neural network based on stacked neurons can be constructed.

A Perspective Method for Multi-SW Separation Problem

In order to establish a precise model for a dynamic process of multi-SW, the celebrated auto regressive model is introduced to implement weight training in ADALINE. An autoregressive model of order p can be expressed as

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_p x_{t-p} + \varepsilon_t. \quad (1)$$

The error term ε_t can be interpreted as stochastic noises. For an MSW series with a mean value of zero, this parameter can be set to zero. The values of rest weight parameters α_i, $i = 0, 1, \ldots, p$, can be obtained using a least mean square (LMS) approach. Supposing that we have an original series of length L. All the equations associated with the AR(p) model in Equation (1) can be expressed in a matrix form as

$$\begin{pmatrix} x_1 & x_2 & \ldots & x_p \\ x_2 & x_3 & \ldots & x_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{L-p} & x_{L-p+1} & \ldots & x_{L-1} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_L \end{pmatrix}. \quad (2)$$

In the above equation, the coefficient matrix A is of the size $(L-p)\times p$. It is a matrix of full column rank and its pseudo inverse

can be computed as $A^{\dagger} = (A^T A)^{-1} A^T$. The unknown column vector containing the AR parameters can be computed using

$$\alpha = A^{\dagger}\left(x_{p+1}\ x_{p+2}\ \ldots\ x_L\right)^T. \quad (3)$$

The above least mean square method is among the various techniques to estimate the AR parameters. Because the order of p is relatively small, this method can be extremely efficient.

After establishing the AR model, data extrapolation can be made. Beside the available data, additional samples can be predicted using Equation (1). The flow chart of the proposed method is displayed in **Figure 1B**. The estimation error significantly affects the prediction accuracy, which is shown in the discussion part.

Since the length of the time series being investigated can be substantially enlarged using the data extrapolation, the spectral resolution in the FFT spectrum also improves. As a result, the MSWs closely spaced in the original spectrum can be separated in the new spectrum.

Discussions

In the proposed method, the training process of ADALINE is completed by AR. However, in the literature, the back propagation (BP) is more popular. By using BP, the cost function can also be blow in value. However, the prediction errors grow significantly as the prediction length increases. Therefore, AR can be regarded as an effective long term prediction tool in multi-SW processes.

We can lengthen the entire sampling interval based on predictions of additional samples, which significantly improves the actual frequency resolution in the FFT spectrum. That is to say, we can separate their counterparts in the spectral domain. However, for SW components which do not satisfy the full period sampling condition, spectrum correction methods should be combined to further investigate their harmonic information.

AUTHOR CONTRIBUTIONS

BC and WS conceived and designed the classification method, reviewed, and edited the manuscript. BZ performed the experiment, preprocess, analyzed the data, and wrote the manuscript. All authors read and approved the manuscript.

FUNDING

This research was supported financially by Fundamental Research Funds for the Central Universities under Grant (No. 20720190009).

REFERENCES

1. Wu HT. Current state of nonlinear-type time–frequency analysis and applications to high-frequency biomedical signals. *Curr Opin Syst Biol.* (2020) 23:8–21. doi: 10.1016/j.coisb.2020.07.013
2. Huang JS, Chen BQ, Yao B, He WP. ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network. *IEEE Access.* (2019) 7:92871–80. doi: 10.1109/ACCESS.2019.2928017
3. Koizumi H, Whitten WB, Reilly PTA, Koizumi E. Derivation of mathematical expressions to define resonant ejection from square and sinusoidal wave ion traps. *Int J Mass Spectrometry.* (2009) 286:64–9. doi: 10.1016/j.ijms.2009.06.011
4. Singhal A, Singh P, Fatimah B, Pachori RB. An efficient removal of power-line interference and baseline wander from ECG signals by employing Fourier decomposition technique. *Biomed Signal Proc Control.* (2020) 57:101741. doi: 10.1016/j.bspc.2019.101741
5. Qiu Y, Huang KJ, Xiao F, Shen HB. A segment-wise reconstruction method based on bidirectional long short term memory for Power Line Interference suppression. *Biocybern Biomed Eng.* (2018) 38:217–27. doi: 10.1016/j.bbe.2018.01.003
6. Taralunga DD, Gussi I, Strungaru R. Fetal ECG enhancement: adaptive power line interference cancellation based on Hilbert Huang Transform. *Biomed Signal Proc Control.* (2015) 19:77–84. doi: 10.1016/j.bspc.2015.03.009
7. Razzaq N, Sheikh SAA, Salman M, Zaidi T. An intelligent adaptive filter for elimination of power line interference from high resolution electrocardiogram. *IEEE Access.* (2016) 4:1676–88. doi: 10.1109/access.2016.2548362
8. Widrow B, Glover JR, McCool JM, Kaunitz J, Williams CS, Hearn RH, et al. Adaptive noise cancelling: principles and applications. *Proc IEEE.* (1975) 63:1692–716. doi: 10.1109/PROC.1975.10036
9. Gokcesu K, Ergeneci M, Ertan E, Gokcesu H. An adaptive algorithm for online interference cancellation in EMG sensors. *IEEE Sensors J.* (2019) 19:214–23. doi: 10.1109/JSEN.2018.2874724
10. Chen BQ, Lan QX, Li Y, Zhuang SQ, Cao XC. Enhancement of fault feature extraction from displacement signals by suppressing severe end distortions via sinusoidal wave reduction. *Energies.* (2019) 12:3536. doi: 10.3390/en12183536
11. Ni GM, Liu LB, Yu XJ, Ge X, Chen S, Liu X, et al. Contrast enhancement of spectral domain optical coherence tomography using spectrum correction. *Comp Biol Med.* (2017) 89:505–11. doi: 10.1016/j.compbiomed.2017.09.004
12. Lin HB, Ding K. A new method for measuring engine rotational speed based on the vibration and discrete spectrum correction technique. *Measurement.* (2013) 46:2056–64. doi: 10.1016/j.measurement.2013.03.017
13. Salor O. Spectral correction-based method for interharmonics analysis of power signals with fundamental frequency deviation. *Electric Power Syst Res.* (2009) 79:1025–31. doi: 10.1016/j.epsr.2008.12.016
14. Zhang WK, Wen DS, Song ZX. Spectrum reconstruction in interference spectrometer based on sparse Fourier transform. *Optik.* (2018) 154:157–64. doi: 10.1016/j.ijleo.2017.09.004
15. Chen BQ, Li Y, Cao XC, Sun W, Zhuang S. Recovery of undersampled force measurement from high-speed milling process using approximate sparsity in frequency domain. *Measurement.* (2021) 175:109143. doi: 10.1016/j.measurement.2021.109143
16. Stoica P, Moses R. *Spectral Analysis of Signals*. New York, NY: Prentice Hall (2005)
17. He WP, Zi YY, Chen BQ, Wu F, He ZJ. Automatic fault feature extraction of mechanical anomaly on induction motor bearing using ensemble super-wavelet transform. *Mechanical Systems Signal Proc.* (2015) 54–55:457–80. doi: 10.1016/j.ymssp.2014.09.007
18. Huang JS, Chen BQ, Li Y, Sun WF. Fractal geometry of wavelet decomposition in mechanical signature analysis. *Measurement.* (2020) 173:108571. doi: 10.1016/j.measurement.2020.108571
19. Cao XC, Chen BQ, Yao B, He WP. Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification. *Comp Industry.* (2019) 106:71–84. doi: 10.1016/j.compind.2018.12.018

20. Cao XC, Chen BQ, Zeng NY. A deep domain adaption model with multi-task networks for planetary gearbox fault diagnosis. *Neurocomputing.* (2020) 409:173–90. doi: 10.1016/j.neucom.2020.05.064

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 Public Health

ORIGINAL RESEARCH
published: 19 May 2021
doi: 10.3389/fpubh.2021.649624

Economic Burden of Major Diseases in China in 2013

Xianyan Song[1], *Lan Lan*[2]*, *Ting Zhou*[3], *Jin Yin*[2] and *Qiong Meng*[4]*

[1] Institute of Hospital Management, West China Hospital, Sichuan University, Chengdu, China, [2] West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China, [3] Department of Epidemiology and Biostatistics, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China, [4] Department of Epidemiology and Health Statistics, School of Public Health, Kunming Medical University, Kunming, China

Studies on the economic burden of disease (EBD) can estimate the social benefits of preventing or curing disease. The majority of studies focus on the economic burden of a single or regional disease; however, holistic or national research is rare in China. Estimating the national EBD can provide evidence for policy makers. We used the top-down method to assess the economic burden of 30 types of diseases between urban and rural areas in China. The two-step model was used to evaluate the direct economic burden of disease (DEBD), while the human capital method was used to assess the indirect economic burden of disease (IEBD). The total economic burden of 30 types of diseases in China was between $13.39 and 803.00 billion in 2013. The average total economic burden of disease (TEBD) in cities was $81.39 billion, while diseases in villages accounted for $50.26 billion. The range of direct and indirect EBD was $5.77–494.52 billion, and the range in urban areas was $0.61–20.34 billion. The direct and indirect EBD in rural areas accounted for $5.88–277.76 billion and $0.59–11.39 billion, respectively. There was a large difference between the economic burden of different diseases. The economic burden of urban diseases was more significant than the burden for the rural. The top five most economically burdensome diseases were myocardial infarction coronary artery bypass, acute myocardial infarction, cerebral hemorrhage, acute upper gastrointestinal bleeding and acute appendicitis.

Keywords: direct economic burden, indirect economic burden, disease, China, decision making

OPEN ACCESS

Edited by:
Yonghong Peng,
Manchester Metropolitan University,
United Kingdom

Reviewed by:
Yue Huang,
Jinan University, China
Jiangbo Pu,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

***Correspondence:**
Qiong Meng
mengqiong2006@163.com
Lan Lan
lanl@scu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *05 January 2021*
Accepted: *26 April 2021*
Published: *19 May 2021*

Citation:
Song X, Lan L, Zhou T, Yin J and Meng Q (2021) Economic Burden of Major Diseases in China in 2013. Front. Public Health 9:649624. doi: 10.3389/fpubh.2021.649624

INTRODUCTION

The rapid development of the economy has improved living and nutritional standards. Simultaneously, it has had a large effect on changing the spectrum of diseases in China (1, 2). Chronic non-communicable diseases are currently the prevailing diseases, increasing number of people paying more attention to their health. Furthermore, the study of the economic burden of disease (EBD) can help estimate the social benefits, resulting in the prevention or cure of disease, and can provide information related to the economic evaluation of disease (3–6). More importantly, this type of research offers a scientific basis for limited medical insurance and a rational allocation of health resources.

China introduced a new medical system reform policy on March 17, 2009. The main purpose of the new policy was to solve the problems that made it difficult and expensive for patients to visit a doctor. The new policy has decreased the damage on patients' bodies and the loss to society caused by illnesses. However, if the main disease can be prevented by implementing certain intervening measures with limited health resources, injuries and losses for individuals, families and societies that are caused by the disease can be avoided. Due to China's unique economic development, there is great economic disparity between urban and rural areas in China, which has greatly affected the constituent of the disease and the treatment options that are available to residents living between urban and rural areas. The majority of researches on economic burden has mostly focused on a single disease or localized diseases (7–11), but an integrated study of multiple diseases is rare in China. Recent studies only estimated the economic burden of lung cancer (12), seasonal influenza (13), chronic diseases (14), children with asthma (15) and Alzheimer's disease (16) in China, and the economic burden of Alzheimer's disease in Zhejiang Province (17), hepatitis E-infected patients in Jiangsu Province (18) and rare diseases in Shanghai (19) etc. Therefore, we intend to estimate the overall EBD and then, to separately measure the EBD between urban and rural areas in China in order to provide evidence for relevant policy makers.

METHODS

We used a top-down approach to estimate EBD in China. The top-down approach (6) finds national or regional total health expenditures in advance, relying on the existing system for information on health expenditures. Then, the various expenses are classified, such as hospitalization costs, outpatient and emergency costs, drug costs, etc., based on a certain percentage corresponding to the allocation of items used to target disease. Lastly, the average cost of individual or various diseases is calculated.

Data Collection

The Analysis Report of Notional Health Services Survey in China 2013 (ARNHSS) provided consultation rates data for 2 weeks, categorized by the diseases and hospitalization rates and constituent ratio of outpatient and inpatient institutions in 2013. The China Health and Family Planning Statistical Yearbook 2014 (CHFPSY) provided the data for the average outpatient costs, hospitalization costs and the length of stay by diseases at different hospital levels and the total national health expenditure in 2013. The China Statistical Yearbook 2014 (CSY) provided the urban and rural population numbers and the Gross Domestic Product (GDP) data in 2013.

Every 5 years, the Chinese government conducts the national health services survey, which is a special sample survey, in order to understand the urban and rural residents' health status, health care utilization, health care costs and burden on the country. In 2013, the fifth national health services survey adopted multi-stage stratified cluster random sampling. The sample covered 31 provinces, 156 counties, 780 towns and 1,560 villages. It collected data from a total of 93,600 families, which is ∼30 million people and is a good representative sample of China. The data from CHFPSY and CSY were reported by the administrative agency that is highly credible and known for high quality work. Furthermore, in order to obtain a consistent statistical caliber, each analytical index in CHFPSY was as closely matched as possible with ARNHSS in China.

Research Disease

We selected 30 types of diseases from CHFPSY for this study to estimate their direct and indirect economic burden because the classification of diseases did not precisely match between CHFPSY and ARNHSS in China. Additionally, the CHFPSY classification of diseases was coded according to ICD-10. Therefore, it can be deduced that data integration was in process. Firstly, we listed the 30 types of diseases from CHFPSY 2014, and then we matched the name of the disease with ARNHSS in China 2013. If we could not find the exact same name of the disease, we classified the diseases into related systems according to the systems of diseases or main symptoms. If they still did not match using this method, the disease was classified as missing.

Direct Economic Burden of Disease (DEBD)

DEBD refers to the total direct costs for the prevention and treatment of diseases, including various costs incurred by the individual, family and society for preventing, diagnosing, treating and rehabilitating the diseases and injuries (20). On the one hand, it includes the costs of health services provided by the health institutions, such as the prevention of capital costs, emergency expenses, outpatient diagnosis and treatment costs, hospital expenses, medical expenses, health technology labor, family bed treatment and care costs, etc. On the other hand, other charges paid by the patients while receiving health services are included, such as nutrition fees, transportation fees, travel expenses, non-prescription costs for purchasing rehabilitation equipment, etc. We estimated the direct economic burden of 30 types of diseases, including outpatient medical expenses and inpatient medical expenses only because we don't have non-medical expenses. The formula is as follows:

$$DEBDi = N \times [\text{2-week consultation rate} \times 26 \times (\Sigma \text{ outpatient facilities constituent ratio} \times \text{average outpatient medical cost in outpatient institutions at different levels}) + \text{hospitalization rate} \times (\Sigma \text{ inpatient facilities constituent ratio} \times \text{average hospitalization medical expense in inpatient institutions at different levels})] \quad (1)$$

In the above formula, *DEBDi* is the direct economic burden for the class *i* disease, and *N* is the number of people in the research population.

Indirect Economic Burden of Disease (IEBD)

IEBD refers to a society and family's loss between the current value and future potential value due to illness that may result in the reduction of effective working time and the ability to

TABLE 1 | Direct economic burden of 30 types of diseases in urban areas in China in 2013.

Name of disease	2-week consultation rate (‰)	Weighted average outpatient costs ($)	Outpatient costs ($billion)	Inpatient rate (‰)	Weighted average inpatient costs ($)	Inpatient costs ($billion)	Direct economic burden ($billion)
Viral hepatitis	0.46	28.73	2.50	0.79	1,237.78	7.19	9.69
Infiltrative pulmonary tuberculosis	0.46	28.73	2.50	0.79	1,217.77	7.08	9.57
Acute myocardial infarction	3.37	28.73	18.39	8.07	2,871.28	169.33	187.72
Congestive heart failure	3.37	28.73	18.39	8.07	1,141.58	67.32	85.71
Bacterial pneumonia	0.51	28.73	2.78	2.77	805.43	16.29	19.07
Chronic pulmonary heart disease	3.37	28.73	18.39	8.07	1,278.13	75.38	93.76
Acute upper gastrointestinal bleeding	7.63	28.73	41.68	9.78	1,241.93	88.83	130.51
Primary nephrotic syndrome	0.34	28.73	1.84	0.53	1,010.48	3.93	5.77
Hyperthyroidism	8.85	28.73	48.32	4.74	865.29	29.97	78.29
Cerebral hemorrhage	2.26	28.73	12.32	7.06	2,475.15	127.79	140.12
Cerebral infarction	2.26	28.73	12.32	7.06	1,302.97	67.27	79.60
Aplastic anemia	0.34	28.73	1.84	0.53	1,130.33	4.40	6.24
Acute leukemia	0.34	28.73	1.84	0.53	2,131.53	8.30	10.14
Nodular goiter	0.34	28.73	1.84	1.98	1,422.72	20.59	22.43
Acute appendicitis	7.63	28.73	41.68	9.78	1,020.75	73.01	114.69
Acute cholecystitis	0.68	28.73	3.73	2.04	1,194.06	17.80	21.53
Inguinal hernia	1.48	28.73	8.11	1.24	947.41	8.57	16.67
Malignant gastric tumor	0.91	28.73	5.00	4.90	2,492.69	89.35	94.35
Pulmonary malignant tumor	0.91	28.73	5.00	4.90	1,807.28	64.78	69.78
Esophageal malignant tumor	0.91	28.73	5.00	4.90	2,406.22	86.25	91.25
Myocardial infarction coronary artery bypass	3.37	28.73	18.39	8.07	8,073.60	476.13	494.52
Malignant neoplasm of bladder	0.91	28.73	5.00	4.90	2,139.75	76.70	81.69
Benign prostatic hyperplasia	3.27	28.73	17.85	5.23	1,553.33	59.43	77.28
Intracranial injury	1.08	28.73	5.90	1.70	1,850.28	23.02	28.92
Prolapse of lumbar intervertebral disc	6.51	28.73	35.58	6.12	1,308.20	58.51	94.09
Bronchial pneumonia in children	3.37	28.73	18.39	2.77	432.84	8.75	27.14
Infectious diarrhea in children	3.37	28.73	18.39	9.78	269.22	19.26	37.64
Leiomyoma of uterus	3.37	28.73	18.39	1.24	1,428.34	12.92	31.30
Cesarean section	0.11	28.73	0.61	9.51	944.71	65.71	66.32
Senile cataract	0.73	28.73	4.01	2.00	907.82	13.28	17.30

work (21). This is also known as indirect costs. Indirect costs represent the value given by society to health and life. While it broadly includes social productivity losses, losses of income, losses of housework, employment costs, training costs, insurance costs, management costs, etc., indirect costs indicate the loss of productivity in this narrow sense. We used the human capital approach (22) to estimate the indirect economic burden of 30 types of diseases and calculate the economic loss of missed working time due to hospitalization. The formula is as follows:

$IEBDi = N \times$ hospitalization rate $\times$ (Σ inpatient facilities constituent ratio $\times$average hospitalization days in inpatient institutions at different levels) $\times$ GDP/365 (2)

In the above formula, *IEBDi* is the indirect economic burden for the class *i* disease, and *N* is the number of people in the research population.

RESULTS

DEBD

We weighted the average outpatient medical costs that were not caused by disease according to the constituent ratio of first diagnosis mechanism for urban outpatients over 2 weeks. The proportions of average outpatient medical costs in tertiary, secondary and primary hospitals were 14.46, 18.18, and 67.36%, respectively. In 2013 for urban patients, there were 16.04% distributed in tertiary hospitals, 69.28% distributed in secondary hospitals, and 14.68% distributed in primary hospitals. This distribution was used to weigh the average medical costs of hospitalization. China had a population of 731.11 million urban residents in 2013. The result of the outpatient costs, hospitalization costs and

TABLE 2 | Direct economic burden of 30 types of diseases in rural areas in China in 2013.

Name of disease	2-week consultation rate (‰)	Weighted average outpatient costs ($)	Outpatient costs ($billion)	Inpatient rate (‰)	Weighted average inpatient costs ($)	Inpatient costs ($billion)	Direct economic burden ($billion)
Viral hepatitis	0.43	25.75	1.80	1.25	1,088.56	8.55	10.35
Infiltrative pulmonary tuberculosis	0.43	25.75	1.80	1.25	993.99	7.81	9.61
Acute myocardial infarction	2.84	25.75	11.96	5.79	2,133.82	77.85	89.81
Congestive heart failure	2.84	25.75	11.96	5.79	1,005.15	36.67	48.63
Bacterial pneumonia	0.63	25.75	2.67	3.11	625.97	12.25	14.92
Chronic pulmonary heart disease	2.84	25.75	11.96	5.79	1,047.51	38.22	50.18
Acute upper gastrointestinal bleeding	9.59	25.75	40.42	10.55	1,031.80	68.53	108.95
Primary nephrotic syndrome	0.48	25.75	2.01	0.72	854.14	3.87	5.88
Hyperthyroidism	3.93	25.75	16.56	2.34	733.45	10.80	27.35
Cerebral hemorrhage	2.19	25.75	9.22	6.74	2,104.12	89.33	98.55
Cerebral infarction	2.19	25.75	9.22	6.74	1,063.90	45.17	54.39
Aplastic anemia	0.48	25.75	2.01	0.72	868.45	3.94	5.95
Acute leukemia	0.48	25.75	2.01	0.72	1,546.24	7.01	9.02
Nodular goiter	0.27	25.75	1.14	2.05	1,184.70	15.26	16.40
Acute appendicitis	9.59	25.75	40.42	10.55	862.14	57.26	97.68
Acute cholecystitis	0.88	25.75	3.73	1.86	934.11	10.94	14.67
Inguinal hernia	1.52	25.75	6.40	1.13	806.33	5.75	12.15
Malignant gastric tumor	0.61	25.75	2.55	2.98	1,862.85	34.94	37.50
Pulmonary malignant tumor	0.61	25.75	2.55	2.98	1,406.94	26.39	28.95
Esophageal malignant tumor	0.61	25.75	2.55	2.98	1,901.10	35.66	38.22
Myocardial infarction coronary artery bypass	2.84	25.75	11.96	5.79	7,285.04	265.80	277.76
Malignant neoplasm of bladder	0.61	25.75	2.55	2.98	1,744.04	32.72	35.27
Benign prostatic hyperplasia	3.12	25.75	13.16	5.58	1,335.90	46.94	60.10
Intracranial injury	0.96	25.75	4.06	1.47	1,480.87	13.69	17.75
Prolapse of lumbar intervertebral disc	7.83	25.75	33.02	5.83	934.17	34.29	67.32
Bronchial pneumonia in children	2.84	25.75	11.96	3.11	363.62	7.11	19.07
Infectious diarrhea in children	2.84	25.75	11.96	10.55	255.41	16.96	28.92
Leiomyoma of uterus	2.84	25.75	11.96	1.13	1,199.35	8.56	20.52
Cesarean section	0.08	25.75	0.33	10.02	810.88	51.16	51.50
Senile cataract	0.64	25.75	2.70	1.72	743.22	8.04	10.74

DEBD in urban China are shown in **Table 1**, which is based on Formula (1). The largest DEBD in urban areas was myocardial infarction coronary artery bypass (MICAB), and the second largest was acute myocardial infarction (AMI), while the smallest DEBD was primary nephrotic syndrome (PNS).

Similar to the method used to process urban medical costs, we weighted the average rural outpatient medical costs. As a result, the proportions of average rural outpatient medical costs in tertiary, secondary and primary hospitals for rural outpatients were 2.53, 16.31, and 81.16% in 2013, respectively. In 2013, there were 13.11% of rural inpatients distributed in tertiary hospitals, 56.61% distributed in secondary hospitals, and 30.28% distributed in primary hospitals. China had a rural population of 629.61 million residents in 2013. The result of the outpatient costs, hospitalization costs and DEBD in rural China are shown in **Table 2**, which is based on Formula (1). The largest DEBD in rural areas was MICAB, and the second largest was acute upper

TABLE 3 | Indirect economic burden of 30 types of diseases in urban areas in China in 2013.

Name of disease	Inpatient rate (‰)	Weighted inpatient days (day)	Daily GDP per capita ($)	Indirect economic burden ($billion)
Viral hepatitis	0.79	15.70	18.65	1.70
Infiltrative pulmonary tuberculosis	0.79	12.84	18.65	1.39
Acute myocardial infarction	8.07	9.08	18.65	9.99
Congestive heart failure	8.07	9.90	18.65	10.88
Bacterial pneumonia	2.77	8.78	18.65	3.31
Chronic pulmonary heart disease	8.07	11.11	18.65	12.22
Acute upper gastrointestinal bleeding	9.78	7.79	18.65	10.39
Primary nephrotic syndrome	0.53	12.12	18.65	0.88
Hyperthyroidism	4.74	8.89	18.65	5.75
Cerebral hemorrhage	7.06	14.93	18.65	14.38
Cerebral infarction	7.06	11.64	18.65	11.21
Aplastic anemia	0.53	8.36	18.65	0.61
Acute leukemia	0.53	14.91	18.65	1.08
Nodular goiter	1.98	8.19	18.65	2.21
Acute appendicitis	9.78	6.87	18.65	9.16
Acute cholecystitis	2.04	8.50	18.65	2.36
Inguinal hernia	1.24	6.94	18.65	1.17
Malignant gastric tumor	4.90	13.61	18.65	9.09
Pulmonary malignant tumor	4.90	14.11	18.65	9.43
Esophageal malignant tumor	4.90	15.79	18.65	10.55
Myocardial infarction coronary artery bypass	8.07	18.50	18.65	20.34
Malignant neoplasm of bladder	4.90	14.06	18.65	9.40
Benign prostatic hyperplasia	5.23	11.67	18.65	8.32
Intracranial injury	1.70	12.27	18.65	2.85
Prolapse of lumbar intervertebral disc	6.12	11.00	18.65	9.17
Bronchial pneumonia in children	2.77	6.88	18.65	2.60
Infectious diarrhea in children	9.78	4.94	18.65	6.59
Leiomyoma of uterus	1.24	9.44	18.65	1.59
Cesarean section	9.51	6.69	18.65	8.68
Senile cataract	2.00	4.26	18.65	1.16

gastrointestinal bleeding (AUGIB), while the smallest DEBD was PNS.

IEBD

We weighted average hospitalization days by disease based on the constituent ratio for China's urban inpatients in 2013. The GDP per capita was $6,807 in China in 2013. The indirect economic losses of urban patients caused by hospitalization are shown in **Table 3**, which is based on Formula (2). The largest IEBD in urban areas was MICAB, which was followed by cerebral hemorrhage (CH), and the smallest IEBD was aplastic anemia (AA).

Similar to the method used for dealing with urban patients' data, we weighted average hospitalization days by disease to obtain rural patients' hospitalization days by disease. The indirect economic losses of rural patients due to hospitalization are demonstrated in **Table 4**, which is based on Formula (2). The largest IEBD in rural areas was CH, which was followed by MICAB, and the smallest IEBD was AA.

Total Economic Burden of Disease (TEBD)

The total health expenditure of China in 2013 was $517.64 billion. The top five diseases among 30 types of diseases that had the heaviest TEBD based on the proportion in total health expenditure were MICAB (15.51%), AMI (5.67%), CH (5.11%), AUGIB (5.01%) and acute appendicitis (AAs) (4.45%). The last five diseases that had the smallest TEBD were AA (0.26%), PNS (0.26%), acute leukemia (AL) (0.41%), infiltrative pulmonary tuberculosis (IPT) (0.43%) and viral hepatitis (VH) (0.46%). The average TEBD for cities was $81.39 billion and for the villages was $50.26 billion, respectively. The total economic burden of 30 types of diseases in urban and rural areas in China is shown in **Table 5**.

DISCUSSION

This study integrally estimated the EBD of China. The TEBD was between $13.39 and 803.00 billion, which included 30 types of common diseases in China. This accounted for 2.54% of

TABLE 4 | Indirect economic burden of 30 types of diseases of rural area in China, 2013.

Name of disease	Inpatient rate (‰)	Weighted average inpatient days (day)	Daily GDP per capita ($)	Indirect economic burden ($billion)
Viral hepatitis	1.25	15.58	18.65	2.28
Infiltrative pulmonary tuberculosis	1.25	12.50	18.65	1.83
Acute myocardial infarction	5.79	8.94	18.65	6.09
Congestive heart failure	5.79	9.86	18.65	6.71
Bacterial pneumonia	3.11	8.46	18.65	3.09
Chronic pulmonary heart disease	5.79	10.74	18.65	7.31
Acute upper gastrointestinal bleeding	10.55	7.61	18.65	9.42
Primary nephrotic syndrome	0.72	11.55	18.65	0.98
Hyperthyroidism	2.34	8.61	18.65	2.36
Cerebral hemorrhage	6.74	14.39	18.65	11.39
Cerebral infarction	6.74	11.16	18.65	8.83
Aplastic anemia	0.72	7.01	18.65	0.59
Acute leukemia	0.72	12.37	18.65	1.05
Nodular goiter	2.05	8.17	18.65	1.96
Acute appendicitis	10.55	6.95	18.65	8.60
Acute cholecystitis	1.86	8.03	18.65	1.75
Inguinal hernia	1.13	7.01	18.65	0.93
Malignant gastric tumor	2.98	13.03	18.65	4.56
Pulmonary malignant tumor	2.98	13.88	18.65	4.86
Esophageal malignant tumor	2.98	15.06	18.65	5.27
Myocardial infarction coronary artery bypass	5.79	15.25	18.65	10.38
Malignant neoplasm of bladder	2.98	13.74	18.65	4.81
Benign prostatic hyperplasia	5.58	11.27	18.65	7.39
Intracranial injury	1.47	11.54	18.65	1.99
Prolapse of lumbar intervertebral disc	5.83	10.43	18.65	7.14
Bronchial pneumonia in children	3.11	6.55	18.65	2.39
Infectious diarrhea in children	10.55	4.73	18.65	5.87
Leiomyoma of uterus	1.13	9.39	18.65	1.25
Cesarean section	10.02	6.62	18.65	7.79
Senile cataract	1.72	4.42	18.65	0.89

the average proportion of the national total health expenses; although, the highest proportion was 15.51%. This proportion is enough to capture our government's attention. The DEBD of cities was $5.77–494.52 billion, while the DEBD of villages was $5.88–277.76 billion. The IEBD for citizens was $0.61–20.34 billion, and the IEBD for the rural population was $0.59–11.39 billion. Whether in cities or countryside, the DEBD was much higher than the IEBD.

The DEBD, IEBD, and TEBD of VH, IPT, and PNS in urban areas were slightly lower than those in rural areas; however, the residual EBD of 27 types of diseases for the cities were higher than those for the countryside. Interestingly, the largest difference for the EBD of MICAB between urban and rural areas, which exceeded $226.72 billion for TEBD, was a relatively large difference for DEBD but a slight difference for IEBD. There is a possible explanation for this difference. One possible reason could be that there were more MICAB patients in cities than in the countryside because rural patients failed to visit a doctor because they were worried about incurring expensive medical bills.

The EBD had a large difference among different diseases. The average total economic burden of 30 types of diseases was $131.65 billion in China. The largest five diseases for TEBD were MICAB, AMI, CH, AUGIB, and AAs. The smallest five diseases for TEBD were AA, PNS, AL, IPT, and VH. The middle position for TEBD was held by malignant tumors.

It is important to note that this study has some limitations. The following costs or intangible economic burdens were not taken into consideration because of limited data: self purchased medical fees, the cost of time in bed, and cost of recovery time at home. Additionally, we could not measure the production value of the loss of life due to premature death and the loss of production value due to long-term disability caused by illness or disability, which will likely result in underestimating the EBD. The misclassification of the diseases may also result in errors.

TABLE 5 | Total economic burden of 30 types of diseases in China in 2013.

Name of disease	National		Urban		Rural	
	Total economic burden ($billion)	Proportion in total health expenses (%)	Total economic burden ($billion)	Proportion in total health expenses (%)	Total economic burden ($billion)	Proportion in total health expenses (%)
Viral hepatitis	24.02	0.46	11.39	0.22	12.63	0.24
Infiltrative pulmonary tuberculosis	22.40	0.43	10.96	0.21	11.44	0.22
Acute myocardial infarction	293.61	5.67	197.71	3.82	95.90	1.85
Congestive heart failure	151.93	2.94	96.59	1.87	55.34	1.07
Bacterial pneumonia	40.39	0.78	22.38	0.43	18.01	0.35
Chronic pulmonary heart disease	163.47	3.16	105.98	2.05	57.49	1.11
Acute upper gastrointestinal bleeding	259.28	5.01	140.91	2.72	118.37	2.29
Primary nephrotic syndrome	13.52	0.26	6.65	0.13	6.86	0.13
Hyperthyroidism	113.75	2.20	84.04	1.62	29.72	0.57
Cerebral hemorrhage	264.44	5.11	154.5	2.98	109.95	2.12
Cerebral infarction	154.03	2.98	90.81	1.75	63.23	1.22
Aplastic anemia	13.39	0.26	6.85	0.13	6.54	0.13
Acute leukemia	21.29	0.41	11.22	0.22	10.07	0.19
Nodular goiter	43.00	0.83	24.64	0.48	18.36	0.35
Acute appendicitis	230.14	4.45	123.85	2.39	106.28	2.05
Acute cholecystitis	40.31	0.78	23.89	0.46	16.42	0.32
Inguinal hernia	30.93	0.60	17.85	0.34	13.09	0.25
Malignant gastric tumor	145.5	2.81	103.44	2.00	42.06	0.81
Pulmonary malignant tumor	113.01	2.18	79.21	1.53	33.80	0.65
Esophageal malignant tumor	145.28	2.81	101.80	1.97	43.48	0.84
Myocardial infarction coronary artery bypass	803.00	15.51	514.86	9.95	288.14	5.57
Malignant neoplasm of bladder	131.17	2.53	91.10	1.76	40.08	0.77
Benign prostatic hyperplasia	153.09	2.96	85.60	1.65	67.49	1.30
Intracranial injury	51.50	0.99	31.76	0.61	19.74	0.38
Prolapse of lumbar intervertebral disc	177.72	3.43	103.26	1.99	74.46	1.44
Bronchial pneumonia in children	51.20	0.99	29.74	0.57	21.46	0.41
Infectious diarrhea in children	79.02	1.53	44.23	0.85	34.79	0.67
Leiomyoma of uterus	54.66	1.06	32.89	0.64	21.77	0.42
Cesarean section	134.29	2.59	75.00	1.45	59.29	1.15
Senile cataract	30.09	0.58	18.46	0.36	11.63	0.22

CONCLUSION

The EBD reflects the burden caused by illness for a society. If we can reduce or eliminate EBD, socioeconomic losses will decrease, and our society will benefit. However, policy makers must make allocations with limited resources. Therefore, this paper provides solid data and research for policy makers to make informed decisions. Overall, there was a large difference in the economic burden of different diseases, and the total economic burden of urban patients' disease was larger than that of rural patients. The top ten diseases were myocardial infarction coronary artery bypass, acute myocardial infarction, cerebral hemorrhage, acute upper gastrointestinal bleeding, acute appendicitis, prolapse of lumbar intervertebral disc, chronic pulmonary heart disease, cerebral infarction, benign prostatic hyperplasia and congestive heart failure.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://www.stats.gov.cn/, http://www.nhc.gov.cn/.

AUTHOR CONTRIBUTIONS

LL and QM contributed to the design of the study and project management and contributed to the final interpretation of data. XS was the lead qualitative researcher and designed the process evaluation. QM developed the coding framework with LL and

TZ. Coding and initial data interpretation was performed by LL, TZ, and JY. The first draft of the manuscript was produced by XS and LL. All authors critically reviewed and edited the draft paper, read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank National Natural Science Foundation of China (No. 81960617).

REFERENCES

1. Zhuang RS, Wang SY, Liang WN, Jing CX, Li B, et al. Studies on direct and indirect economic burden of disease and related factor in countryside of Qingdao city in 2001. *Zhonghua liuxingbingxue zazhi.* (2003) 24:196. doi: 10.3760/j.issn:0254-6450.2003.03.009
2. He QY, Zhou X, Xie CM, Liang ZA, Chen P, Wu CG. Impact of chronic obstructive pulmonary disease on quality of life and economic burden in Chinese urban areas. *Chin J Tuberculosis Respir Dis.* (2009) 32:253–7. doi: 10.3760/cma.j.issn.1001-0939.2009.04.006
3. Carlsson P. Health technology assessment and priority setting for health policy in Sweden. *Int J Technol Assess Health Care.* (2004) 20:44–54. doi: 10.1017/S0266462304000777
4. Perleth M, Jakubowski E, Busse R. What is 'best practice' in health care? State of the art and perspectives in improving the effectiveness and efficiency of the European health care systems. *Health Policy.* (2001) 56:235–50. doi: 10.1016/S0168-8510(00)00138-X
5. Eichler HG, Kong SX, Gerth WC, Mavros P, Jönsson B. Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge? *Value Health.* (2004) 7:518–28. doi: 10.1111/j.1524-4733.2004.75003.x
6. Simoens S. Health economic assessment: a methodological primer. *Int J Environ Res Public Health.* (2009) 6:2950–66. doi: 10.3390/ijerph6122950
7. Lou P, Zhu Y, Chen P, Zhang P, Yu J, Zhang N, et al. Vulnerability, beliefs, treatments and economic burden of chronic obstructive pulmonary disease in rural areas in China: a cross-sectional study. *BMC Public Health.* (2012) 12:1–7. doi: 10.1186/1471-2458-12-287
8. Wang B, Chen Y, Zhang J, Li J, Guo Y, Hailey D. A preliminary study into the economic burden of cerebral palsy in China. *Health Policy.* (2008) 87:223–34. doi: 10.1016/j.healthpol.2008.01.001
9. Oh IH, Yoon SJ, Seo HY, Kim EJ, Kim YA. The economic burden of musculoskeletal disease in Korea: a cross sectional study. *BMC Musculoskeletal Disord.* (2011) 12:157. doi: 10.1186/1471-2474-12-157
10. Liu JL, Maniadakis N, Gray A, Rayner M. The economic burden of coronary heart disease in the UK. *Br Heart J.* (2002) 88:597–603. doi: 10.1136/heart.88.6.597
11. Woo J, Ho SC, Chan SG, Yu AL, Yuen YK, Lau J. An estimate of chronic disease burden and some economic consequences among the elderly Hong Kong population. *J Epidemiol Community Health.* (1997) 51:486–9. doi: 10.1136/jech.51.5.486
12. Liu C, Shi J, Wang H, Yan X, Wang L, Ren J, et al. Population-level economic burden of lung cancer in China: provisional prevalence-based estimations, 2017–2030. *Chin J Cancer Res.* (2021) 33:79–92. doi: 10.21147/j.issn.1000-9604.2021.01.09
13. Gong H, Shen X, Yan H, Lu WY, Zhong GJ, Dong KG, et al. Estimating the disease burden of seasonal influenza in China, 2006-2019. *Zhong hua yi xue za zhi.* (2021) 101:560–7. doi: 10.3760/cma.j.cn112137-20201210-03323
14. Bloom DE, Chen S, Kuhn M, McGovern ME, Oxley L, Prettner K. The economic burden of chronic diseases: estimates and projections for China, Japan, and South Korea. *J Econ Ageing.* (2020) 17:100163. doi: 10.1016/j.jeoa.2018.09.002
15. Wu P, Xu B, Shen A, He Z, Zhang CJ, Ming WK, et al. The economic burden of medical treatment of children with asthma in China. *BMC Pediatr.* (2020) 20:386. doi: 10.1186/s12887-020-02268-6
16. Clay E, Zhou J, Yi ZM, Zhai S, Toumi M. Economic burden for Alzheimer's disease in China from 2010 to 2050: a modelling study. *J Market Access Health Policy.* (2019) 7:1667195. doi: 10.1080/20016689.2019.1667195
17. Yu E, Liao Z, Fan W, Hu W, Tian G, Chen K, et al. The economic burden of Alzheimer's disease in Zhejiang province. *J Alzheimers Dis.* (2021) 80:539–53. doi: 10.3233/JAD-201285
18. Cui T, Zhang X, Wang Q, Yue N, Zheng M, Wang D, et al. Disease burden concerning hepatitis E-infected inpatients in Jiangsu province, China. *Vaccine.* (2020) 38:673–9. doi: 10.1016/j.vaccine.2019.10.045
19. Cai X, Yang H, Genchev GZ, Lu H, Yu G. Analysis of economic burden and its associated factors of twenty-three rare diseases in Shanghai. *Orphanet J Rare Dis.* (2019) 14:233. doi: 10.1186/s13023-019-1168-4
20. Adamiak G. Methods for the economic evaluation of health care programmes, 3rd ed. *J Epidemiol Community Health.* (2006) 60:822–3. doi: 10.1136/jech.2005/040253corr2
21. Muller CF. Economic costs of illness and health policy. *Am J Public Health.* (1980) 70:1245–6. doi: 10.2105/AJPH.70.12.1245
22. Van Beeck EF, van Roijen L, Mackenbach JP. Medical costs and economic production losses due to injuries in the Netherlands. *J Trauma Acute Care Surg.* (1997) 42:1116–23. doi: 10.1097/00005373-199706000-00023

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 Public Health

BRIEF RESEARCH REPORT
published: 19 May 2021
doi: 10.3389/fpubh.2021.671070

Reducing False-Positives in Lung Nodules Detection Using Balanced Datasets

Jinglun Liang[1], *Guoliang Ye*[1], *Jianwen Guo*[1], *Qifan Huang*[1,2] *and Shaohui Zhang*[1]*

[1] School of Mechanical Engineering, Dongguan University of Technology, Dongguan, China, [2] School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou, China

Malignant pulmonary nodules are one of the main manifestations of lung cancer in early CT image screening. Since lung cancer may have no early obvious symptoms, it is important to develop a computer-aided detection (CAD) system to assist doctors to detect the malignant pulmonary nodules in the early stage of lung cancer CT diagnosis. Due to the recent successful applications of deep learning in image processing, more and more researchers have been trying to apply it to the diagnosis of pulmonary nodules. However, due to the ratio of nodules and non-nodules samples used in the training and testing datasets usually being different from the practical ratio of lung cancer, the CAD classification systems may easily produce higher false-positives while using this imbalanced dataset. This work introduces a filtering step to remove the irrelevant images from the dataset, and the results show that the false-positives can be reduced and the accuracy can be above 98%. There are two steps in nodule detection. Firstly, the images with pulmonary nodules are screened from the whole lung CT images of the patients. Secondly, the exact locations of pulmonary nodules will be detected using Faster R-CNN. Final results show that this method can effectively detect the pulmonary nodules in the CT images and hence potentially assist doctors in the early diagnosis of lung cancer.

Keywords: pulmonary nodule detection, lung image classification, deep learning, convolutional neural network, transfer learning

OPEN ACCESS

Edited by:
Nianyin Zeng,
Xiamen University, China

Reviewed by:
Zhuyun Chen,
South China University of Technology, China
Shuai Li,
Swansea University, United Kingdom

***Correspondence:**
Shaohui Zhang
zhangsh@dgut.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *23 February 2021*
Accepted: *12 April 2021*
Published: *19 May 2021*

Citation:
Liang J, Ye G, Guo J, Huang Q and Zhang S (2021) Reducing False-Positives in Lung Nodules Detection Using Balanced Datasets. Front. Public Health 9:671070. doi: 10.3389/fpubh.2021.671070

INTRODUCTION

Lung cancer is one of the most common cancers in the world and has the highest proportion of new cases (11.6%) and deaths (18.4%) among all cancers in 2018 (1–3). A lack of obvious clinical symptoms in the early stage of the disease are one of the main reasons for the high mortality rate of lung cancer, because most patients have missed the best time for treatment when they go to see a doctor with symptoms. At that time, lung cancer may be too advanced to be effectively treated with surgery. Therefore, the early screening, diagnosis, and treatment of lung cancer is a major focus of lung cancer prevention and control. In the early stage, malignant pulmonary nodules are one of the main manifestations of lung cancer on CT images. Pulmonary nodules are a common disease, which is a small round or oval tissue growing in the lungs. However, it does not mean all pulmonary nodules are malignant, so it is necessary to detect the malignant pulmonary nodules in the lung.

Previous studies show that low-dose spiral CT (CT) is an effective method to analyze lung cancer nodules and reduce mortality compared with chest X-ray photography (4). The whole lung usually needs to be scanned to detect tiny pulmonary nodules, which will produce a huge amount of images

for each patient. It is a serious challenge for radiologists to detect pulmonary nodules from so many CT images by excluding uncorrelated tissues such as bronchi and blood vessels. It could easily lead to misdiagnosis due to the radiologist's fatigue and distraction caused by the overwhelming number of images. Therefore, it is urgent to develop an automated efficient system for pulmonary nodules screening.

A Computer-Aided-Detection (CAD) system is one of the feasible methods to assist doctors to detect and classify lung images. Furthermore, a promising CAD system not only can reduce the missed detection of pulmonary nodules but also can improve the accuracy and efficiency of the analysis of CT images. There are several commercial CAD systems in clinical use, such as the Lung Care commercial CAD system developed by Siemens and the Image Checker CTLN-1000 commercial CAD system developed by R2 (5). A typical CAD system mainly consists of 3 parts, including pre-processing, nodule candidate detection, and false-positive reduction (6). The main purpose of pre-processing is to standardize data, enhance images and reduce noise. The detection stage after pre-processing uses very high sensitivity to detect all nodules as much as possible, forming candidate nodules with many false positives. The previous stage produces a large number of candidate nodules, so this stage is mainly to reduce the number of false positives in the candidate and generate the final result.

In the past decade, especially in combination with image processing methods (7), machine learning methods have been studied extensively in the medical detection field. In addition, researchers have also obtained many related achievements (8, 9). The pulmonary nodules detection in lung CT images is one of the complex and highly concerning problems in medical image processing. The conventional lung nodule detection algorithms detect the pulmonary nodules by processing and analyzing CT images mainly through combining the pulmonary nodule characteristics such as size, location, and shape with the image processing algorithm and machine learning (10). These detection algorithms include two steps. The region of interest (ROI) was firstly extracted by image segments to identify suspected targets, and secondly, the ROI will be classified as malignant and non-malignant pulmonary nodules. Messay et al. pre-processed the image by combining the multi-layer gray threshold method with the open operation in morphology and then extracted the ROI of pulmonary nodules by rule-based analysis method (11). Murphy et al. used cluster structure and cluster merging to extract the nodules by calculating the shape index and curviness of each pixel in the image of the lung parenchyma, and setting thresholds for these two parameters to obtain seed points (10). Although conventional machine learning methods can achieve high accuracy in one type of target, they are still hard to achieve good results in other types of modules. This is mainly due to the complexity of different types of pulmonary nodules and the complex situations in classification.

With the development of artificial intelligence (AI) and the increasingly reliable large-scale annotated datasets, a series of deep learning algorithms has experienced an enormous advance in the image processing and video processing field, especially the deep convolutional neural network (CNN). Meanwhile, many improvements and optimization algorithms were proposed to improve the effect (12), and it has gradually made achievements in the field of medical image recognition (13, 14). CNN can automatically learn image features from sufficient training data. Automatic extraction of pulmonary nodules features can adapt to different types of pulmonary nodules. It can avoid the problems of omission or incomplete features in artificial extraction features and also improve the automation of pulmonary nodules detection (15). Therefore, compared with the traditional computer-aided lung image diagnosis method the deep learning algorithm has great advantages. It shows a significant improvement in the detection of the pulmonary nodules and it is gradually becoming the mainstream method in the field of pulmonary nodules detection and more and more research results have been achieved.

Anirudhi et al. proposed a method based on a 3D CNN for pulmonary nodules detection. It used weakly labeled data to train the neural networks in the field of pulmonary nodules detection. The experimental results were also superior to the traditional methods and the method produced fewer false positives (16, 17). Golan et al. proposed a method for lung nodule detection based on a deep CNN, which used the open data set LIDC and IDRI with a CNN to extract the characteristics of the lung nodules. This method does not use the segmentation and false-positive method, but still obtains a good result (15). Fu et al. proposed a feature extraction method of pulmonary nodules based on CNN. This method can effectively extract the brightness, shape, material, and other features of pulmonary nodules. This method can effectively combine the features extracted by hand with the features extracted by the neural network (18). Li et al. used an integrated CNN to solve the problem of high false-positives. This method combined three CNNs and achieved good results on the JSRT data set (19). Although deep learning can detect pulmonary nodules with various characteristics, it still produces many false positives. Thus, the high false-positive rate is the key issue of using deep learning techniques in pulmonary nodules.

This paper intends to solve the mentioned problem from the perspective of preliminary screening. The main idea is to introduce a filtering step to remove irrelevant images before testing to reduce false positives. This new method is based on deep learning which is divided into two steps. The first step is to screen the images with pulmonary nodules from the whole lung CT image of the patients and the second step is to detect the exact locations of pulmonary nodules using Faster R-CNN. In the first step, a classification network is trained to screen the target images with suspected pulmonary. The experimental results on a balanced LUNA16 dataset have achieved an accuracy of above 98%. In this way, the suspected pulmonary nodules in the CT image of the case can be quickly selected by the classification network, and the location range of the suspected nodules in the CT image can be detected by the FasterRCNN. This can improve the efficiency of the detector, increase the reliability of the detection, and reduce the false positive of the detector.

The main framework of this paper is as follows. In section Background, we briefly introduce the relative networks, including Alexnet, Resnet, and FasterRCNN. In section Materials and Methods, we describe the LUNA16 dataset, the CT

image processing, and the experimental method. In section Experimental Results and Analysis we explain the experimental results and analysis. The last section is the conclusion of this paper.

BACKGROUND

Transfer Learning

Because of the increasing popularity of deep neural networks, more and more scholars are adopting deep learning to solve complex problems in the field of medical imaging. At present, there are mainly three feasible methods to successfully apply CNNs to medical images: (1) Training CNN from the ground up, (2) conducting unsupervised CNN pre-training with the supervised fine-tuning base on off-the-shelf pre-trained CNN features, (3) transfer learning (20). Since the pre-trained network has learned a wealth of image features, it is usually faster and easier to use the pre-trained network with transfer learning compared to training the network from scratch. ImageNet is a famous database in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), and the majority of the pre-trained networks are trained on it (21). It can learn features specific to the new data set when fine-tuning the network. In the following, we introduce the networks that were used in this paper, Alexnet Faster R-CNN and ResNet.

Alexnet

For screening suspected pulmonary, we train a classification model base on Alexnet in the first step. Alexnet is a typical deep convolutional neural network and the champion of the ImageNet 2012 Image Recognition Challenge. The Alexnet network has a total of 60 million parameters and 650,000 neurons, consisting of five convolutional layers and three fully-connected layers. After the first, third, and fifth convolutional layers are followed by the max-pooling layer, the last fully-connected layer has a 1000-way softmax (22). The main feature of the Alexnet networks includes the following four parts:

(1) The Relu activation function is added at the end of each Conv layer, which solves the problem of gradient disappearance of Sigmoid and makes the convergence faster. The common linear rectifier functions are the ramp function, which is

$$f(x) = \max(0, x) \tag{1}$$

Here x is the input to the neuron.

(2) To reduce overfitting, it employed the random discard technique (Dropout) in the fully-connected layers. Dropout is regularization method in which some neural network units are temporarily discarded (their weights are retained) from the network according to a certain probability and no longer respond to the forward and reverse transmission of data in the training process of the deep learning network. At the same time, the data set is artificially enlarged, including image translation, horizontal reflection, and changing the intensity of the RGB channel in the training image.

(3) It has also added a layer of Normalization (Local Response Normalization), which makes it more accurate. ReLU function does not have a limited range like tanh and Sigmoid, so it needs to be normalized after ReLU. The idea of LRN originated from a concept called lateral inhibition in neurobiology, which means that the activated neuron inhibits the surrounding neurons.

The core idea of Local Response Normalization is to normalize by using neighboring data, as shown in the following formula:

$$b^i_{x,y} = a^i_{x,y} / (k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a^i_{x,y})^2)^\beta \tag{2}$$

Where $a^i_{x,y}$ acts on the convolution kernel at position (x, y), and then performs ReLU, the resulting neuron output. N is the total number of convolution kernels of this layer. n is the number of convolution kernels adjacent to the same location. k, α, β are hyper-parameters.

(4) It used the overlapping maximum pool and obtain greater performance. In general, there will not be overlap between adjacent sliding windows, which means the pooling unit's step size s equal to the pooling unit's size z. On the contrary, it set s < z in Alexnet networks.

Faster-RCNN

In fact, detecting pulmonary nodules is one of the target detection of lung images. The target detection methods based on deep learning are mainly divided into two categories: (1) end-to-end, such as YOLO and SSD network architecture, (2) based on Region Proposal like Faster R-CNN. The former has a relatively fast detection speed, while the latter has a relatively high detection rate and accuracy. Therefore, the detection network of this paper adopted Faster R-CNN as the basic network structure.

The feature of the target detection method based on Regional Proposal is to extract the regional proposal from the input image at first, which is to get the Region Of Interest (ROI) of the target. Before the Faster R-CNN was proposed, the common method for obtaining Regional Proposal is Selective Search. Target detection methods such as R-CNN, SPP-NET, Fast R-CNN, etc., use selective search algorithms to extract target regions. Compared with the former algorithm, the groundbreaking development of Faster R-CNN directly computes the candidate box by Region Proposal networks, which makes the target detection speed significantly improved.

The network structure of Faster R-CNN mainly includes four parts:

1) Conv layers. Faster R-CNN as a detection method based on the CNN network target, this part is the basis of the extraction of the features of images, consisting of Conv + ReLU + pooling layer. These feature maps obtained from Conv layers are shared for subsequent Region Proposal Networks (RPN) layer and full connection layer.
2) Region Proposal Networks (RPN). The Region Proposal is generated by the RPN network. This part uses Softmax to classify anchors as positive or negative proposals and

then obtains accurate proposals through bounding box regression correction.

3) Anchors. The key in RPN is anchors, which is a set of rectangular boxes, in which each row has four values (x1, y1, x2, y2), representing the coordinates of the upper left corner and the lower right corner of the rectangle. Input candidate boxes of different sizes can be obtained by setting different aspect ratios of the anchor. These anchors actually introduce the multi-scale approach that is often used in detecting.
4) ROI Pooling. This part combines the feature maps and Region Proposals from the output of Conv layers and RPN layer, integrating the information to extract proposal feature maps and send them to the subsequent full connection layer to determine the target category. The exact position of the anchor is obtained again by the bounding Box regression.

The Loss used by the entire network is as follows:

$$L\left(\{p_i\},\{t_i\}\right)=\frac{1}{N_{cls}}\sum_i L_{cls}\left(p_i,p_i^*\right)+\lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}\left(t_i,t_i^*\right) \tag{3}$$

In the above formula, i represents the anchor index in the mini-batch and p_i is positive softmax probability. Moreover, p_i^* represents the corresponding ground-truth predict probability, that is, when IOU >0.7 is between the anchor and ground-truth, the anchor is considered to be positive and $p_i^* = 1$; on the contrary, when IOU <0.3, this anchor is considered as negative and $p_i^* = 0$; anchors with 0.3< IOU <0.7 are excluded. t_i represents the predicted bounding box and t_i^* represents the ground-truth box corresponding to the positive anchor. The total loss is divided into two parts. For the regression loss, it is activated only for the positive anchor. The two terms using N_{cls} and N_{reg} to normalize, since in the actual process, the gap between N_{cls} and N_{reg} is too large, use the parameter λ to balance (23).

ResNet

He Kaiming's team proposed a deep convolutional neural network structure ResNet in 2015. Furthermore, the network won the champion of image classification, object detection, and target positioning in the ImageNet (ILSVRC2015) competition.

As CNN can extract the features of multiple levels from the data, and that the more network layers mean can extract richer features at different levels. Therefore, the deeper the network structure, the stronger the ability to extract abstract features, and the richer the semantic information obtained. However, when the depth of the network is increased, the effect of the stochastic gradient descent algorithm will become weaker, which will eventually lead to a gradient disappearance or gradient explosion. The previous network structure can train dozens of layers of the network by standard initialization and regularization layer, but with the further increase of the network layer, the degradation phenomenon will eventually appear, namely as the increase of the number of the network layer the accuracy of both the training set and the test set decreases. It is not caused by overfitting, but by redundant network layers learning parameters that are not identical mappings.

The idea of ResNet is to assume that there is an optimized network layer in a network layer, so often the deep network we design has many redundant network layers. So we want these redundant layers to be able to do identity mapping so that the inputs and outputs that go through the identity layer are exactly the same (24).

The ResNet still allows the non-linear layer to satisfy $H(x,\omega_h)$, and then introduces a short connection directly from the input to the output of the non-linear layer, making the whole mapping to

$$y = H(x,\omega_h) + x \tag{4}$$

This is the core formula of the ResNet. In other words, the ResNet is an operation of network construction, and any network that uses this operation can be called the residual network. Through the experiment on ImageNet, the ResNet can be deepened to hundreds of layers, and higher accuracy of previous convolutional neural networks such as VGGNet and GoogLeNet can be obtained.

MATERIALS AND METHODS

Image Datasets

In this study, we make use of the LUNA16 (LUng Nodule Analysis) datasets for training and testing the deep learning model, which is one of the most representative international data sets in the field of pulmonary nodule detection, providing a publicly available data set about pulmonary nodules in the lung. In LUNA16, the data was collected from LIIDC-IDRI data sets, which is the largest public database for lung nodules (25–27). Based on LIDC-IDRI, LUNA16 datasets retain scans with a thickness of ≤3 mm and eliminate inconsistent or missing slices. Thus, 888 CT scans were screened out of 1,018 scans in LIDC-IDRI. In addition, four experienced thoracic radiologists annotated all LIDC-IDRI scans and there were 36,378 annotations made by the radiologists in these 888 scans. However, only the annotations were categorized as pulmonary nodules larger than 3 mm as relevant lesions. So only the pulmonary nodules that had been annotated by at least three out of the four radiologists were selected.

Finally, 888 scans were screened in the LUNA16 data set. A total of 1,186 nodules were annotated by at least three radiologists, which are lesions that the algorithm should detect. For ease of download, the data is not stored in the generic DOM format, but in Metalmage (.mhd) format, with each image the size of 512×512 pixels.

Image Pre-processing

According to the principle of CT, the value of CT represents that X-ray beams illuminate different parts of the body with different densities to distinguish different tissues and organs. The higher the value of CT, the greater the density of the substance, and HU (Hounsfield unit) is the unit of CT value. The lung CT value is between −600 and −450 HU, the body fat is between −20 and −10 HU, blood is between 13 and 32 HU. For reference, in the state of nature, the CT value of the air is about −1,000 HU, and water is about 0 HU, the bone of the human body is about 1,000

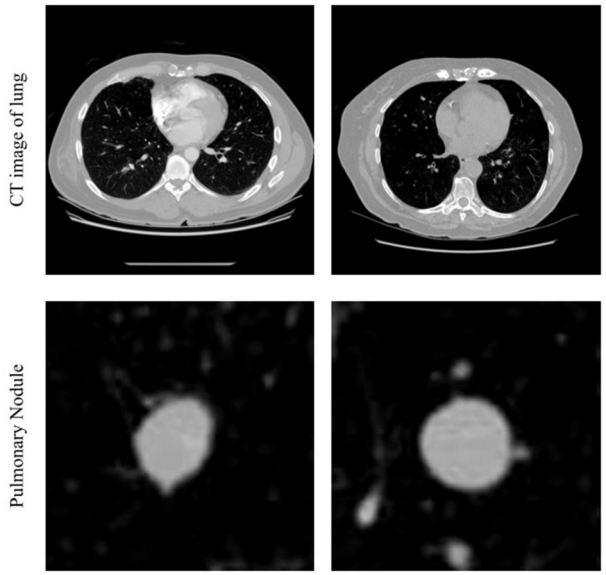


FIGURE 1 | Example of CT images in lung and pulmonary nodule.

HU. The raw CT images include other substances like air and skeleton that we do not focus on, so we filter out the irrelevant substance by a threshold of the CT value to obtain clear images.

We read the CT image from each subset of Luna16, then filtered the values of $< -1{,}000$ HU and >500 HU, and standardized to the range of 0–1. After that, we were able to get the CT images with clear pulmonary nodules, which are shown in **Figure 1** first row. The center of the image in the second row in **Figure 1** shows the pulmonary nodules.

Training

To accurately screen out images with pulmonary nodules, images with pulmonary nodules were taken as positive samples, whereas images without pulmonary nodules were taken as negative samples. To test the effect of sample balance on the model, several models were trained according to different positive and negative sample ratios. In this study, all experiments are conducted on a computer with an Intel Xeon Silver 4210 CPU and two NVIDIA GeForce RTX 2080 Ti GPU, and the software environment is Win10 and MATLAB 2020a.

The details of the experiment setting are as follows:

1) In the training dataset, images from LUNA16 datasets were mainly divided into the following two parts: The positive samples taken from the 1,186 CT images with annotations by the radiologists mentioned earlier. The negative samples mainly consist of total lung CT images from several cases that were not annotated. According to the different proportions of the number of positive samples and negative samples, several different groups were constructed for training and testing their respective effects.
2) To adapt the pre-training network to the new data set, the network was fine-tuned as follows: The last three layers of the network were removed and replaced with a fully connected layer, a softmax layer, and a classification output layer. The output of the fully connected layer was set to 2, which is the same size as the number of classes in the new data, and both WeightLearnRateFactor and BiasLearnRateFactor were 20.
3) All the samples used for training were resized to 227*227 pixels to adapt the input layer of the pre-train network Alexnet. The data set is divided into the training set and test set for cross-validation. The mini-batch size was set to 128, the learning rate was 0.001, and the max epoch number was 200.

After filtering the images, we trained the Faster R-CNN model to detect pulmonary nodules. According to the annotation information of 1,186 pulmonary nodules, the pulmonary nodules were annotated to generate the annotation files needed for training. After the selective node ROI was extracted from the RPN, Resnet101 was used to determine whether the region was a nodule or not. The training option was mainly as follows, the optimization method was stochastic gradient descent method, 500 epoch, 1 mini-batch size, 0.0001 initial learning rate, 0.1 "LearnRateDropFactor," 100 "LearnRateDropPeriod."

EXPERIMENTAL RESULTS AND ANALYSIS

Results on Difference Proportion Sample

We conducted computer experiments on the LUNA16 dataset to prove the effectiveness of the method. **Table 1** summarizes the performance of the model with different ratios of training data, including accuracy, True Positive Rate (TPR), and False Positive Rate (FPR). It is obvious that the accuracy could reach a maximum of 99.43% in the first row of **Table 1** when the ratio of Pulmonary nodule and Non-Pulmonary nodule is near 1, that is, the number of positive samples and negative samples is balanced (28).

In addition to this, it does not perform very well in the case of unbalanced proportions of samples in the training set. Due to the serious imbalance of samples, networks are prone to treat the test sets as a category with a large sample size. Although the accuracy performance is high, the recall rate is low. As can be seen from **Table 1**, when the ratio increases from 1:1 to 1:10 and then to 1:60, the TPR also decreases continuously. Because of the unbalanced sample proportion, increasing the sample size will lead to a worse effect. However, there were only 1,086 positive samples, so the generalization ability of the model is not strong.

Results on Detection

Then we introduced the CT images of the patients into the classification network and classified several CT images before and after the location of pulmonary nodules. After determining the range of pulmonary nodules, the Faster RCNN model was used to detect pulmonary nodules.

We visualize the results of the test on the test set to verify their accuracy. The test results were shown in **Figure 2**, in which CT images with pulmonary nodules were selected from the test set for testing. The first row in the figure is marked with the annotation

TABLE 1 | Summary of inference different ratio of training data, presenting the accuracy.

Number of images		Evaluate		
Positive	**Negative**	**Accuracy**	**TPR**	**FPR**
1,086	1,171 (4 cases)	99.43%	98.90% (347 TP/351 P)	0% (0 FP/351 N)
1,086	1,353 (5 cases)	99.08%	98.00% (344 TP/351 P)	0% (0 FP/406 N)
1,086	1,029 (4 cases)	99.39%	98.90% (347 TP/351 P)	0% (0 FP/309 N)
1,086	10,166 (50 cases)	97.68%	78.06% (274 TP/351 P)	0% (0 FP/3050 N)
1,086	67,845 (287 cases)	98.91%	32.48% (114 TP/351 P)	0.01% (3 FP/20350 N)
201	201(1 case)	99.20%	98.30% (59 TP/60 P)	0% (0 FP/60 N)
494	494(2 cases)	98.00%	95.90% (142 TP/148 P)	0% (0 FP/148 N)
753	753(3 cases)	99.30%	98.70% (223 TP/226 P)	0% (0 FP/226 N)

TP, True Positive; P, Positive; FP False Positive; N, Negatives.

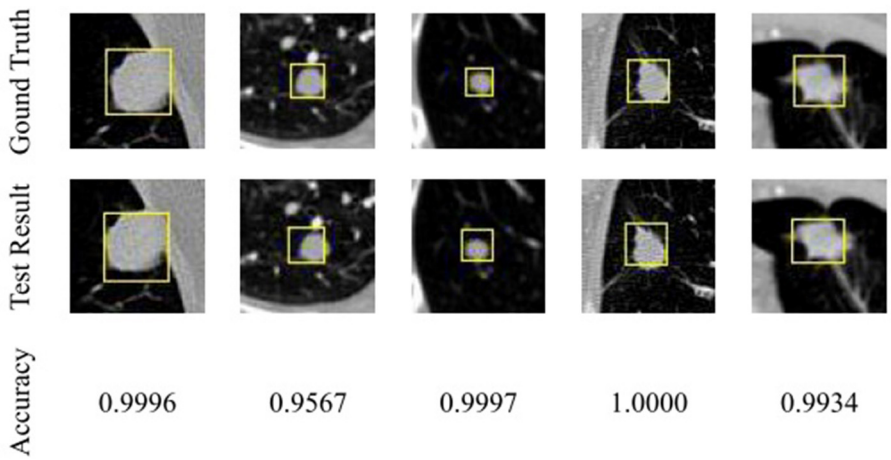


FIGURE 2 | Comparison of pulmonary nodules test results and actual locations.

by radiologists, the second row is detection network test results, and the third row is the nodules detection accuracy.

As you can see from **Figure 2**, the detected nodule is approximately the same as the actual location, and the size of the detection frame is approximately the same as the actual size. Detection accuracy has reached a high value, which can prove that based on the detection of the network of the nodules, the test set has a good detection effect.

Discussion

According to the result of the tests on the LUNA16 datasets, the pulmonary nodule classification model performance is satisfactory and stable when the proportions of samples are relatively balanced. It means that CT images with positive pulmonary nodules can be efficiently and accurately classified through the classification network. Therefore, a simple filtering process can remove a lot of unrelated images so that the Faster R-CNN model could detect the pulmonary nodule in a small quantity of CT images that may contain pulmonary nodules without examining all CT images. So, it is a feasible method and is expected to help solve the problems of low detection rates and high false-positives in pulmonary nodules detection.

In this way, the suspected pulmonary nodules in the CT image of the case would quickly be selected through the classification network, and the location range of the suspected nodules in the CT image can be determined, and then detected by the detector, so as to improve the efficiency of the detector, increase the reliability of the detection, and reduce the false positive of the detector. It means that it could effectively detect the positive pulmonary nodules in the total CT image to assist doctors in the early screening of pulmonary nodules and has certain research value.

CONCLUSIONS

In this paper, deep learning and related algorithms are used to study and discuss lung image classification, a new method based on the deep convolutional neural network to detect pulmonary nodules in CT images is proposed. This method

can be regarded as a different way to reduce the false-positive rate of detection by Screening targets before detection. The main contribution of this method greatly reduces the detection time and efficiency and reduces the false-positive rate. The following conclusions can be drawn from the experimental results:

1) Under the training set of positive and negative sample balance, the classification network can accurately classify lung CT images with pulmonary nodules or not.
2) A simple filtering process can remove a large number of unrelated images so that the Faster R-CNN model could detect the pulmonary nodule in a small quantity of CT images without examining all CT images. Therefore, the proposed method can reduce the time spent on testing and improve efficiency.
3) Because the number of samples of pulmonary nodules is insufficient, the generalization ability of the model is not strong.

In the next work, we intend to improve the generalization ability. Ideally speaking, if we can introduce more positive samples to the training dataset, the predictive performance can be more accurate. Moreover, we hope to integrate the information about benign or malignant pulmonary nodules into the deep network so that it can automatically complete testing and evaluation.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://luna16.grand-challenge.org.

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

JL proposed the research question, conducted the analysis, and edited the manuscript. GY and JG did the investigation, validated the results, discussed, and gave advice. QH accomplished the dataset preparation, model training, and wrote the original draft. SZ contributed to the associated analysis and design of experiments, reviewed, and edited the manuscript. All authors contributed to the article and determined the submitted version together.

FUNDING

This work was supported by the Dongguan Social Science and Technology Development Project (Grant No. 2019507140209) and the Department of Education of Guangdong in China under Grants 2020ZDZX3029, 2017KZDXM082, and 2020ZDZX2016.

ACKNOWLEDGMENTS

Thanks to all the authors involved in this study for their valuable suggestions, as well as thanks to the computational resources provided by the Dongguan University of Technology.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, et al. Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin*. (2018) 68:394–424. doi: 10.3322/caac.21492
2. Feng RM, Zong YN, Cao SM, Xu RH. Current cancer situation in china: good or bad news from the2018 global cancer statistics? *Cancer Commun.* (2019) 39:6. doi: 10.1186/s40880-019-0368-6
3. Center MM. Global cancer statistics. *CA Cancer J Clin.* (2015) 65:87–108. doi: 10.3322/caac.21262
4. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med.* (2011) 365:395–409. doi: 10.1056/NEJMoa1102873
5. Niemeijer M, Loog M, Abramoff MD, Viergever MA, Prokop M, van Ginneken B. On combining computer-aided detection systems. *IEEE Trans Med Imaging.* (2011) 30:215–23. doi: 10.1109/TMI.2010.2072789
6. Arindra Adiyoso Setio A, Traverso A, de Bel T, Berens MSN, van den Bogaard C, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Med Image Anal.* (2016) 42:1–13. doi: 10.1016/j.media.2017.06.015
7. Zeng N, Wang Z, Zineddin B, Li Y, Du M, Xiao L, et al. Image-based quantitative analysis of gold immunochromatographic strip *via* cellular neural network approach. *IEEE Trans Med Imag.* (2014) 33:1129–36. doi: 10.1109/T.M.I.2014.2305394
8. Torres EL, Fiorina E, Pennazio F, Peroni C, Saletta M, Camarlinghi N, et al. Large scale validation of the m5l lung cad on heterogeneous ct datasets. *Med Phys.* (2015) 42:7370. doi: 10.1118/1.4907970
9. Zeng N, Wang Z, Zhang H. Inferring nonlinear lateral flow immunoassay state-space models *via* an unscented Kalman filter. *Sci China Inform Sci.* (2016) 59:112204. doi: 10.1007/s11432-016-0280-9
10. Murphy K, van Ginneken B, Schilham AM, de Hoop BJ, Gietema HA, Prokop M. A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbor classification. *Med Image Anal.* (2009) 13:757–70. doi: 10.1016/j.media.2009.07.001
11. Messay T, Hardie RC, Rogers SK. A new computationally efficient cad system for pulmonary nodule detection in ct imagery. *Med Image Anal.* (2010) 14:390–406. doi: 10.1016/j.media.2010.02.004
12. Zeng N, Song D, Li H, You Y, Alsaadi FE. A competitive mechanism integrated multi-objective whale optimization algorithm with differential evolution. *Neurocomputing.* (2020) 432:65. doi: 10.1016/j.neucom.2020.12.065
13. Lee J-G, Jun S, Cho Y-W, Lee H, Kim GB, Seo JB, et al. Deep learning in medical imaging: general overview. *Kor J Radiol.* (2017) 18:570–84. doi: 10.3348/kjr.2017.18.4.570
14. Zeng N, Wang Z, Zhang H, Kim K-E, Li Y, Liu X. An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips. *IEEE Trans Nanotechnol.* (2019) 18:819–29. doi: 10.1109/T.N.A.N.O.2019.2932271

15. Golan R, Jacob C, Denzinger J. Lung nodule detection in CT images using deep convolutional neural networks. In: *International Joint Conference on Neural Networks*. Vancouver, BC: IEEE (2016).
16. Tourassi GD, Armato SG, Anirudh R, Thiagarajan JJ, Bremer T, Kim H. Lung nodule detection using 3D convolutional neural networks trained on weakly labeled data. *Spie Med Imaging*. (2016) 9785:978532. doi: 10.1117/12.2214876
17. Huang X, Shan J, Vaidya V. Lung nodule detection in CT using 3D convolutional neural networks. In: *The 2017 IEEE International Symposium on Biomedical Imaging*. Melbourne, VIC: IEEE (2017).
18. Fu L, Ma J, Ren Y, Han Y, Zhao J. Automatic detection of lung nodules: false positive reduction using convolution neural networks and handcrafted features. In: *Medical Imaging 2017: Computer-Aided Diagnosis*, Vol. 10134 (2017). p. 101340A. doi: 10.1117/12.2253995
19. Li C, Zhu G, Wu X, Wang Y. False-positive reduction on lung nodules detection in chest radiographs by ensemble of convolutional neural networks. *IEEE Access*. (2018) 6:16060–7. doi: 10.1109/ACCESS.2018.2817023
20. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: cnn architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imag*. (2016) 35:1285–98. doi: 10.1109/TMI.2016.2528162
21. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *Int J Comp Vision*. (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y
22. Krizhevsky A, Sutskever I, Hinton, G. ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, Vol. 25. Curran Associates Inc. (2012).
23. Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Machine Intell*. (2017) 39:1137–49. doi: 10.1109/TPAMI.2016.2577031
24. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV: IEEE (2016).
25. Armato SG, Roberts RY, McNitt-Gray MF, Meyer CR, Reeves AP, McLennan G, et al. The lung image database consortium (LIDC). *Acad Radiol*. (2007) 14:1455–63. doi: 10.1016/j.acra.2007.08.006
26. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P,et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *J Digital Imaging*. (2013) 26:1045–57. doi: 10.1007/s10278-013-9622-7
27. Tourassi GD, Armato SG, Bergtholdt M, Wiemker R, Klinder T. Pulmonary nodule detection using a cascaded SVM classifier. *Spie Med Imaging*. (2016) 9785:978513. doi: 10.1117/12.2216747
28. Sóñora-Mengana A, Papavasileiou E, Naranjo JC, Jansen B, Vandemeulebroucke J. Evaluation of data balancing techniques. Application to CAD of lung nodules using the LUNA16 framework. *RIELAC*. (2018) 39:57–67.

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 Public Health

ORIGINAL RESEARCH
published: 21 May 2021
doi: 10.3389/fpubh.2021.685596

Gaussian Process Autoregression for Joint Angle Prediction Based on sEMG Signals

Jie Liang[1], *Zhengyi Shi*[2,3], *Feifei Zhu*[2,3], *Wenxin Chen*[2,3], *Xin Chen*[1]* *and Yurong Li*[2,3]*

[1] Department of Rehabilitation, Fuzhou Second Hospital Affiliated to Xiamen University, Fuzhou, China, [2] College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China, [3] Fujian Key Laboratory of Medical Instrumentation and Pharmaceutical Technology, Fuzhou, China

There is uncertainty in the neuromusculoskeletal system, and deterministic models cannot describe this significant presence of uncertainty, affecting the accuracy of model predictions. In this paper, a knee joint angle prediction model based on surface electromyography (sEMG) signals is proposed. To address the instability of EMG signals and the uncertainty of the neuromusculoskeletal system, a non-parametric probabilistic model is developed using a Gaussian process model combined with the physiological properties of muscle activation. Since the neuromusculoskeletal system is a dynamic system, the Gaussian process model is further combined with a non-linear autoregressive with eXogenous inputs (NARX) model to create a Gaussian process autoregression model. In this paper, the normalized root mean square error (NRMSE) and the correlation coefficient (CC) are compared between the joint angle prediction results of the Gaussian process autoregressive model prediction and the actual joint angle under three test scenarios: speed-dependent, multi-speed and speed-independent. The mean of NRMSE and the mean of CC for all test scenarios in the healthy subjects dataset and the hemiplegic patients dataset outperform the results of the Gaussian process model, with significant differences ($p < 0.05$ and $p < 0.05$, $p < 0.05$ and $p < 0.05$). From the perspective of uncertainty, a non-parametric probabilistic model for joint angle prediction is established by using Gaussian process autoregressive model to achieve accurate prediction of human movement.

Keywords: sEMG, Gaussian process, joint angle prediction, NARX, neurorehabilitation

OPEN ACCESS

Edited by:
Kathy Clawson, University of Sunderland, United Kingdom

Reviewed by:
Yanzheng Zhu, Huaqiao University, China
Xunyuan Yin, University of Alberta, Canada

***Correspondence:**
Xin Chen
cxin2670@126.com
Yurong Li
liyurong@fzu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *25 March 2021*
Accepted: *20 April 2021*
Published: *21 May 2021*

Citation:
Liang J, Shi Z, Zhu F, Chen W, Chen X and Li Y (2021) Gaussian Process Autoregression for Joint Angle Prediction Based on sEMG Signals. Front. Public Health 9:685596. doi: 10.3389/fpubh.2021.685596

1. INTRODUCTION

Rehabilitation robots and other rehabilitation equipment have developed rapidly and are widely used for therapeutic training of patients suffering from neurological disorders including stroke, cerebral palsy and spinal cord injury. Patients with impaired neurological function are able to use rehabilitation equipment for a variety of exercises to restore strength and flexibility in their extremities (1). Neurorehabilitation techniques can help both passive and active training of patients with neurological injuries. Compared to passive training through rehabilitation equipment, training that involves the patient's own will can improve the effectiveness of treatment and restore motor function through active movement (2). Electromyography (EMG) represents the sum of subcutaneous motor action potentials generated through muscle contraction (3), which can represent neuromuscular activity and is a way to reflect the patient's voluntary effort.

Surface electromyography (sEMG) is a non-invasive EMG signal that is mainly acquired through non-invasive electrodes, which has the advantage of being more accessible and less likely to impede the normal activity of the user and will override the actual joint motion (4), and is often used as a control command to realize the human-robot interface (HRI) of portable or wearable assisted rehabilitation equipment. The surface electromyography (sEMG) is a non-invasive EMG signal collected by surface electrodes, which has the advantage of being more accessible than electroencephalography (EEG) and less likely to impede the user's normal activities. The forces generated by muscles in response to neural control signals depend on a large number of variables distributed over many spatiotemporal scales (5), which makes it difficult to predict the muscle force. While the nervous system clearly has knowledge of some of these variables, such as muscle length and velocity, other aspects of muscle dynamics (e.g., the detailed dynamics of molecules at the level of individual myofibrils and sarcomeres) are much more difficult to measure and estimate. With knowledge of the nervous system, it is possible to model the relationship between neural control signals and muscle force and use it to predict or simulate muscle force production. However, this relationship itself cannot be accurately estimated, and the remaining unaccounted aspects of muscle dynamics will result in seemingly random fluctuations in force, also known as motor noise. Therefore, there are two main sources of uncertainty in the neuromusculoskeletal system: irreducible noise during the motor system and variability in the relationship between neural control signals and muscle outputs (6). The impact of uncertainty in neuromusculoskeletal models on joint motion prediction can be mitigated, but not completely eliminated, by different modeling approaches. The human neuromuscular system is a highly non-linear and time-varying system with uncertainty (7). In order to allow models to handle the dynamic high-dimensional nature of the neuromuscular system, it is not enough to rely on traditional model structures and faster computational processing. Therefore, a central issue for further research on neuromusculoskeletal systems, or any artificial controller, is how to command muscles effectively in the presence of uncertainty.

Non-linear system modeling and identification can be divided into parametric and non-parametric models from the perspective of model structure (8–12). From the perspective of the Bayesian statistical framework, probability distributions over the functional space can be considered and modeled by optimizing these distributions to characterize uncertainty. This type of model has no explicit modeling mechanism or constraints and is referred as non-parametric modeling (13). The non-parametric model does not depend on a specified set of parameters or a fixed model structure, and is an estimation method based on statistical principles, which can generate functions to fit the data without the constraints of the model mechanism. The number and nature of the parameters of the non-parametric model are flexible and variable, which can change accordingly with the change of the data set. It can be used for modeling and analyzing high-dimensional time-varying systems, taking the uncertainty into account during the modeling process, and characterizing the uncertainty. The non-parametric methods mainly include spectral estimation, spectral analysis, correlation analysis and kernel-based analysis methods. Gaussian process (GP) model is a non-parametric kernel method in the framework of Bayesian model, which is simple to implement, computationally efficient, and most importantly, GP model can describe the posterior distribution of the model function, which in turn can be used to describe the uncertainty of the model, and is a recent research hotspot for non-parametric methods (14, 15). Kang et al. proposed an effective method for generating suboptimal motion of a multi-body system using a GP dynamics model to achieve dimensionality reduction of the system and deal with motion optimization problems (16). Schearer and Ullauri achieved the estimation of joint moments by building semi-parametric and non-parametric models through GP model, respectively (17, 18). Xiloyannis et al. used a Gaussian autoregressive model for decoding neural information to achieve multidimensional decoding control of 11 joint movements (19). Yang et al. proposed a proportional pattern recognition control of arm muscles using a wearable ultrasound sensor to achieve both gesture recognition and muscle contraction force estimation based on statistical features and Gaussian process regression models (20).

It has been shown that joint motion can reflect the intrinsic dynamics of human movement (21), so motion signals can also be used in the modeling of the neuromusculoskeletal system for building autoregressive models for prediction. The non-linear autoregressive with eXogenous inputs (NARX) model is an effective method for solving non-linear sequential problems, and modeling in conjunction with the NARX model can better incorporate the non-linear spatiotemporal correlation structure of muscle-driven control signals and natural human motion (22). Dynamic recurrent neural networks based on the NARX model are widely used for joint angle estimation, decoding shoulder, elbow and wrist motions and prosthesis model control (23–25). Gupta proposed an ankle joint angle estimation model based on the NARX model using sEMG signals and knee joint angle signals, and the performance of the model proved its applicability to ankle joint angle estimation for active prosthesis, orthosis and lower limb rehabilitation (26). Liu et al. used the NARX model to train and identify the EMG signals motion mapping relationship between a rehabilitation training bed and sEMG signals based motion prediction to achieve the identification of upper body tilt in different directions (27). Raj et al. proposed a multilayer perceptron neural network model based on the NARX model for estimating elbow joint angle and elbow joint angular velocity, and the proposed model estimated elbow joint angle and elbow joint angular velocity with high accuracy (28).

Since the neuromusculoskeletal system is a dynamic time-varying system, the GP model is only a mapping of the input to the output distribution, which is a basic static system. By combining the NARX model into the modeling, the resulting model can not only adapt to non-linear discrete-time processes, but also to different noise models, and the resulting dynamic model, which will better fit the physiological properties of the neuromusculoskeletal system. Thus, in this paper, a knee joint angle prediction model based on sEMG signals is proposed by considering the physiological properties of microscopic

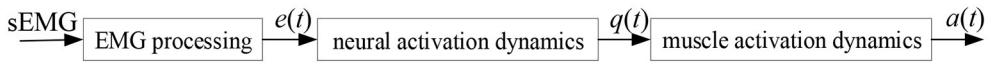


FIGURE 1 | Muscle activation dynamics.

muscle activation and combining NARX with GP model. First, considering the physiological properties of sEMG, the muscle activation kinetic model is used to extract features from the sEMG signals of a pair of antagonist muscles controlling knee joint motion, and the muscle activation intensity of this pair is obtained. Then a joint angle prediction model based on GP model is proposed for the uncertainty of the neuromusculoskeletal system with the muscle activation intensity as the input signal. Since the neuromusculoskeletal system is a time-varying nonlinear system, the one-step ahead (OSA) prediction of the NARX model is introduced to construct a Gaussian process autoregressive model, which uses the confidence interval of the prediction to describe the uncertainty, reduces the influence of model uncertainty on the prediction results, and improves the prediction rationality, accuracy, and efficiency of the joint angle prediction model.

2. MATERIALS AND METHODS

2.1. Muscle Activation Dynamics

The sEMG signal is the sum of action potentials recruited to the muscle by surface electrodes and is used to reflect the activation level of the muscle. sEMG signals can be considered as a form of characterization of neuromotor control commands and are widely used to analyze musculoskeletal models. The feature extraction of the EMG signals by the existing studied models only considers the macroscopic characteristics of the EMG signals, without considering the microscopic physiological characteristics of muscle activation. In order to characterize the time-varying features of the sEMG signals, respond to micro-physiological properties, and reflect the relationship between EMG signals, neural activation and muscle activation, a muscle activation kinetic model was established to achieve feature extraction of the sEMG signals (29–31). The muscle activation kinetics is mainly expressed as the transformation process between EMG signals and muscle activation, as shown in **Figure 1**, where $a(t)$ is the muscle activation, $e(t)$ is the processed sEMG signal, $q(t)$ the neural activation, and detailed in Li et al. (32).

2.2. Gaussian Process

Gaussian process (GP) is defined as a random process consisting of infinite high-dimensional random variables in a high-dimensional space, in which the joint distribution among any finite number of random variables is a Gaussian distribution. GP model can be derived from the weight-space view or the function-space view. Since each set of weights implies a specific function and the distribution of the weights implies the distribution of the function, the distribution of the GP can be obtained from the function-space view to obtain the equivalent of the weight-space view (33), which is the more commonly used derivation method for Gaussian process models.

Suppose the sample set $\mathcal{D}$ has N samples:

$$\mathcal{D} = (\mathbf{X}, \mathbf{Y}) = \left\{ (\mathbf{x}_i, \mathsf{y}_i) | \mathbf{x}_i \in \mathbb{R}^d, \mathsf{y}_i \in \mathbb{R}, i = 1, \ldots, N \right\} \quad (1)$$

where $\mathbf{x}_i$ denotes input vector,y_i denotes output vector.

A Gaussian process is completely specified by its mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$ (34):

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (2)$$

$$m(\mathbf{x}) = \mathbb{E}\left[f(\mathbf{x})\right] \quad (3)$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}\left[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}))\right] \quad (4)$$

where the random variable function $f(\mathbf{x})$ represents the distribution of y_i at $\mathbf{x}_i$, the mean function $m(\mathbf{x})$ reflects the expected function value at input $\mathbf{x}$. The covariance function models the dependence between the function values at different input points $\mathbf{x}$ and $\mathbf{x}'$, which is often referred as the kernel function of a GP model.

In the Gaussian process regression, considering the following model:

$$\mathbf{Y} = f(\mathbf{X}) + \boldsymbol{\varepsilon} \quad (5)$$

where $\mathbf{X}$ denotes the input vector,$\mathbf{Y}$ denotes the observed vector with noise, the noise follows a Gaussian distribution $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2)$, the random variable function $f(\mathbf{X})$ follows a Gaussian distribution:

$$f(\mathbf{X}) \sim \mathcal{N}(\mu(\mathbf{X}), \mathbf{K}(\mathbf{X}, \mathbf{X})) \quad (6)$$

Thus:

$$\mathbf{Y} \sim \mathcal{N}(\mu(\mathbf{x}), \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 I) \quad (7)$$

For the prediction input $\mathbf{X}^* = (\mathbf{x}_1^*, \ldots, \mathbf{x}_N^*)^T$, the joint distribution of the predicted values $f(\mathbf{X}^*)$ and the training data output is:

$$\begin{pmatrix} \mathbf{Y} \\ f(\mathbf{X}^*) \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(\mathbf{X}) \\ \mu(\mathbf{X}^*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I} & \mathbf{K}(\mathbf{X}, \mathbf{X}^*) \\ \mathbf{K}(\mathbf{X}^*, \mathbf{X}) & \mathbf{K}(\mathbf{X}^*, \mathbf{X}^*) \end{bmatrix} \right) \quad (8)$$

where $\mathbf{K}(\mathbf{X}, \mathbf{X})$ denotes the covariance matrix of the input signal, which is a symmetric semi-positive definite matrix of $N \times N$ order:

$$\mathbf{K}(\mathbf{X}, \mathbf{X}) = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{pmatrix} \quad (9)$$

Knowing the joint high-dimensional distribution, the posterior distribution is obtained by finding the conditional probability $p(f(\mathbf{X}^*)|\mathbf{Y},\mathbf{X},\mathbf{X}^*)$ according to Bayes' theorem:

$$f(\mathbf{X}^*)|\mathbf{X}^* \sim \mathcal{N}(\mu^*, \Sigma^*) \tag{10}$$

$$\mu^* = \mathbf{K}(\mathbf{X}^*,\mathbf{X})[\mathbf{k}(\mathbf{X},\mathbf{X}+\sigma^2\mathbf{I}]^{-1}(\mathbf{Y}-\mu(\mathbf{X})) + \mu(\mathbf{X}^*) \tag{11}$$

$$\Sigma^* = \mathbf{K}(\mathbf{X}^*,\mathbf{X}^*) - \mathbf{K}(\mathbf{X}^*,\mathbf{X})[\mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma^2\mathbf{I}]^{-1}\mathbf{K}(\mathbf{X},\mathbf{X}^*) \tag{12}$$

The covariance matrix accounts for the major part of the posterior distribution of the Gaussian process, and the covariance matrix is a key component of the Gaussian process prediction. The kernel function is the main structure of the covariance matrix, so it is also a central part of the Gaussian process model. In the modeling and identification of dynamic systems, the dimensions of the inputs are relatively high, which makes the description of the mapping function complex. Considering the smoothness and continuity of dynamic systems, the squared exponential (SE) kernel function is often used in the modeling process (35), which is defined as:

$$k(\mathbf{x},\mathbf{x}') = \sigma_f^2 \exp\left[-\frac{(\mathbf{x}-\mathbf{x}')^T(\mathbf{x}-\mathbf{x}')}{2\sigma_l^2}\right] \tag{13}$$

where hyperparameter σ_l is the characteristic length scale, which determines the relative weights of the distances of the input variables. σ_f is the signal standard deviation, which reflects the magnitude of the function change.

The Gaussian process model is mainly determined by the kernel function and its hyperparameters, and its learning process is a process of training through the data to obtain the posterior probability distribution, which mainly includes the selection (or design) of the kernel function and the determination of the hyperparameters.

The kernel functions of Gaussian processes often contain unknown and indefinite hyperparameters, such as length scales, signal and noise variances, etc. These need to be inferred from the data, resulting in posterior distributions of the hyperparameters that are not easily obtained. Therefore, the full Bayesian derivation of hyperparameters is not commonly used in practical applications. The usual practice is to obtain point estimates of the hyperparameters by maximizing the log marginal likelihood.

Given a sample set $\mathcal{D}$ and the hyperparameter of the Gaussian process is θ, the marginal likelihood is as shown in Equation (14):

$$p(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta}) = \int p(\mathbf{Y}|\mathbf{X},f,\boldsymbol{\theta})p(f|\mathbf{X},\boldsymbol{\theta})df \tag{14}$$

The marginal likelihood is mainly a marginalization of the function. In the Gaussian process model, the prior $f|\mathbf{X},\boldsymbol{\theta}$ of the model is a Gaussian distribution, i.e., $p(f|\mathbf{X},\boldsymbol{\theta})=\mathcal{N}(0,\mathbf{K}_\theta)$. When the observed likelihood function $p(\mathbf{Y}|\mathbf{X},f,\boldsymbol{\theta})$ of the sample set is also Gaussian distributed, i.e., $p(\mathbf{Y}|\mathbf{X},f,\boldsymbol{\theta})=\mathcal{N}(f,\sigma^2\mathbf{I})$, then $p(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta})$ is also Gaussian distributed:

$$p(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta}) = \int \mathcal{N}(0,\mathbf{K}_\mathbf{Y})\mathcal{N}(f,\sigma^2\mathbf{I})df = \mathcal{N}(0,\mathbf{K}_\mathbf{Y}+\sigma^2\mathbf{I}) \tag{15}$$

According to Equation (15), the log marginal likelihood is obtained as:

$$\log p(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta}) = -\frac{1}{2}\mathbf{Y}^T{\mathbf{K}_\mathbf{Y}}^{-1}\mathbf{Y} - \frac{1}{2}\log|\mathbf{K}_\mathbf{Y}| - \frac{N}{2}\log 2\pi \tag{16}$$

where $\mathbf{K}_\mathbf{Y} = \mathbf{K}(\mathbf{X},\mathbf{X})+\sigma^2\mathbf{I}$ is the output covariance matrix.

The maximum likelihood estimation combined with the conjugate gradient method is commonly used for the Gaussian process model to achieve the estimation of the hyperparameters of the model, and the computational complexity of this method is $\mathcal{O}(N^2)$ for each hyperparameter, and the computational complexity is small. The hyperparameter estimates of the Gaussian process model are obtained by maximizing the log marginal likelihood function through a gradient ascent based optimization tool:

$$\begin{aligned}\frac{\partial}{\partial\theta_i}\log p(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta}) &= -\frac{1}{2}\mathbf{Y}^T{\mathbf{K}_\mathbf{Y}}^{-1}\frac{\partial\mathbf{K}_\mathbf{Y}}{\partial\theta_i}{\mathbf{K}_\mathbf{Y}}^{-1}\mathbf{Y} - \frac{1}{2}tr({\mathbf{K}_\mathbf{Y}}^{-1}\frac{\partial\mathbf{K}_\mathbf{Y}}{\partial\theta_i}) \\ &= \frac{1}{2}tr\left((\boldsymbol{\alpha}\boldsymbol{\alpha}^T - {\mathbf{K}_\mathbf{Y}}^{-1})\frac{\partial\mathbf{K}_\mathbf{Y}}{\partial\theta_i}\right)\end{aligned} \tag{17}$$

where $\boldsymbol{\alpha} = \mathrm{K}_\mathrm{Y}^{-1}\mathrm{Y}$.

2.3. Non-parametric Model for Joint Angle Prediction Based on sEMG Signals

2.3.1. Joint Angle Prediction Based on Gaussian Process Model

The hamstrings and quadriceps are antagonistic muscles that together control the flexion and extension of the knee joint. The hamstrings are the muscles of the posterior thigh and consist mainly of the semitendinosus, semimembranosus and biceps femoris, while the quadriceps are the muscles of the anterior thigh and consist mainly of the vastus lateralis, vastus medialis, vastus intermedius and rectus femoris. In this paper, the sEMG signals of a pair of muscles in this antagonistic muscle group, the semimembranosus and the lateral femoris, were selected for the development of a non-parametric model for joint angle prediction. The physiological properties of muscle activation are combined with the GP model, and the squared exponential kernel is selected for subsequent modeling and analysis to establish a non-parametric model for joint angle prediction based on the GP model, as shown in **Figure 2**, where k denotes kth time step, $e_{1,k}$ and $e_{2,k}$ are the preprocessed sEMG signals of semimembranosus and lateral femoris muscles, respectively, which are then subjected to muscle activation dynamics to calculate the muscle activation $a_{1,k}$ and $a_{2,k}$. $\boldsymbol{u}_k$ denotes the input of the GP model, $\boldsymbol{u}_k = [a_{1,k}, a_{2,k}]^T$. The output of the Gaussian process model $\hat{y}_k$ is the predicted value of the joint angle.

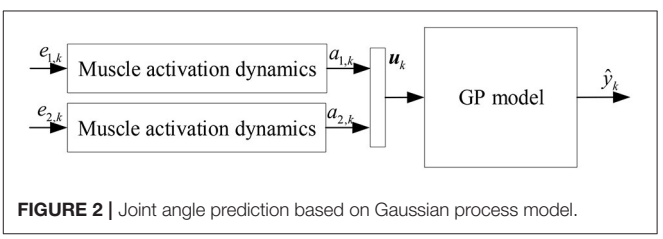


FIGURE 2 | Joint angle prediction based on Gaussian process model.

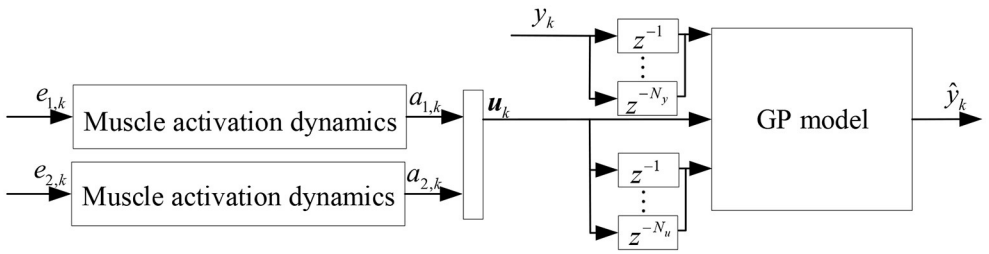


FIGURE 3 | Joint angle prediction based on Gaussian process autoregressive model.

2.3.2. Joint Angle Prediction Based on Gaussian Process Autoregressive Model

Since the neuromusculoskeletal system is a dynamic time-varying system, the joint angle prediction based on GP model only maps the input to the output distribution, which is a basic static system. Therefore, this paper further considers the time series of input and output signals to establish dynamic Gaussian process regression.

Non-linear autoregressive with eXogenous inputs (NARX) model is an effective method for solving non-linear sequence problems that accommodates non-linear discrete time processes and noisy models (36, 37). NARX is a dynamic recurrent network that predicts the current value of the system output by using a non-linear function f with previous inputs and outputs. The NARX model based on the actual measured values of the output is called One-step ahead (OSA) prediction:

$$y_k^* = f(u_k, u_{k-1} \ldots, u_{k-n_u}, y_{k-1}, y_{k-2}, \ldots, y_{k-n_y}) \tag{18}$$

where $f(\cdot)$ is the non-linear function between the input u_k and the estimated value y_k^* and y_k denotes measured value of model output, k represents kth time step, n_u and n_y are the maximum lags for model input and output, respectively.

The joint angle signal can be easily collected by inertial measurement unit (IMU), etc., and the joint angle prediction system can be established by EMG signals to achieve further advance prediction of joint angle. Since the high accuracy of joint angle prediction is required in practical applications, this paper improves the joint angle prediction method based on GP model by using the NARX model and muscle activation dynamics, which establishes a Gaussian dynamic model with NARX structure, i.e., Gaussian process autoregressive model, for joint angle OSA prediction, as shown in **Figure 3**, where y_k denotes measured value of joint angle, $\hat{y}_k$ denotes the joint angle prediction of Gaussian process autoregressive model, N_y and N_y are the maximum lags for model input and output, respectively.

3. EXPERIMENTS AND RESULTS

3.1. Datasets

In this paper, publicly available datasets (dataset 1 and dataset 2) were cited to validate and analyze the proposed knee joint angle prediction model.

3.1.1. Healthy Subjects Dataset

Dataset 1 contains gait data from 10 healthy subjects in running condition, and Dataset 1 is published in https://simtk.org/projects/nmbl_running. The dataset measured knee moment signals, EMG signals and motion data of 10 healthy subjects running on a treadmill at four speeds (2.0, 3.0, 4.0, and 5.0 m/s). Subjects were all male (age: 29 ± 5 years; height: 1.77 ± 0.04 m; weight: 70.9 ± 7.0 kg), all provided informed consent, and each subject was experienced in long-distance running, at least 50 km per week. Fifty-four reflective markers were placed on each subject, and the trajectory of the markers was recorded using eight Vicon MX40+ cameras with a data acquisition frequency of 100 Hz. Ground reaction forces and moments were acquired using a Bertec Corporation treadmill with a sampling frequency of 1,000 Hz. Motion data and ground reaction forces were preprocessed with 4th order zero-phase hysteresis Butterworth low-pass filtering (cutoff frequency 15 Hz) and critical damping low-pass filtering (cutoff frequency 15 Hz), respectively. The Delsys Bangoli system was used to collect EMG signals. A total of 11 muscles including the gluteus maximus, biceps femoris long head, medial femoris, lateral femoris, semimembranosus, and tibialis anterior muscles were collected. Hamner et al. gave a complete description of the dataset (38).

3.1.2. Hemiparetic Subject Dataset

Joint angle prediction is mainly applied to the development of rehabilitation equipment and rehabilitation training, so a gait dataset containing a male patient with high-functioning right hemiparesis, which is publicly available at https://simtk.org/projects/emgdrivenmodel, was also selected for further testing and analysis of the knee angle prediction model proposed in this paper. The subject was 79 years old, height 1.7 m, mass 80.5 kg, with a LE Fugl-Meyer motor assessment score of 32/34 and right-sided hemiparesis. All experimental procedures were approved by the University of Florida Health Sciences Center Institutional Review Board (IRB-01), and the subject signed a written informed consent prior to participation in the experiment. The dataset collected gait data from subjects walking on a split-belt instrumented treadmill (Bertec Corp., Columbus, OH) at five different speeds (0.4, 0.5, 0.6, 0.7, and 0.8 m/s), with over 50 gait cycles collected for each speed. Motion capture in the experiments was mainly performed by an optical motion capture system (Vicon Corp., Oxford, UK) and ground reaction force detection was measured using the treadmill with

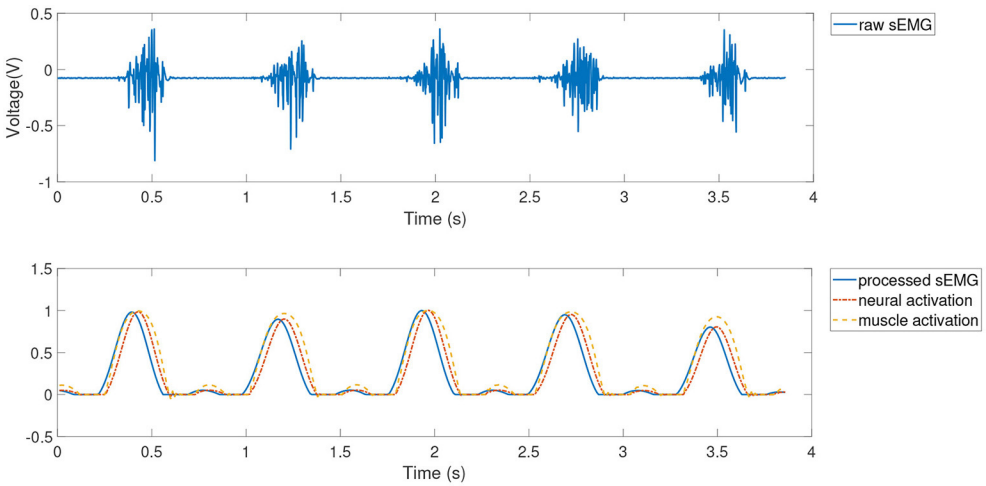


FIGURE 4 | Feature extraction of the lateral femoral.

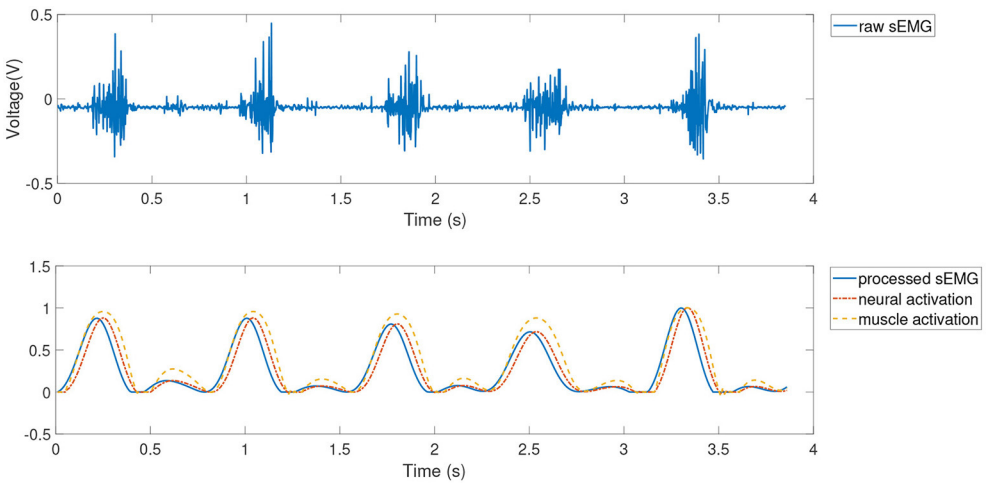


FIGURE 5 | Feature extraction of semimembranosus.

sampling frequency of 100 and 1,000 Hz, respectively. The EMG signal acquisition was performed using Motion Lab Systems with a sampling frequency of 1,000 Hz. Ground response and marker motion data was filtered using a fourth-order zero-phase lag Butterworth filter with a cutoff frequency of 7 divided by the gait period. EMG signal data were collected for 16 muscle groups of the lower extremity, including the anterior tibialis, semimembranosus, long head of the biceps femoris, medial femur, and lateral femur. A full description of this dataset is provided by Meyer et al. (39).

3.1.3. Pre-processing and Feature Extraction Results

The sEMG signals of dataset 1 and dataset 2 are preprocessed. Raw sEMG signals was firstly filtered using Butterworth zero phase shift bandpass filter (4th order, cutoff frequency 40 Hz) to eliminate low-frequency noise, then full-wave rectified and low-pass filtered (4th order, cutoff frequency = 3.5/step period), and finally sEMG data from each muscle were normalized to the maximum value over all trials to obtain the processed sEMG signals $e(t)$.

Feature extraction was then further performed using the muscle activation kinetic model in section 2.1. The results of pre-processing and feature extraction for dataset 1 (subject 1, 2 m/s) for the lateral femoral and semimembranosus muscles are shown in **Figures 4**, **5**.

3.2. Data Allocation Strategy

To study the effect of different speeds on joint angle prediction, joint angle prediction results were analyzed for two datasets of healthy subjects and hemiplegic subject in three conditions: speed-dependent, multi-speed and speed-independent. The normalized root mean square error (NRMSE) and correlation

coefficient (CC) were used as evaluation indicators of the prediction performance of the joint angle, as shown in Equations (19) and (20).

$$NRMSE = \frac{1}{y_{real}^{\max}} \sqrt{\frac{1}{num} \sum_{i=1}^{num} (y_{est} - y_{real})^2} \tag{19}$$

$$CC = \frac{cov(y_{est}, y_{real})}{\sigma_{y_{est}} \cdot \sigma_{y_{real}}} \tag{20}$$

where num denotes the number of samples tested, y_{est} denotes the predicted value of joint angle, y_{real} denotes the actual value of joint angle, and $y_{real}^{\max}$ denotes the maximum magnitude of the actual joint angle value.

One-way analysis of variance (ANOVA) was conducted to assess the statistical difference of estimation errors obtained by different models (40). The level of statistical significance was set to $p < 0.05$.

Dataset 1 contained 10 subjects, each speed containing five gait cycles, and three conditions were analyzed for each subject. speed-dependent took the first three cycles of each speed as a training set, and the last two cycles at the corresponding speed as test set; multi-speed took the first three cycles of each speed together as a training set, and the last two cycles of each speed separately as a test set; the speed-independent took the last two cycles of one speed in turn as test set, and the first three cycles of each unselected speed together as training set.

Dataset 2 contained the left and right leg gait data of a subject with right-sided hemiplegia, and the left and right leg gait data were analyzed for three conditions. speed-dependent tested the angle prediction results of each speed, using 10 gait cycles of a single speed as the training set and another

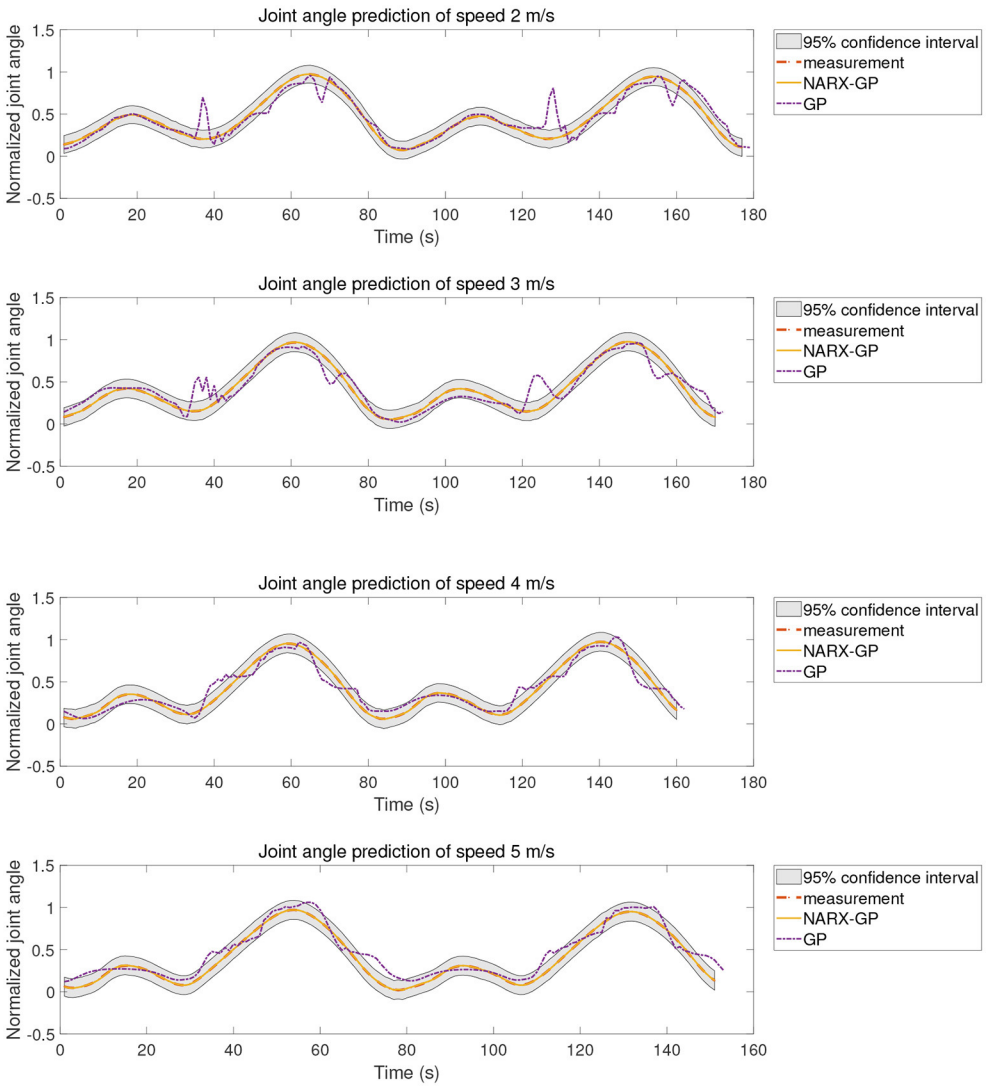


FIGURE 6 | Joint angle prediction under speed-dependent of subject 9.

10 cycles of the same speed as the test set; multi-speed used data from 5 m/s, each with 10 gait cycles, for a total of 50 gait cycles as the training set, and 10 gait cycles for each speed as the test set; speed-independent test took 10 gait cycles of one speed in turn as the test set, and 10 gait cycles of each of the other unselected speeds together as the training set.

Before performing the joint angle prediction based on Gaussian process model and Gaussian process autoregressive model, the Gaussian process model needs to be trained offline for different datasets according to the data allocation strategy. In this paper, the dataset was trained and tested on the MATLAB platform, and the Gaussian process regression model was trained offline using the "fitrgp" function. The method of estimating the model parameters was set to "exact," and the point estimates of the hyperparameters were obtained by maximizing the log marginal likelihood, and the kernel function was set to the squared exponential kernel function. The input signal for offline training of joint angle prediction based on Gaussian process model was the muscle activation of lateral femoral and semimembranosus muscles, and the output was the normalized joint angle signal. Given that muscle dynamics is a second-order model, the joint angle prediction based on the Gaussian process autoregressive model is set to second order, which lead to the number of the maximum lags for model input and output be 2, i.e., $n_u = n_y = 2$. Therefore, the output signal of offline training was the joint angle signal, and the input was the muscle activation. The model was trained offline and joint angle prediction was performed according to the data allocation strategy for different datasets under different data allocation strategies, respectively.

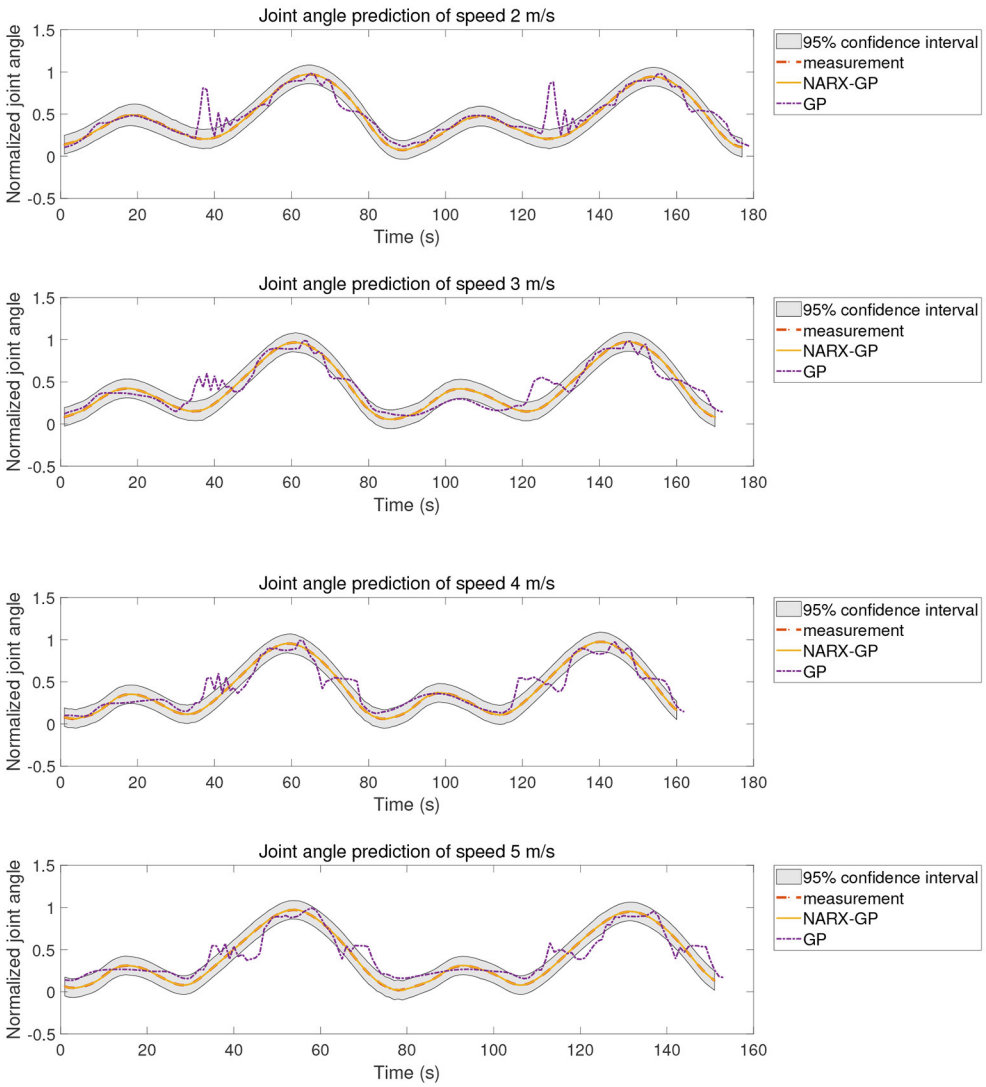


FIGURE 7 | Joint angle prediction under multi-speed of subject 9.

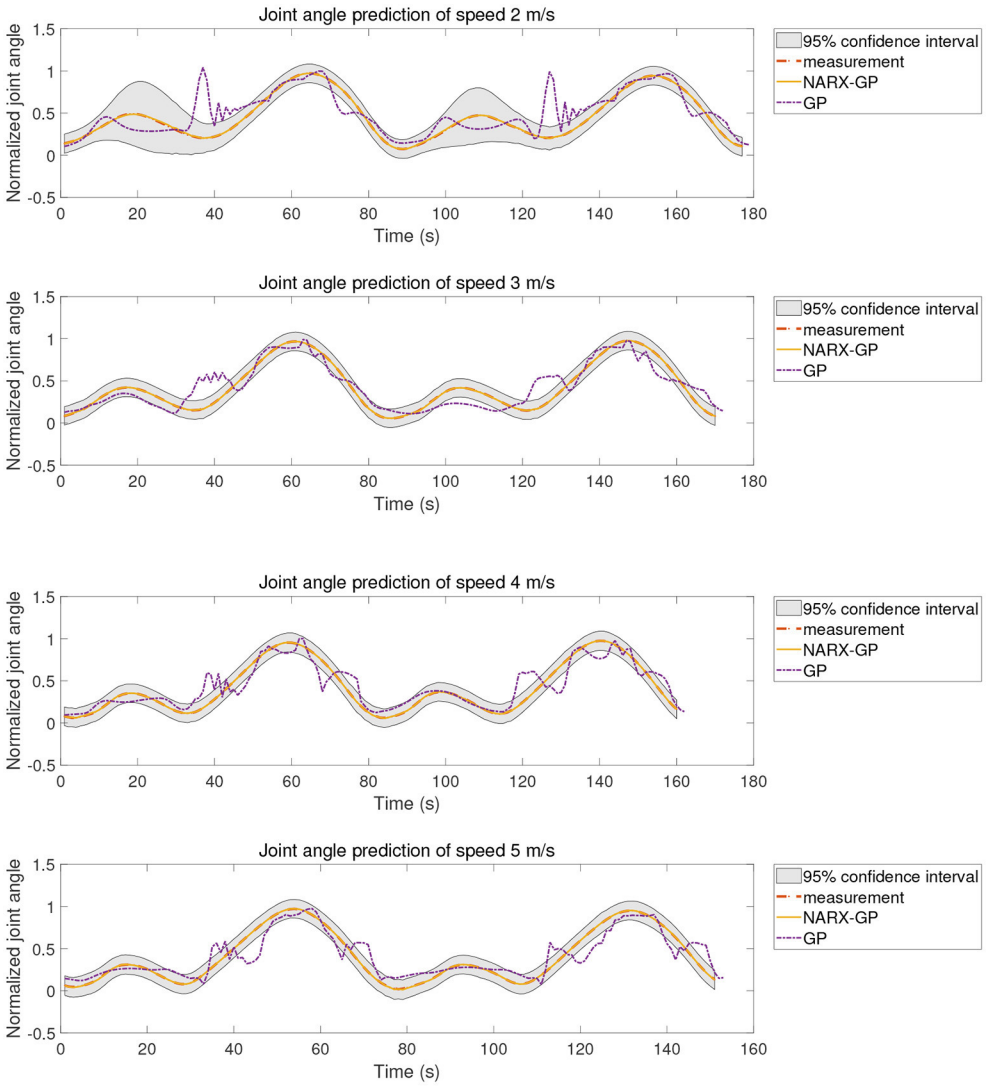


FIGURE 8 | Joint angle prediction under speed-independent of subject 9.

3.3. Estimation Results

3.3.1. Joint Angle Prediction Results for Dataset 1

The proposed method was tested using the data in dataset 1. Subjects were tested for knee angle prediction in three cases, speed-dependent, multi-speed, and speed-independent, according to the data allocation strategy, and the test results for subject 9 knee angle prediction were shown in **Figures 6–8**, where "NARX-GP" was the joint angle prediction based on the Gaussian process autoregressive model, "GP" was the joint angle prediction based on the Gaussian process model, and "measurement" was the actual measurement of joint angle. The gray shading was the 95% confidence interval ($\mu \pm 2\sigma$) for the prediction of the joint angle based on the Gaussian process autoregressive model to describe the uncertainty. From **Figures 6–8**, it can be concluded that the direct joint angle prediction by Gaussian process model cannot describe the relationship between sEMG signal and joint angle well, and the knee joint angle prediction results have a large error. Establishing a Gaussian process autoregressive model for OSA prediction of joint angle can significantly improve the prediction accuracy and can approximate the actual joint angle signal. OSA prediction incorporates the actual values of the previous moments of output into the model structure with high prediction accuracy, and is suitable for scenarios where the actual measurements of the output are easy to collect and where high prediction accuracy is required. The joint angle signal can be easily collected by inertial measurement unit (IMU), etc., and the joint angle prediction system can be established by EMG signals to achieve further advance prediction of joint angle.

Further error assessment and statistical analysis of the prediction results were performed, and the mean NRMSE and CC between the predicted and actual measurements for different velocity joint angles of the subjects are shown in **Figures 9**, **10**.

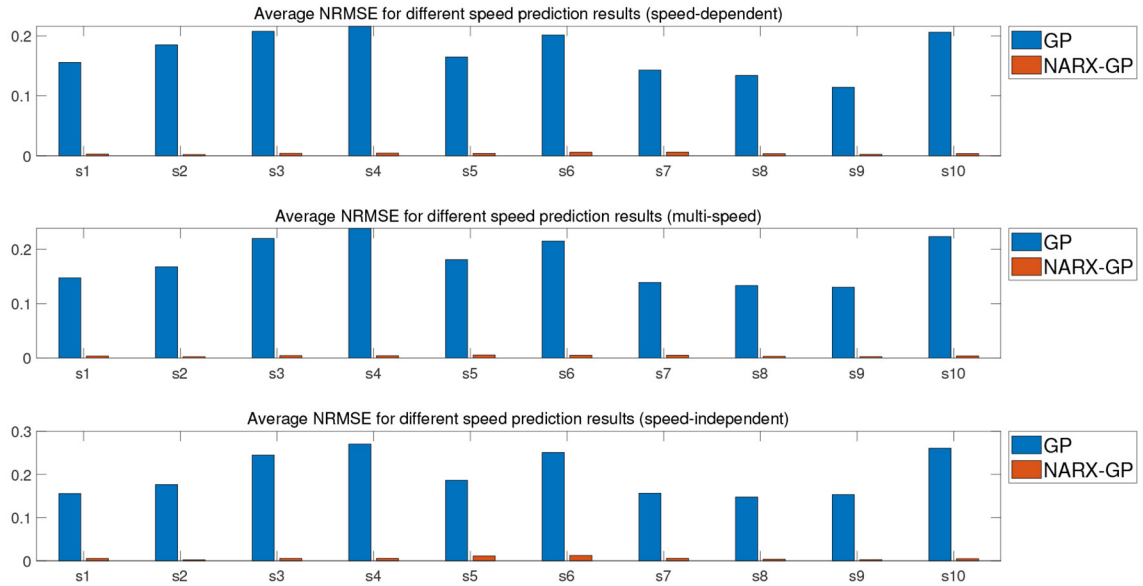


FIGURE 9 | Average NRMSE for prediction of joint angle under different speed.

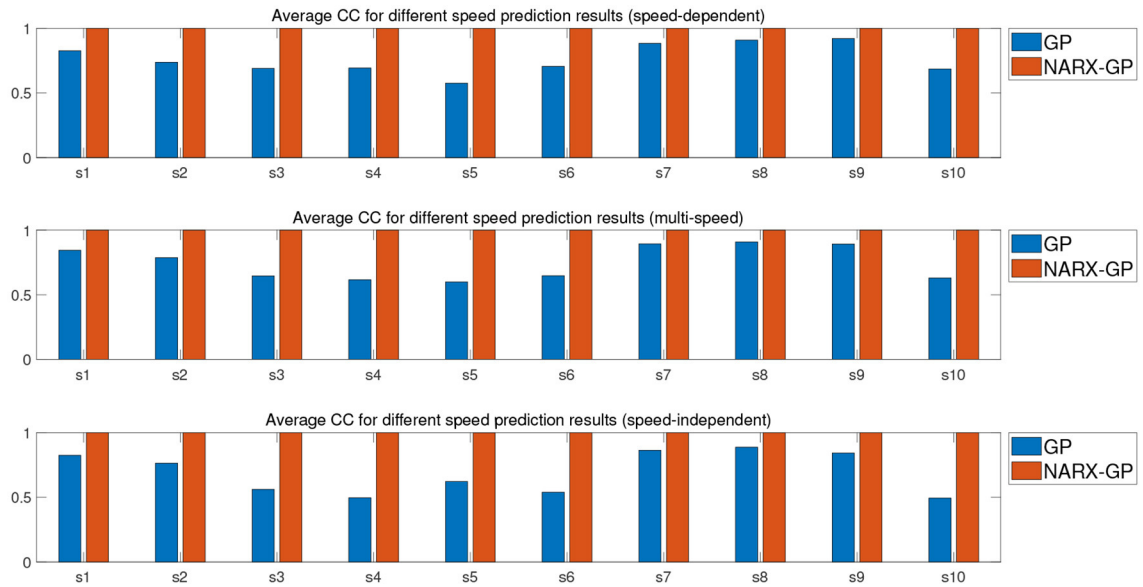


FIGURE 10 | Average CC for prediction of joint angle under different speed.

It can be seen from the figures that the prediction results of the NARX-GP model were significantly better than those of the GP model, the NRMSE between the prediction results of the NARX-GP model for knee joint angle and the actual knee joint angle was small and significantly smaller than that of the GP model, and the strong correlation between the prediction results of the NARX-GP model and the actual values of the joint angle with a higher correlation coefficient than that of the GP model.

The means and standard deviations of NRMSE and CC between predicted and actual values of joint angles for all subjects in speed-dependent, multi-speed and speed-independent conditions are shown in **Tables 1**, **2**. From **Tables 1**, **2**, it can be seen that the predictions of the NARX-GP model were highly correlated, and the NRMSE of the predictions was significantly lower than that of the GP model. The mean NRMSE was further calculated as 0.0039 ± 0.019, 0.0038 ± 0.0019, and 0.0059 ± 0.0061 for all subjects in the NARX-GP

model in the load-dependent, multi-load and load-independent conditions, 0.1728 ± 0.0543, 0.1795 ± 0.0563, and 0.2003 ± 0.0659 for the GP model. ANOVA was used to evaluate NRMSE of NARX-GP and GP model predictions, showing significant differences ($p < 0.05$, $p < 0.05$, and $p < 0.05$). The NARX-GP model prediction errors were smallest for all three scenarios, and slightly larger for load-independent. The joint angle prediction results of the GP model for all three scenarios were optimal for speed-dependent, with the smallest NRMSE, followed by multi-speed, and worst for speed-independent. The variability of joint motion at different speeds affected the prediction results of the model, and the experimental results also demonstrated that the speed-dependent results were optimal and the speed-independent results were the worst. The average NRMSE of NARX-GP model prediction results for all scenarios was 0.0045 ± 0.0040, which was better than the GP model results (0.1842 ± 0.0602), with a significant difference ($p < 0.05$).

TABLE 1 | NRMSE between the estimated joint torque of different models and the measurements ("true" values) of all subjects (mean ± std).

		2 m/s	3 m/s	4 m/s	5 m/s
Speed-dependent	GP	0.1788 ± 0.0756	0.1802 ± 0.0475	0.1604 ± 0.0483	0.1718 ± 0.0356
	NARX-GP	0.0046 ± 0.0016	0.0039 ± 0.0023	0.0031 ± 0.0023	0.0041 ± 0.0021
Multi-speed	GP	0.1850 ± 0.0735	0.1745 ± 0.0489	0.1712 ± 0.0457	0.1874 ± 0.0510
	NARX-GP	0.0052 ± 0.0028	0.0033 ± 0.0010	0.0030 ± 0.0011	0.0035 ± 0.0011
Speed-independent	GP	0.2075 ± 0.0697	0.1852 ± 0.0560	0.1910 ± 0.0563	0.2174 ± 0.0746
	NARX-GP	0.0086 ± 0.0086	0.0039 ± 0.0014	0.0046 ± 0.0030	0.0066 ± 0.0072

TABLE 2 | CC between the predicted joint angle of different models and the measurements ("true" values) of all subjects (mean ± std).

		2 m/s	3 m/s	4 m/s	5 m/s
Speed-dependent	GP	0.6276 ± 0.4151	0.7819 ± 0.1202	0.8241 ± 0.1120	0.8170 ± 0.1080
	NARX-GP	0.9999 ± 0.0001	0.9999 ± 0.0002	0.9999 ± 0.00004	0.9999 ± 0.0001
Multi-speed	GP	0.6381 ± 0.3708	0.7744 ± 0.1573	0.8026 ± 0.1141	0.7710 ± 0.1591
	NARX-GP	0.9998 ± 0.0002	0.9999 ± 0.00003	0.9999 ± 0.00003	0.9999 ± 0.0001
Speed-independent	GP	0.5970 ± 0.3333	0.7387 ± 0.2172	0.7493 ± 0.1540	0.6733 ± 0.2494
	NARX-GP	0.9991 ± 0.0020	0.9999 ± 0.00006	0.9998 ± 0.0002	0.9996 ± 0.0009

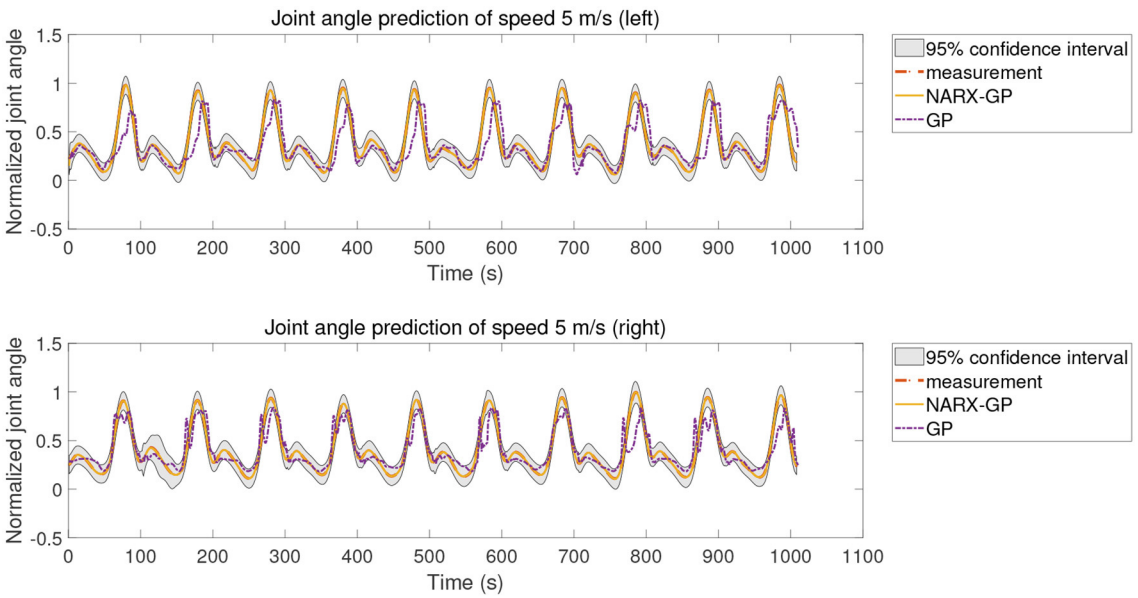


FIGURE 11 | Joint angle prediction of 5 m/s (speed-dependent).

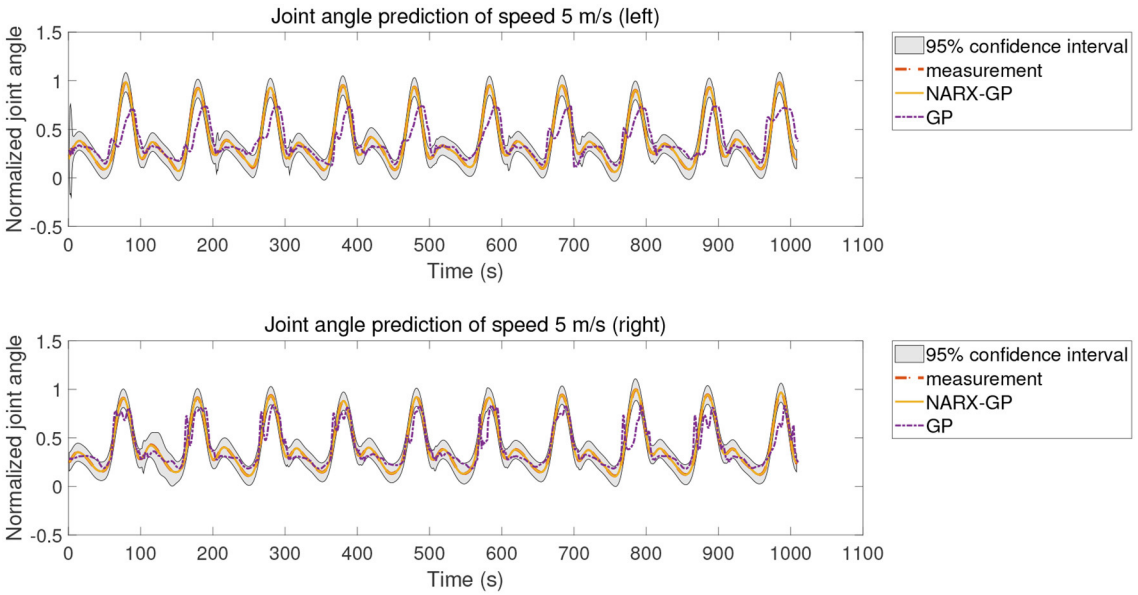


FIGURE 12 | Joint angle prediction of 5 m/s (multi-speed).

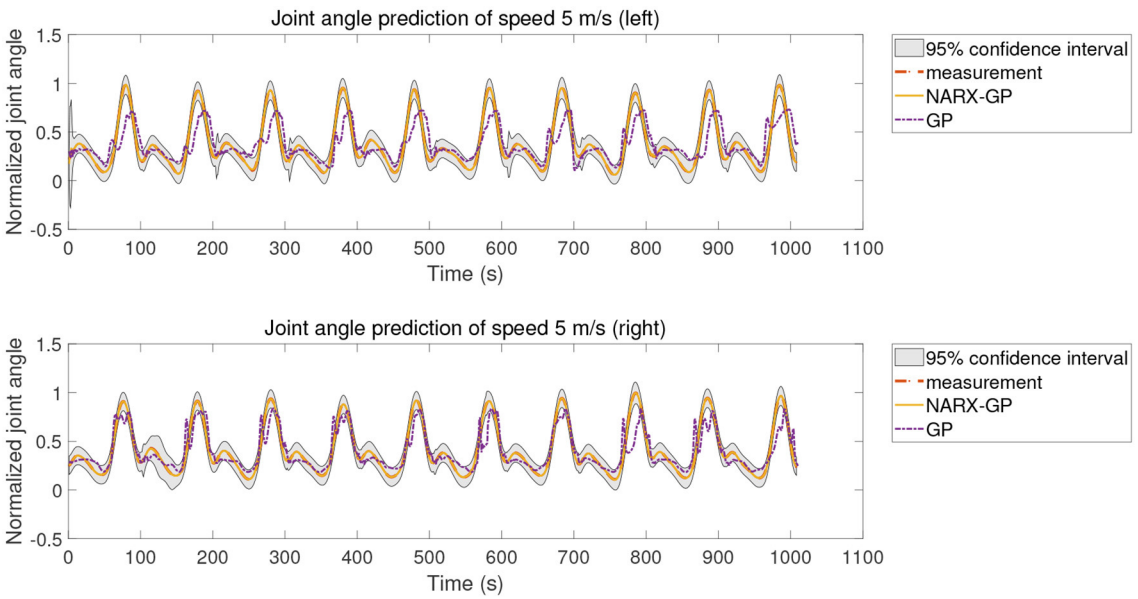


FIGURE 13 | Joint angle prediction of 5 m/s (speed-independent).

3.3.2. Joint Angle Prediction Results for Dataset 2

The proposed method was further tested using the data in dataset 2 to validate the accuracy of the proposed method for predicting the knee joint angle of patients. To reduce the influence of different muscle selections on the prediction results, the sEMG signals of the same pair of muscles, semimembranosus and lateral femoris, were selected for testing in dataset 2 as in dataset 1 and used to build a non-parametric model for joint angle prediction. Dataset 2 included data from the left and right legs of patients with different speeds, so the joint angle prediction results of subjects in three cases of speed-dependent, multi-speed and speed-independent were tested separately for the left and right legs according to the data allocation strategy, and the test results of joint angle prediction for speed of 5 m/s are shown in **Figures 11–13**, where "NARX-GP" was the joint angle prediction based on the Gaussian process autoregressive model, "GP" was the joint angle prediction based on the Gaussian process model, and "measurement" was the actual measurement

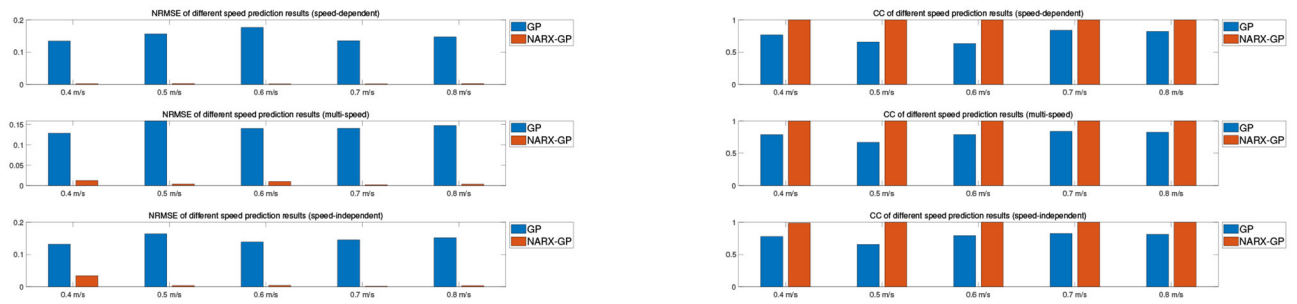


FIGURE 14 | Angle prediction results of NARX-GP model and GP model for hemiplegic subject dataset.

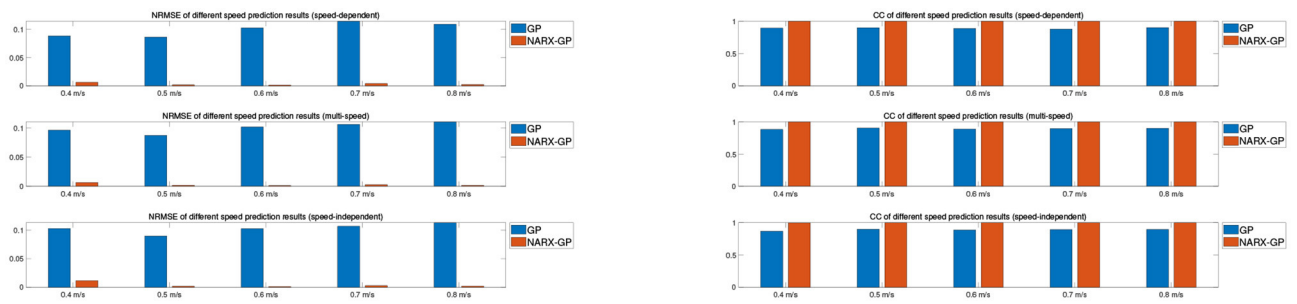


FIGURE 15 | Angle prediction results of NARX-GP model and GP model for hemiplegic subject dataset.

of joint angle. The gray shading was the 95% confidence interval for the prediction of the joint angle based on the Gaussian process autoregressive model to describe the uncertainty. From **Figures 11–13**, it can be concluded that the patient's knee joint NARX-GP model has better angle prediction than the GP model and can achieve good prediction results, and the joint angle prediction can approximate the actual joint angle for both the healthy side and the affected side (right side) of the patient.

The error assessment and statistical analysis of the prediction results of the NARX-GP model and the GP model for the hemiplegic subject dataset, the NRMSE and CC between the predicted and actual values of joint angles at different speeds are shown in **Figures 14**, **15**. The errors between the predicted results and the actual values of joint angles for both the left and right leg NARX-GP models were small, highly correlated, and significantly better than the GP model.

Further evaluation and analysis of variance of the prediction results for the left and right legs of the dataset showed that the mean NRMSE of the prediction results of the NARX-GP model for the left and right legs were 0.0063 ± 0.0081 and 0.0032 ± 0.0028, respectively, which were significantly better than those of the GP model (0.1466 ± 0.0127 and 0.1012 ± 0.0092), with significant differences ($p < 0.05$ and $p < 0.05$). Joint angle prediction using the NARX-GP model for both the healthy and affected side of the patient was able to have high accuracy with no significant difference ($p = 0.1919 > 0.05$). the NRMSE of the NARX-GP model prediction results for all scenarios was 0.0047 ± 0.0063 on average, which was better than the GP model results (0.1239 ± 0.0253) with a significant difference ($p < 0.05$).

4. DISCUSSION AND CONCLUSION

The EMG signal contains abundant motion information, which is ahead of the actual joint motion, and is often used as a control signal to predict joint motion. Therefore, EMG signals are widely used in applied scientific research related to the development of intelligent rehabilitation technologies and devices. EMG signal based modeling of the neuromusculoskeletal system, as an important component of joint motion prediction, has become a hot topic of research as it is important to help the development of rehabilitation techniques and equipment for patients with sports injuries. There is uncertainty in the neuromusculoskeletal system, and in order for the model to provide a description of the uncertainty, this paper proposes to model the uncertainty using a Gaussian autoregressive model. The muscle activation dynamics model was first introduced into the Gaussian process model to establish a joint angle prediction model based on Gaussian process. Due to the high requirement for joint angle prediction accuracy in practical applications and the fact that the neuromusculoskeletal system is a dynamic non-linear system, the NARX model was introduced into the Gaussian process model to establish a Gaussian autoregressive model to achieve

OSA prediction of knee joint angle. The results of different test scenarios on the healthy subjects and hemiplegic subject datasets showed that the designed Gaussian autoregressive model had significantly better prediction accuracy than the Gaussian process model, and there was no significant difference in the prediction accuracy between the affected and healthy sides of the hemiplegic subject, both of which were able to achieve more accurate prediction results for knee angles and could provide uncertainty information.

In this paper, a non-parametric model for knee joint angle prediction was developed from a predictive value-based NARX model approach by mixing a muscle activation kinetic model with a data-driven model. The proposed modeling approach was validated with a publicly available dataset. The proposed method utilizes only the EMG signals of a pair of antagonistic muscles, reducing the cost of EMG signal detection and the complexity of the model. However, there are still some shortcomings in this paper and there are many problems that have not yet been studied with some need for improvement. In this paper, the performance of only one pair of antagonist muscles in the hamstrings and quadriceps was tested, and the effect of sEMG signals from other muscles in the hamstrings and quadriceps as input on the accuracy of knee joint angle prediction can be further tested. In addition, although the knee joint is used as the research object for the study and validation of the model in this paper, the proposed joint angle prediction method is not limited to the knee joint angle prediction. In the subsequent research, the proposed method can be applied to the angle prediction of other joints for relevant testing and validation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

Conception of the study was conducted by JL, ZS, and WC. FZ processed the data. XC analyzed and interpreted the data with input and support from XC and JL. WC drafted the manuscript. YL, ZS, and WC revised the manuscript critically for important intellectual content. All authors read and approved the manuscript.

FUNDING

This work was supported in part by the Fujian Province Nature Science Foundation of China under Grant 2019J01544, in part by the National Nature Science Foundation of China under Grant 61773124.

ACKNOWLEDGMENTS

We thank the members of Fuzhou Second Hospital Affiliated to Xiamen University for valuable discussion and thanks to the Fujian Key Laboratory of Medical Instrumentation and Pharmaceutical Technology for providing the experimental environment and experimental equipment.

REFERENCES

1. Liu J, Kang SH, Xu D, Ren Y, Lee SJ, Zhang LQ. EMG-based continuous and simultaneous estimation of arm kinematics in able-bodied individuals and stroke survivors. *Front Neurosci*. (2017) 11:480. doi: 10.3389/fnins.2017.00480
2. Furukawa Ji, Noda T, Teramae T, Morimoto J. Human movement modeling to detect biosignal sensor failures for myoelectric assistive robot control. *IEEE Trans Robot*. (2017) 33:846–57. doi: 10.1109/TRO.2017.2683522
3. Côté-Allard U, Fall CL, Drouin A, Campeau-Lecours A, Gosselin C, Glette K, et al. Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE Trans Neural Syst Eng Rehabil*. (2019) 27:760–71. doi: 10.1109/TNSRE.2019.2896269
4. Corcos DM, Gottlieb GL, Latash ML, Almeida GL, Agarwal GC. Electromechanical delay: an experimental artifact. *J Electromyogr Kinesiol*. (1992) 2:59–68. doi: 10.1016/1050-6411(92)90017-D
5. Faisal AA, Selen LP, Wolpert DM. Noise in the nervous system. *Nat Rev Neurosci*. (2008) 9:292–303. doi: 10.1038/nrn2258
6. Berniker M, Jarc A, Kording K, Tresch M. A probabilistic analysis of muscle force uncertainty for control. *IEEE Trans Biomed Eng*. (2016) 63:2359–67. doi: 10.1109/TBME.2016.2531083
7. Chopp-Hurley JN, Langenderfer JE, Dickerson CR. Probabilistic evaluation of predicted force sensitivity to muscle attachment and glenohumeral stability uncertainty. *Ann Biomed Eng*. (2014) 42:1867–79. doi: 10.1007/s10439-014-1035-3
8. Lennart L. *System Identification: Theory for the User*. Upper Saddle River, NJ: PTR Prentice Hall (1999).
9. Zeng N, Wang Z, Liu W, Zhang H, Hone K, Liu X. A dynamic neighborhood-based switching particle swarm optimization algorithm. *IEEE Trans Cybernet*. (2020). doi: 10.1109/TCYB.2020.3029748. [Epub ahead of print].
10. Zeng N, Wang Z, Zhang H, Kim KE, Li Y, Liu X. An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips. *IEEE Trans Nanotechnol*. (2019) 18:819–29. doi: 10.1109/TNANO.2019.2932271
11. Zeng N, Wang Z, Zineddin B, Li Y, Du M, Xiao L, et al. Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach. *IEEE Trans Med Imaging*. (2014) 33:1129–36. doi: 10.1109/TMI.2014.2305394
12. Zeng N, Wang Z, Li Y, Du M, Cao J, Liu X. Time series modeling of nano-gold immunochromatographic assay via expectation maximization algorithm. *IEEE Trans Biomed Eng*. (2013) 60:3418–24. doi: 10.1109/TBME.2013.2260160
13. Roberts S, Osborne M, Ebden M, Reece S, Gibson N, Aigrain S. Gaussian processes for time-series modelling. *Philos Trans R Soc A Math Phys Sci Eng*. (2013) 371:20110550. doi: 10.1098/rsta.2011.0550
14. Eslamy M, Alipour K. Synergy-based Gaussian process estimation of ankle angle and torque: conceptualization for high level controlling of active robotic foot prostheses/orthoses. *J Biomech Eng*. (2019) 141:021002. doi: 10.1115/1.4041767
15. Hamaya M, Matsubara T, Noda T, Teramae T, Morimoto J. Learning assistive strategies for exoskeleton robots from user-robot physical interaction. *Pattern Recogn Lett*. (2017) 99:67–76. doi: 10.1016/j.patrec.2017.04.007
16. Kang H, Park F. Motion optimization using Gaussian process dynamical models. *Multibody Syst Dyn*. (2015) 34:307–25. doi: 10.1007/s11044-014-9441-8
17. Schearer EM, Liao YW, Perreault EJ, Tresch MC, Memberg WD, Kirsch RF, et al. Identifying inverse human arm dynamics using a robotic testbed. In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Chicago, IL: IEEE (2014). p. 3585–91. doi: 10.1109/IROS.2014.6943064

18. Ullauri JB, Peternel L, Ugurlu B, Yamada Y, Morimoto J. On the EMG-based torque estimation for humans coupled with a force-controlled elbow exoskeleton. In: *2015 International Conference on Advanced Robotics (ICAR)*. Istanbul: IEEE (2015). p. 302–7. doi: 10.1109/ICAR.2015.7251472
19. Xiloyannis M, Gavriel C, Thomik AA, Faisal AA. Gaussian process autoregression for simultaneous proportional multi-modal prosthetic control with natural hand kinematics. *IEEE Trans Neural Syst Eng Rehabil*. (2017) 25:1785–801. doi: 10.1109/TNSRE.2017.2699598
20. Yang X, Yan J, Chen Z, Ding H, Liu H. A proportional pattern recognition control scheme for wearable a-mode ultrasound sensing. *IEEE Trans Ind Electron*. (2019) 67:800–8. doi: 10.1109/TIE.2019.2898614
21. Belić JJ, Faisal AA. Decoding of human hand actions to handle missing limbs in neuroprosthetics. *Front Comput Neurosci*. (2015) 9:27. doi: 10.3389/fncom.2015.00027
22. Zeng Y, Yang J, Yin Y. Gaussian process-integrated state space model for continuous joint angle prediction from EMG and interactive force in a human-exoskeleton system. *Appl Sci*. (2019) 9:1711. doi: 10.3390/app9081711
23. Raj R, Sivanandan K. Comparative study on estimation of elbow kinematics based on EMG time domain parameters using neural network and ANFIS NARX model. *J Intell Syst Fuzzy*. (2017) 32:791–805. doi: 10.3233/JIFS-16070
24. Li Z, Hayashibe M, Fattal C, Guiraud D. Muscle fatigue tracking with evoked EMG via recurrent neural network: toward personalized neuroprosthetics. *IEEE Comput Intell Mag*. (2014) 9:38–46. doi: 10.1109/MCI.2014.2307224
25. Raj R, Ramakrishna R, Sivanandan KS. A real time surface electromyography signal driven prosthetic hand model using PID controlled DC motor. *Biomed Eng Lett*. (2016) 6:276–86. doi: 10.1007/s13534-016-0240-4
26. Gupta R, Dhindsa IS, Agarwal R. Continuous angular position estimation of human ankle during unconstrained locomotion. *Biomed Signal Process Control*. (2020) 60:101968. doi: 10.1016/j.bspc.2020.101968
27. Liu Z, Wang X, Su M, Le L. Research on rehabilitation training bed with action prediction based on NARX neural network. *Int J Imaging Syst Technol*. (2019) 29:539–46. doi: 10.1002/ima.22334
28. Raj R, Sivanandan K. Elbow joint angle and elbow movement velocity estimation using NARX-multiple layer perceptron neural network model with surface EMG time domain parameters. *J Back Rehabil Musculoskel*. (2017) 30:515–25. doi: 10.3233/BMR-160525
29. Zajac FE. Muscle and tendon: properties, models, scaling, and application to biomechanics and motor control. *Crit Rev Biomed Eng*. (1989) 17:359–411.
30. Buchanan TS, Lloyd DG, Manal K, Besier TF. Neuromusculoskeletal modeling: estimation of muscle forces and joint moments and movements from measurements of neural command. *J Appl Biomech*. (2004) 20:367–95. doi: 10.1123/jab.20.4.367
31. Lloyd DG, Besier TF. An EMG-driven musculoskeletal model to estimate muscle forces and knee joint moments *in vivo*. *J Biomech*. (2003) 36:765–76. doi: 10.1016/S0021-9290(03)00010-1
32. Li Y, Chen W, Yang H, Li J, Zheng N. Joint torque closed-loop estimation using NARX neural network based on sEMG signals. *IEEE Access*. (2020) 8:213636–46. doi: 10.1109/ACCESS.2020.3039983
33. Schulz E, Speekenbrink M, Krause A. A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions. *J Math Psychol*. (2018) 85:1–16. doi: 10.1016/j.jmp.2018.03.001
34. Williams CKI. *Gaussian Processes for Machine Learning*. Taylor & Francis Group (2006).
35. Stein ML. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media (2012).
36. Leontaritis I, Billings SA. Input-output parametric models for non-linear systems part I: deterministic non-linear systems. *Int J Control*. (1985) 41:303–28. doi: 10.1080/0020718508961129
37. Leontaritis I, Billings SA. Input-output parametric models for non-linear systems part II: stochastic non-linear systems. *Int J Control*. (1985) 41:329–44. doi: 10.1080/0020718508961130
38. Hamner SR, Delp SL. Muscle contributions to fore-aft and vertical body mass center accelerations over a range of running speeds. *J Biomech*. (2013) 46:780–7. doi: 10.1016/j.jbiomech.2012.11.024
39. Meyer AJ, Patten C, Fregly BJ. Lower extremity EMG-driven modeling of walking with automated adjustment of musculoskeletal geometry. *PLoS ONE*. (2017) 12:e0179698. doi: 10.1371/journal.pone.0179698
40. Babak S. *Biostatistics With R: An Introduction to Statistics Through Biological Data*. New York, NY: Springer New York (2012).

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 Public Health

ORIGINAL RESEARCH
published: 12 July 2021
doi: 10.3389/fpubh.2021.712827

Research on the Construction and Application of Breast Cancer-Specific Database System Based on Full Data Lifecycle

Yin Jin[1,2], *Wang Junren*[2,3], *Jiang Jingwen*[2,3], *Sun Yajing*[2,3], *Chen Xi*[4]* *and Qin Ke*[1]

[1] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China, [2] West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China, [3] Medical Big Data Center, Sichuan University, Chengdu, China, [4] Chengdu Zhixin Electronic Technology Co., Ltd, Chengdu, China

Relying on the Biomedical Big Data Center of West China Hospital, this paper makes an in-depth research on the construction method and application of breast cancer-specific database system based on full data lifecycle, including the establishment of data standards, data fusion and governance, multi-modal knowledge graph, data security sharing and value application of breast cancer-specific database. The research was developed by establishing the breast cancer master data and metadata standards, then collecting, mapping and governing the structured and unstructured clinical data, and parsing and processing the electronic medical records with NLP natural language processing method or other applicable methods, as well as constructing the breast cancer-specific database system to support the application of data in clinical practices, scientific research, and teaching in hospitals, giving full play to the value of medical big data of the Biomedical Big Data Center of West China Hospital.

Keywords: breast cancer, disease-specific database, metadata, data governance, data security sharing, knowledge graph

OPEN ACCESS

Edited by:
Yonghong Peng,
Manchester Metropolitan University,
United Kingdom

Reviewed by:
Gang Liao,
Chongqing Medical University, China
Lingzhong Meng,
Institute of Software, Chinese Academy of Sciences, China

***Correspondence:**
Chen Xi
chenxi@standata.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *21 May 2021*
Accepted: *14 June 2021*
Published: *12 July 2021*

Citation:
Jin Y, Junren W, Jingwen J, Yajing S, Xi C and Ke Q (2021) Research on the Construction and Application of Breast Cancer-Specific Database System Based on Full Data Lifecycle. Front. Public Health 9:712827. doi: 10.3389/fpubh.2021.712827

INTRODUCTION

With the rapid development of new technologies such as big data and artificial intelligence, the medicine overlaps with such disciplines as information technology, computer science, and cyber security in more and more aspects. Particularly, thanks to the constant advancement of medical technology, the process for screening, diagnosis and treatment of diseases is being expanded to generate various new data. Based on different data modalities, artificial intelligence technology has been widely applied in the field of medicine (1–5). This research focuses on breast cancer which results in the second highest cancer mortality in women (6) and its screening, diagnosis and treatment strategies which have developed from single surgical therapy to a comprehensive treatment mode that combines surgical therapy, chemotherapy, radiotherapy, endocrinotherapy, and targeted therapy, forming a multi-disciplinary team (MDT) of breast cancer. Multi-source heterogeneous data, such as electronic medical record data, image data and gene data, were generated in the whole diagnosis and treatment process to drive the disease diagnosis and treatment and disease research into a big data era of disease-specific research (7).

Currently, the most fundamental challenge confronting the medical research institutions in the development of disease-specific data research is how to integrate multi-source heterogeneous data and build a disease-specific database, so as to support the discovery of potential diagnosis and treatment knowledge patterns from the massive medical data. For example, the data are too much to be screened manually by clinicians; the attributes of the same data element are described in different ways across hospitals or even across information systems of the same hospital, e.g., the same drug may be identified by different codes. As unstructured data such as electronic medical records contain medical information of important value, the difficulty lies in the effective and accurate extraction of such information. Based on the medical big data governance activities in China, Li et al. proposed a big data governance framework for medical data in China through literature review, expert consulting and structural modeling, providing an important reference for data governance framework in this research (8). This paper aims to integrate two heterogeneous clinical data sources, i.e., unstructured medical records and structured clinical data, through clinical text analysis and knowledge extraction; to break the information barriers within the organization and between clinical departments and to promote data sharing among medical centers in combination with patient information from multiple clinical data sources; to establish the disease-specific data standards in accordance with international industry standards and then to construct a multi-modal knowledge graph specific to breast cancer; finally to build a disease-specific database system for the purpose of analyzing disease characteristics, thus providing supports in clinical decision-making and rational drug use to clinicians in the diagnosis and treatment of breast cancer.

THE INVOLVEMENT OF BREAST CANCER PATIENTS

Patients pathologically diagnosed with breast cancer are prospectively registered in the Breast Cancer-specific Database System at West China Hospital, Sichuan University since 2008 (9, 10). Medical records, diagnostic pathology reports, treatment records are recorded by oncologists. All patients are followed by outpatient visit or telephone at 3–4-month intervals within 3 years after diagnosis, 6-month intervals within 4–5 years, and then annually. The characteristics description of breast cancer patients included in the database is shown in **Table 1**.

TABLE 1 | Characteristics description among breast cancer patients at recruitment.

	Group	Value
Age, median (IQR)		41.0 (47.0, 55.0)
Sex	Female	7697 (99.6%)
BMI, median (IQR)		20.83 (22.86, 24.97)
Menopause status, No (%)	Yes	2999 (38.8%)
	No	4693 (60.7%)
	Unknown	38 (0.5%)
Stage, No (%)	0	0
	I	1619 (25%)
	II	3427 (33.5%)
	III	1927 (29.8%)
	Unknown	757 (11.8%)
pT status, No (%)	0	308 (4.0%)
	1	2532 (32.8%)
	2	3633 (47.0%)
	3	336 (4.3%)
	4	440 (5.7%)
	Unknown	481 (6.2%)
pN status, No (%)	0	3748 (48.5%)
	1	2227 (28.9%)
	2	856 (11.1%)
	3	783 (10.1%)
	Unknown	116 (1.4%)
ER	Negative–	2323 (30.1%)
	Positive+	5094 (65.9%)
	Unknown	313 (4.0%)
PR	Negative–	2680 (34.7%)
	Positive+	4737 (61.3%)
	Unknown	313 (4.0%)
HER2	Negative–	4555 (58.9%)
	Positive+	1856 (24.0%)
	Unknown	1319 (17.1%)
Ki67	<14%	1370 (17.8%)
	≥14%	5752 (74.4%)
	Unknown	608 (7.8%)

ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor-2; Ki-67, marker of proliferation Ki-67.

OVERALL DESIGN SCHEME FOR BREAST CANCER-SPECIFIC DATABASE SYSTEM

Overall design thought of the breast cancer-specific database system is shown in **Figure 1**.

Several important oncology patient data systems of China and foreign countries (e.g., cancer registration software CanReg designed and developed by the Descriptive Epidemiology Unit of IARC\the cancer screening database of Chinese Anti-Cancer Association) were referred in the overall design and construction of the breast cancer-specific database system (11, 12). The database is composed of four parts: a patient standard database, a breast cancer malignancy-specific database, a diagnostic imaging database and a breast cancer patient follow-up visit database. Data governance is based on these four parts. The construction of the disease-specific database system involves governance, extraction and application. Governance is performed firstly to collect the current disease-specific data assets of the West China Hospital and sort out their meanings, ownership, etc. The next step is to conduct data classification and quality control to ensure the accuracy of data processing. The last step is to provide a unified standard interface services based on the governed and integrated disease-specific data center.

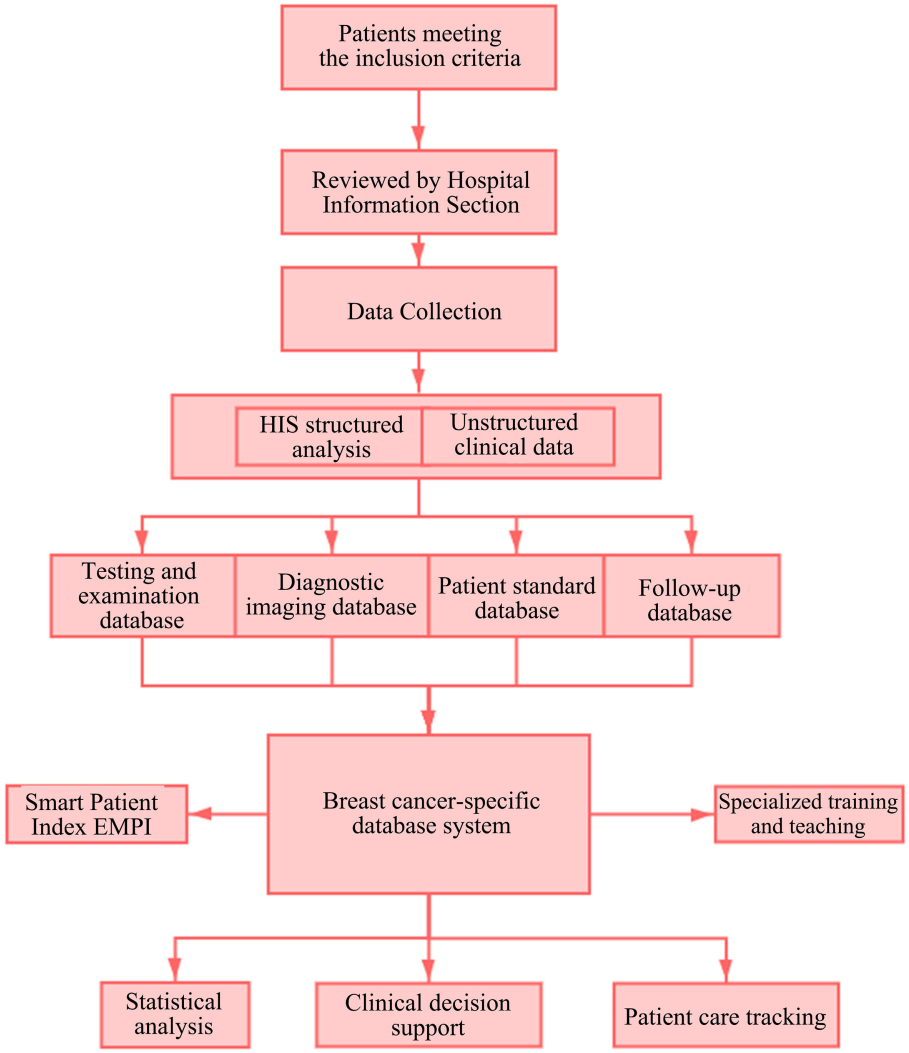


FIGURE 1 | Overall design of breast cancer-specific database system. HIS, Hospital Information System; EMPI, Enterprise Master Patient Index.

In addition, main considerations in design of the disease-specific database include the application of the standardized and governed data in clinical diagnosis and treatment assistance and scientific research. The patient's medical record data can be viewed as a time series that captures the entire clinical process of collecting the patient's medical history, analyzing the condition, diagnosing, and treating the patient. Different data sources have different time spans, resulting in complex timing dependencies between events (13). Therefore, combined with the actual application scenarios and for better support to scientific research, the overall design and final data presentation of the disease-specific database are logically linked through the processes of admission registration, inpatient treatment, checkout and discharge, etc. in the chronological sequence, so as to establish a view for diagnosis and treatment based on full data lifecycle.

METADATA-CENTERED DATA GOVERNANCE SCHEME

Standards–Establishment of Data Standards

Establishment of Dataset Standards

A regional breast cancer-specific database should include a complete range of datasets in the uniform format and meeting the normative standards. In addition, the database should incorporate the national and medical industry standards and all system datasets of medical institutions. A synonym database of the dataset names should be created. A standard database of breast cancer datasets should be created for the Center to provide a dataset graph for its application and corpus support for automatic identification of dataset names. The Center's database includes the scope, normative references, term abbreviations,

TABLE 2 | Some referenced dataset standards.

Classification	Standard name
National health industry standards	Basic Dataset of Basic Information–Personal Information (WS 371-2012) Basic Items Data Set of Health Examination (T/CHIA 2-2018) Guidelines for Data Schema Description of Health Information (WS/T 304-2009) Classification and Coding for Value Domain of Health Data Element (WS364-2011)
International oncology specialized standards	SEER Program Coding and Staging Manual of National Cancer Institute (NCI) International Classification of Diseases, Tenth Revision (ICD-10) International Classification of Diseases for Oncology, Third Edition (ICD-O-3), etc.
Other standards	Supplement to current hospital datasets

WS and WS/T refers to standards issued by Chinese Professional Committee on Health Information Standards; T/CHIA refers to standards issued by Chinese Health Information Association.

dataset metadata attributes, and data element attributes. Standard information for each dataset includes the dataset name, identifier, classification, field description, definition, etc. The datasets are saved in four different databases based on data types, and subdivided into four modules and about 20 submodules. The oncology datasets and breast cancer-specific datasets are created accordingly with reference to different health industry standards. Some referenced dataset standards are shown in **Table 2**, and some collected fields of the breast cancer-specific dataset are shown in **Table 3**.

Establishment of Data Element Standards

A local database of data element standards is established based on the national and industrial standards and in combination with the specific situation of the hospital. The local database includes data element indicators, normative references, term abbreviations, and data element directory. Data element standards specify the Chinese name, English name, identifier, definition, classification, data type, representation format, data threshold value, allowable value type and allowable value of data fields in the data dictionary, which are used to ensure the data quality. In the management of metadata, data elements may be classified and labeled, so as to establish a synonym database of the data elements. The local database describes the attributes of each data element, including Chinese field name, English field name, field name abbreviation, field type, field length, required or not, range or reference standards, notes and remarks. If there are any relevant international standards for the range of data elements, they can be referenced directly; otherwise, the range will be set by physicians and other professionals in combination with clinical experience. The set range standards will be saved together with other collected standards in the local database for the convenience of version management and subsequent updates. **Table 4** shows partial attributes of some data elements.

Quality Control–Fusion and Governance of Multi-Source Heterogeneous Data

The structured medical data from HIS (Hospital Information System), LIS (Laboratory Information Management System), and follow-up visit system are integrated with the image data from PACS (Picture Archiving and Communication Systems). These data are acquired by building an ETL (Extract-Transform-Load) automation platform to perform incremental extraction at regular intervals on a daily basis, and complete data standardization and other processes during the extraction process. The unstructured data in the electronic medical record are structured through natural language processing and machine learning after data source access, and then saved in the disease-specific database.

Afterwards, the data in the four module databases are linked primarily based on the patient ID, thus breaking the information barriers within the organization and between clinical departments. Finally, the front-end application is supported by breast cancer-specific data for fully mining the data of full lifecycle about single disease and providing support for data analysis of multi-center joint scientific research projects. Specific processing methods are described below. The data governance framework is shown in **Figure 2**.

Structured Data Processing

Data acquisition (data reception or data capture) is performed through the data fusion platform for the data of breast cancer patients which are structured but exist in different systems. The source data are extracted, integrated and saved in the target database as per the following steps: (1) Establish a data source directory, and determine the connection mode, access permission, data storage directory, and interfaces of each data source; (2) Data cleaning and filtering: Establish data review rules, e.g., the gender can only be male, female, or unknown, the ID number can only be 18 digits, the patient ID cannot be blank, etc. Then, filter the data according to these rules, and save the unqualified data in a temporary database, with no need for data fusion. The cleaned data should not contain missing or incomplete data, repeated data and nonstandard data. See **Figure 3** for the statistics of some cleaned data; (3) Map the original data in the data source database with the standard datasets in accordance with the specified data standards, and complete the range conversion of data elements at the same time to standardize the processing of breast cancer standard data, so as to complete the collection, and collation of multi-source data; (4) In the process of timed automatic incremental extraction of medical data, monitor the log for each extraction, and count the number of extraction records and completion for later failure rollback (14).

Unstructured Data Processing

Electronic medical records contain highly valuable medical data. The unstructured breast cancer data are parsed by the standard medical structure based on natural language text data, and the structured data correction, annotation and association tools are provided for clinicians to manage the annotation tasks (either by manual or automatic annotation) of the text data to be

TABLE 3 | Some collected fields of the breast cancer-specific dataset.

Classification	Field name	Database type
Basic information	Patient ID, patient name, home address, contact number, first contact name, relationship with patient, ID number, gender, date of birth, height, weight, gender, age, marital status, occupation, ethnicity, ancestral home, nationality, etc.	Patient standard database
Inpatient information	Patient ID, time of medical records, reliability, medical unit admitted, nursing unit admitted, time of admission by the current medical unit, current medical unit, current nursing unit, date of diagnosis, time of admission, length of stay, time of discharge, transfer of departments, date of transfer from current medical unit, medical unit transferred, nursing unit transferred, attending physician, way of discharge, number of operations, etc.	Patient standard database
Progress note	Patient ID, chief complaint, clinical pathway ID, medical advice, observations, results of ward round, dosing regimen, summary opinions, etc.	Disease-specific database
Nursing assessment	Details of occupational exposure, smoking status, duration of smoking, average number of cigarettes, smoking cessation, duration of smoking cessation, drinking, duration of drinking, average number of drinks, allergy history, details of allergy history, diet, general health condition, vaccination history, past history of serious illness, details of serious illness, history of blood transfusion, trauma history, history of infectious diseases, details of infectious diseases, history of surgery, details of surgery, etc.	Patient standard database
Diagnostic information	Diagnosis category, diagnosis code, diagnosis name, pre- and post-operation diagnostic accordance, outpatient diagnostic accordance, clinical case diagnostic accordance, radiopathological diagnostic accordance, discrepancy between admission diagnosis and primary discharge diagnosis, cataloged diagnosis name splicing, cataloged diagnosis code splicing, first page diagnosis name splicing, tumor morphological code name, tumor morphological code, etc.	Disease-specific database
Physical examination	Body temperature, pulse rate, respiratory rate, blood pressure, general condition, skin mucosa, lymph nodes, head, hair distribution, eyes, ears, nose, mouth, face, neck, chest, lungs, heart, blood vessels, abdomen, genitalia, anorectum, spine and extremities, nervous system, routine examinations, specialist examinations, etc.	Patient standard database
Testing	LIS reported DR, test time, item number in test results, item name in test results, sample code, sample name, reference value range, quantitative result, item unit, label, result, etc.	Patient standard database
Examination	Mass size, distribution of lesions (single or multicenter), tumor location, presence of distant metastasis, etc.	Image database
Surgical anesthesia	Date, operation level, anesthesia level, incision type, anesthesiologist, operation code & name, operation time, surgeon, preoperative and postoperative diagnosis, preoperative chemotherapy, radiotherapy, anesthesia method, intraoperative bleeding, blood transfusion, etc.	Disease-specific database
Treatment information	Inpatient diagnosis and treatment plan, type of medical advice, item name of medical advice, frequency, usage, implementation date of medical advice, invalidation date of medical advice (for long-term treatment), source of medical advice, treatment means, etc.	Disease-specific database
Postoperative radiotherapy for tumor	Measurement of radiotherapy (single and cumulative), start and end time, adverse reactions, etc.	Disease-specific database
Postoperative chemotherapy for tumor	Chemotherapy regimen (i.e., drug type and dosage, route of administration), cycle, start and end time, adverse reactions, etc.	Disease-specific database
Disease progression and outcome	Conditions at admission, chief complaint, summary of medical record, course of disease (not available), discharge summary, main discharge diagnosis and treatment, etc.	Disease-specific database
Follow-up visit	Time, survival status, recurrence, metastasis, adverse reactions, etc.	Follow-up visit database
Charges	Charges for outpatient service, hospitalization, operation, examination, testing, drugs, etc.	Disease-specific database

LIS, Laboratory Information Management System; DR, Digital radiography.

processed, so that the unstructured text data in breast cancer pathology reports and present medical history are transformed into analyzable structured data, providing a data basis for the construction of a subsequent consensus data link mining engine, an analysis tool for self-defined data link risk factors and a breast disease knowledge graph. Natural language texts (such as current medical history, color Doppler ultrasound description and pathological description) are annotated by professional physicians for entities and relations until 200 annotations, and preliminary training is developed, then back-annotation is performed using the trained model to assist the physicians in annotation of subsequent samples. At present, 1,000 samples have been annotated and trained, and completed for model training and model evaluation using NLP to make profound adjustment to model parameters. The model ability is evaluated, with the recognition accuracy of 80–85%, reaching the level of manual recognition by general physicians. The parsing results of some electronic medical records are shown in **Figure 4**.

Image Data Processing

Medical imaging technology has increasingly become an indispensable means for disease diagnosis, providing quick and

TABLE 4 | Partial attributes of some standard data elements.

Field name	Field type	Field length	Required or not	Range or reference standard
Gender code	varchar	10	Y	GB/T 2261.1-2003
Marital status code	varchar	10	Y	GB/T 2261.2-2003
Health insurance category code	varchar	10	N	CVO2.01.204 Table for Health Insurance Category Code
Medical history	varchar	200	N	0-No, 1-Yes
Registration category	varchar	20	Y	01-General clinic, 02-Emergency, 03-Specialty clinic, 04-Specialist clinic, 05-VIP clinic, 06-Disease-specific clinic, 09-Others
Diagnosis basis	varchar	20	Y	CT05.01.001
Prescription type/name	varchar	20	Y	01-General prescription, 02-Pediatric prescription, 03-Emergency prescription, 04-Narcotic drug prescription (Class I psychotropic drug prescription), 05-Narcotic drug prescription (Class II psychotropic drug prescription), 99-Others
Dosing frequency code	varchar	20	N	CV06.00.228
Examination site code	varchar	60	N	CV06.00.227
Surgical procedure code	varchar	20	Y	ICD-9-CM-3
Anesthesia mode code	varchar	20	Y	CV06.00.103
Surgical position code	varchar	10	Y	CV06.00.223
ASA physical status classification code	varchar	10	Y	CV05.10.021

GB/T refers to standard issued by China National Institute of Standardization; CV and CT refer to classification code table in standard issued by China National Institute of Standardization.

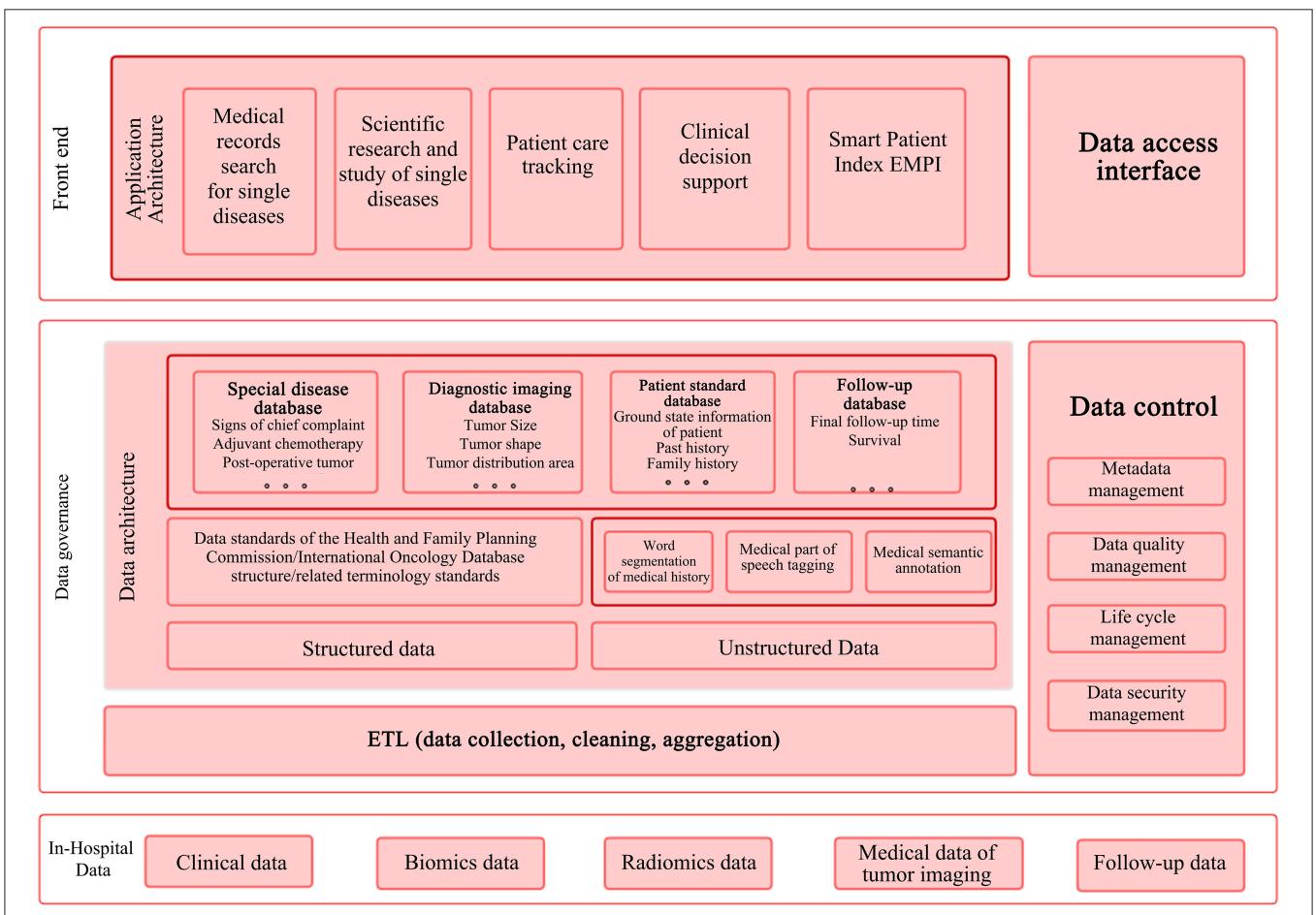


FIGURE 2 | Data governance framework. ETL, Extract-Transform-Load.

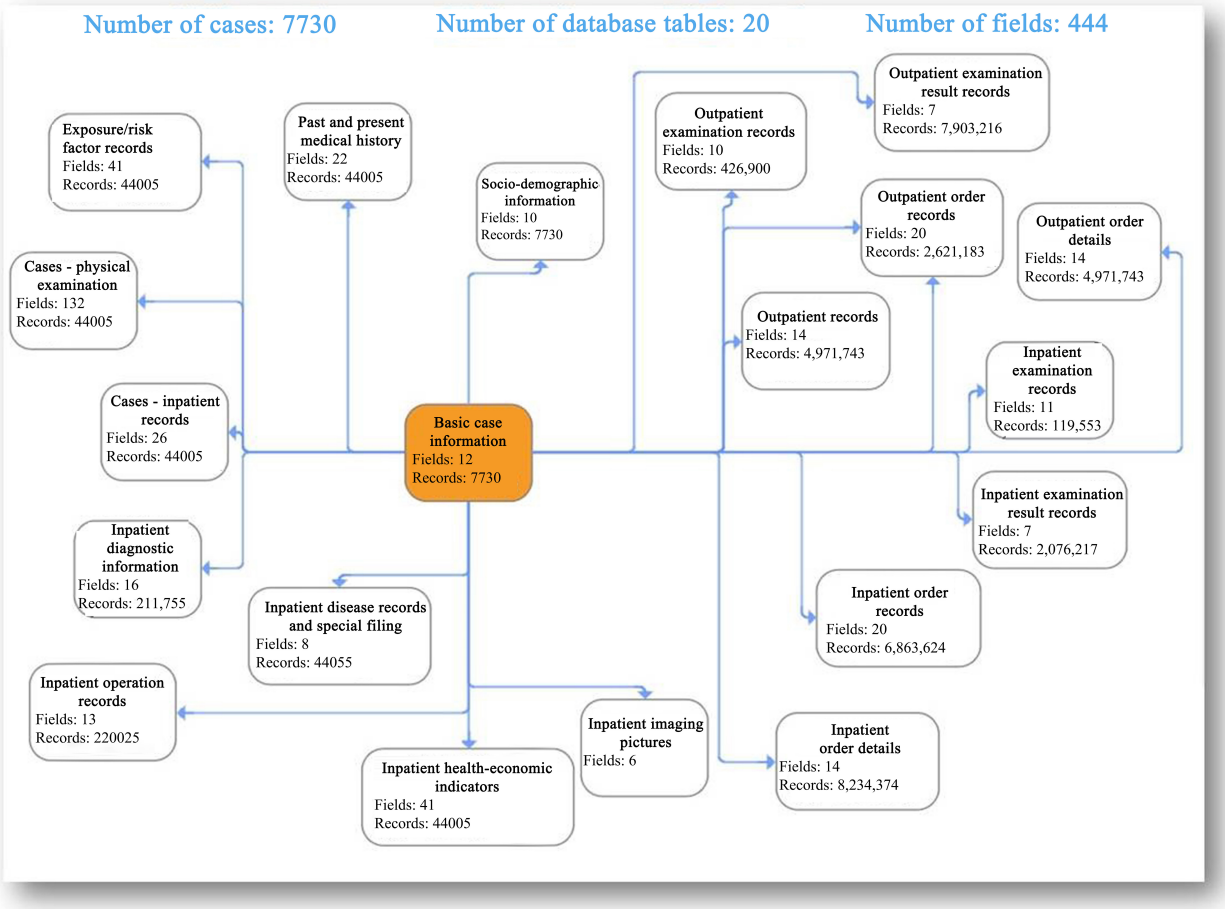


FIGURE 3 | Database cleaning results.

accurate support for clinical practice. The imaging information about a patient's tumor site and the examination report issued by the radiologist are saved in the medical image data. In the process of image data governance, it is necessary to eliminate unqualified image data. With AI deep neural network machine learning technique, the machine can automatically distinguish the unqualified image data and annotate such data (15). The filtered unqualified images are saved in the temporary cache database and manually verified later. The qualified image data are extracted and saved together with basic patient information and report results in the diagnostic imaging database, and finally linked with the patient standard database and the disease-specific database based on the patient ID to connect the whole treatment process. The image data governance process is shown in **Figure 5**.

Model-Building - Multi-Modal Breast Cancer-Specific Knowledge Graph

On the basis of the breast cancer-specific database system, a multi-modal breast cancer-specific knowledge graph is constructed to integrate texts, medical images, and even videos, voices and other rich media information, and to reflect the hierarchical relation among the entities and relations related to breast cancer such as pathogenesis, symptom characteristics, complications, treatment means, medical history, and medication in the form of node network graph. Such a centralized and clear structure can help researchers quickly clarify the relations and differences among numerous and complex knowledge points (16). AI mining engine is constructed to identify valuable hidden relations from the huge breast cancer-specific database and analyze such relations through clustering, attribute comparison and AI active learning, and the results are reviewed by experts and incorporated into the knowledge graph if passing the review. After deep knowledge data mining, the cross-departmental and even cross-hospital knowledge relations can be established only for dynamic knowledge graphs based on multivariate knowledge graph, thus expanding the entity set, relation set and triple set of knowledge graph. Meanwhile, the entity is not limited to the single representation only in breast cancer-related terms. The traditional knowledge graph is out of use, and the current knowledge graph integrates multi-modal

Type of entity	Entity
Date	2013/8/21
Site	Right breast mass
Surgery	Right breast mass excision
Examination	Pathological examination
Examination results	Invasive ductal carcinoma
Examination results	Size 4x2cm
Date	2013/8/21
Date	2013/8/27
Surgery	Modified radical mastectomy for right breast cancer
Examination	Postoperative pathological examination
Examination results	Metastasis of cancer was detected in 1/7 of the right axillary lymph nodes
Date	2013/8/30
Examination	Consultation
Examination results	Invasive ductal carcinoma of right breast
Examination results	With intraductal cancer component
Examination	ER
Examination results	+
Examination results	90%
Examination	PR
Examination results	+
Examination results	60%
Examination	HER-2
Examination results	3+
Examination	Ki-67

Entity a	Entity b	Relation
Postoperative pathological examination	Metastasis of cancer was detected in 1/7 of the right axillary lymph nodes	Examination - results
Postoperative pathological examination	Tumor size 4x2cm	Examination - results
Right breast mass excision	41507	Surgery - date
Modified radical mastectomy for right breast cancer	41513	Surgery - date
ER	+	Examination - results
ER	0.9	Examination - results
PR	+	Examination - results
PR	0.6	Examination - results
HER-2	3+	Examination - results
Ki-67	+	Examination - results
Ki-67	0.2	Examination - results
Targeted therapy	Herceptin	Targeted therapy - drug name
Epirubicin	160mgdl	Drug name - dose
Epirubicin Hydrochloride for Injection	160mgdl	Drug name - dose
EC-T regimen chemotherapy	Epirubicin	Chemotherapy - drug name
EC-T regimen chemotherapy	Epirubicin Hydrochloride for Injection	Chemotherapy - drug name
EC-T regimen chemotherapy	Cyclophosphamide	Chemotherapy - drug name
Cyclophosphamide	900mgdl	Drug name - dose
EC-T regimen chemotherapy	Docetaxel	Chemotherapy - drug name
EC-T regimen chemotherapy	Docetaxel Injection	Chemotherapy - drug name
Docetaxel Injection	160mg	Drug name - dose
Docetaxel	160mg	Drug name - dose
CT	4.1x3.9 cm low density mass	Examination - results
Upper abdomen CT	5.0*3.8cm low density block shadow	Examination - results
Upper abdomen CT	Small nodule shadows visible around	Examination - results
Upper abdomen CT	Similar low-density nodular shadows visible in the tail lobe	Examination - results
Upper abdomen CT	About 1.6 cm in diameter	Examination - results
Upper abdomen CT	42720	Examination - date
Upper abdomen CT	Weakly reinforced nodules visible in the liver	Examination - results

FIGURE 4 | Parsing results of some electronic medical records. ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor-2; Ki-67, marker of proliferation Ki-67.

knowledge, and displays, represents and utilizes medical data in various forms to the largest extent for the convenience of learning and understanding by researchers (17, 18). The partial knowledge graph constructed is shown in **Figure 6**.

MAINTENANCE OF DATA SECURITY AND SHARING OF THE ACHIEVEMENTS IN DISEASE-SPECIFIC DATABASE

After the breast cancer-specific database system is constructed, in order to maximize its value in clinical practice and scientific research, rather than being limited to the inquiry and use in the Center and the Hospital, the data should be shared to multiple parties. However, medical data are particularly sensitive, so the following solutions will be adopted to share the data under the premise of ensuring the data security and patient privacy.

Due to the high security requirements of medical information data and the small volume of data in a single institution, federated learning, or multi-party secure computation can be considered to achieve joint use of data by multiple parties while ensuring data privacy and security. Essentially, both approaches limit the data use to the specified scope, which is effective to avoid data leakage and abuse. Federated learning is a distributed machine learning technique which collaborates the data modeling by multiple parties without data exchange (19). This model ensures the privacy of medical data and allows scientific research. Multi-party secure computation technically ensures that multiple parties of data cannot obtain the original data, and realizes collaborative computation without data leakage, that is, multiple parties run a

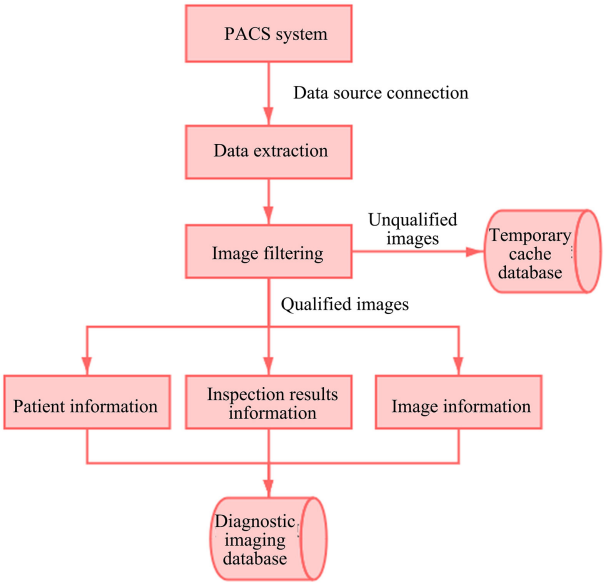


FIGURE 5 | Image data governance process. PACS, Picture Archiving and Communication Systems.

computing task, machine learning task and data retrieval so as to obtain the final results based on common data, but data and intermediate calculation results will not be disclosed to any party in this process.

DEVELOPING DATA VALUE TO SUPPORT CLINICAL RESEARCH

Support scientific research. The disease-specific database system will provide massive datasets for scientific research. Researchers can precisely filter target data according to their different research needs. Data analysis tools are also provided to improve data processing capabilities of the Hospital. Researchers can customize screening modes from the aspects of data sources, data time periods, data label types, verified knowledge, etc. to efficiently and accurately retrieve target data from massive data, thus indirectly improving clinical scientific research capabilities.

Support clinical practice. The diagnosis and treatment data of a breast cancer patient can be synchronized in real time through the "Breast Cancer-specific Database System" to form a data file which includes diagnostic imaging data, clinical pathology data, basic patient information, medical advice information, medication, surgery, radiotherapy, chemotherapy, cost settlement, etc., as well as the associated breast cancer knowledge database, etc. Combined with the "multi-modal breast cancer-specific knowledge graph" and based on the database-wide medical big data, various quantitative or qualitative big data machine learning algorithms are utilized for data analysis (20–22) to output the holographic knowledge portrait analysis reports of the patient's breast cancer risk profile, disease trend, clinical protocol, etc., such as the possibility of certain conclusion and the proportion of certain therapeutic regimen, providing the physicians with multi-dimensional and rich reference information, improving the ability of junior physicians in identification, diagnosis and treatment, and reducing the probability of missed diagnosis and misdiagnosis. Physicians can intuitively view, analyze and integrate multi-dimensional and multi-level holographic knowledge portrait, thus providing reference knowledge for accurate diagnosis and treatment of breast cancer based on the full-volume data. With the help of the auxiliary diagnostic system, physicians can provide more accurate therapeutic regimen based on the stage of the cancer and the patient's physical condition. Meanwhile, the intelligent auxiliary diagnosis system for breast cancer also provides a whole course management tool covering the examination, treatment and follow-up visit, so that the physicians can optimize the therapeutic regimen as appropriate in a timely manner, improve the treatment effect, and also provide more valuable data for the normalization and standardization of breast cancer treatment while using it (23).

Support teaching. Based on the whole-process therapeutic regimen in the breast cancer-specific database and the real physiological data of patients, theoretical learning and practice are carried out simultaneously for teachers and students. The most important thing is that the data are real and updated in real time update, so they are more instructive.

CONCLUSION

A breast cancer-specific database system based on full data lifecycle, by integrating the data and processes of existing clinical data systems, accumulates knowledge database, provides standard access interface and back feeds business integration to promote the optimization and transformation of existing disease-specific research processes and form a closed loop for sustainable development. The disease-specific database system covers several disease-specific databases for conveniently saving and managing patient data in a systematic, standardized and accurate manner, so as to realize the tracking of breast cancer cases, and effectively develop teaching, scientific research and evaluation on the effects of various therapies for breast cancer. A scientific platform is created for research on breast cancer pathogenesis and etiology through comprehensive long-term longitudinal tracking and data comparison/analysis.

Clinical text analysis and knowledge extraction are conducted to integrate two heterogeneous clinical data sources, that is, unstructured medical record data and structured clinical data. New-generation information technologies, such as big data, NLP text parsing, data mining and knowledge graph, are deeply fused and applied to build a disease-specific database system based on full data lifecycle for the purpose of breast cancer disease characteristic analysis, so as to effectively develop teaching, scientific research and evaluation on the diagnosis and treatment of breast cancer and the follow-up visit tracking of cases, conduct comprehensive long-term longitudinal tracking and data comparison and analysis, and create a scientific platform

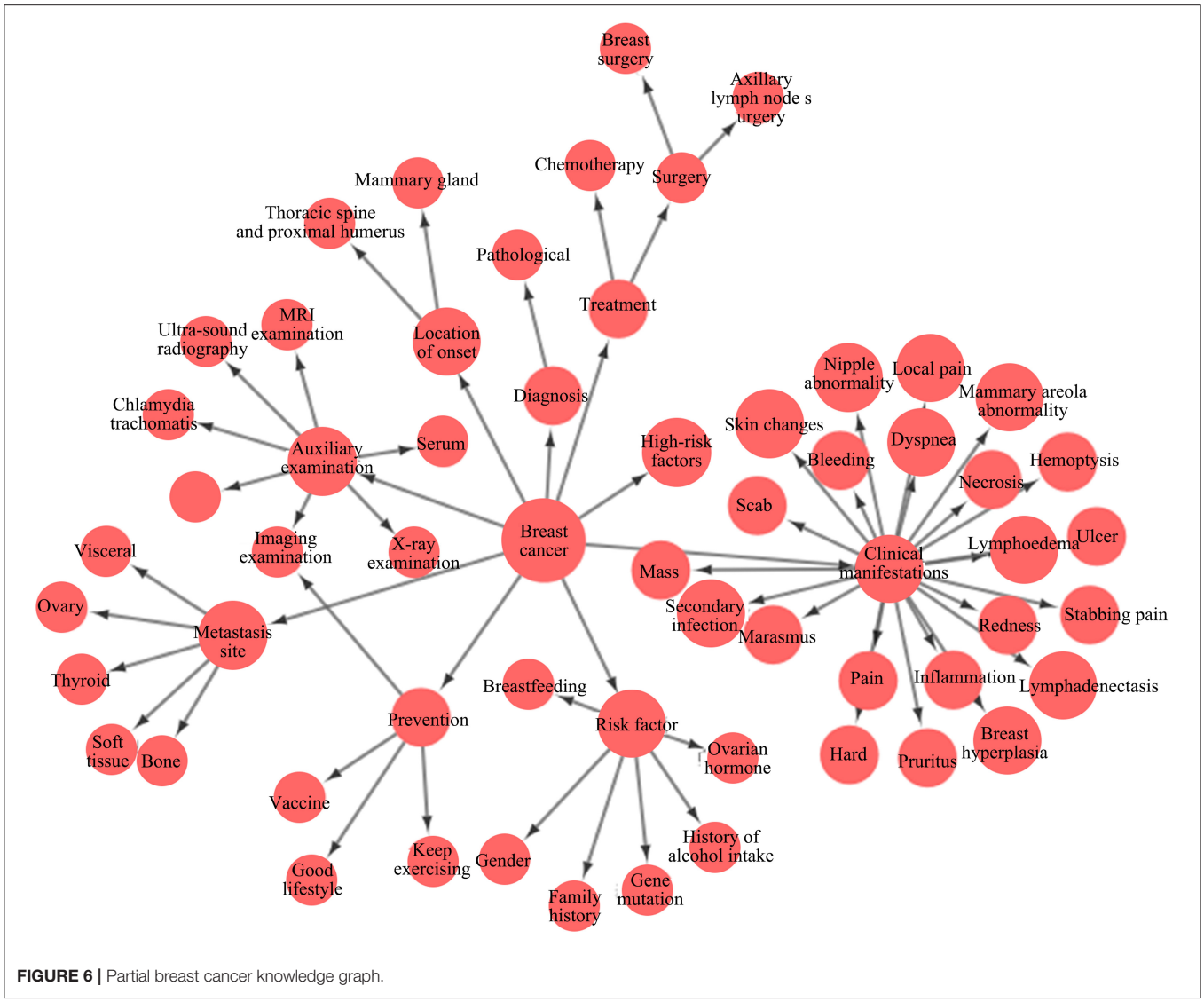


FIGURE 6 | Partial breast cancer knowledge graph.

for research on cancer pathogenesis and etiology. Big data and AI technology are utilized to provide continuous help for single disease of breast cancer before, during and after surgery, enable the physicians to deeply participate in the whole path of disease diagnosis and treatment, truly achieve accurate diagnosis and treatment planning, and break the data barriers between clinical departments. The governance and application of image data are emphasized to explore the image optimization algorithm and image recognition tool through database feedback and cyclic iteration optimization. The occurrence and development rules of relevant diseases are analyzed based on population categories to provide big data-based analysis and services for better clinical diagnosis and treatment, health management and clinical evidence-based medical research. Specialized research and disease-specific database are the focus of the connotation construction of the Hospital. The comprehensive hospitals in China can win a competitive advantage only by strengthening the construction of disciplines and also better meet the health service requirements of society and the country.

DATA AVAILABILITY STATEMENT

The datasets for this study are available from the corresponding author on reasonable request. Requests to access these datasets should be directed to yinjin@wchscu.cn.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the biomedical research ethics committee of West China Hospital (reference number: 20200427). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

CX and YJ were responsible for the study's concept. YJ, WJ, JJ, and SY were responsible for data and project management. YJ and WJ performed the data cleaning and database construction. YJ, WJ, JJ, SY, QK, and CX drafted and revised the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

FUNDING

This work was supported by the Science and Technology Department of Sichuan Province (No. 2020YFG0034).

ACKNOWLEDGMENTS

We thank the team members of the West China Biomedical Big Data Center for their support.

REFERENCES

1. Zeng N, Wang Z, Zineddin B, Li Y, Du M, Xiao L, et al. Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach. *IEEE Trans Med Imaging.* (2014) 33:1129–36. doi: 10.1109/TMI.2014.2305394
2. Lu MY, Chen TY, Williamson DF, Zhao M, Shady M, Lipkova J, Mahmood F. AI-based pathology predicts origins for cancers of unknown primary. *Nature.* (2021):1–5. doi: 10.1038/s41586-021-03512-4
3. Zeng N, Wang Z, Zhang H, Kim K-E, Li Y, Liu X. An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips. *#N/A.* (2019) 18:819–29. doi: 10.1109/TNANO.2019.2932271
4. Qiu J, Peng S, Yin J, Wang J, Jiang J, Li Z, et al. Radiomics Signature to Quantitatively Analyze COVID-19-Infected Pulmonary Lesions. *Interdiscip Sci.* (2021) 13:61–72. doi: 10.1007/s12539-020-00410-7
5. Qiu JJ, Yin J, Qian W, Liu JH, Huang ZX Yu HP, Ji L, et al. Novel multiresolution-statistical texture analysis architecture: radiomics-aided diagnosis of PDAC based on plain CT images. *IEEE Trans Med Imaging.* (2021) 40:12–25. doi: 10.1109/TMI.2020.3021254
6. Ullah MF. Breast cancer: current perspectives on the disease status. *Breast Cancer Metastasis and Drug Resistance. (*2019) 51–64. doi: 10.1007/978-3-030-20301-6_4
7. Issa NT, Byers SW, Dakshanamurthy S. Big data: the next frontier for innovation in therapeutics and healthcare. *Expert Rev Clin Pharmacol.* (2014) 7:293–8. doi: 10.1586/17512433.2014.905201
8. Li Q, Lan L, Zeng N, You L, Yin J, Zhou X, et al. framework for big data governance to advance RHINs: a case study of China. *IEEE Access.* (2019) 7:50330–8. doi: 10.1109/ACCESS.2019.2910838
9. Liu P, Fu B, Yang SX, Deng L, Zhong X, Zheng H. Optimizing survival analysis of xgboost for ties to predict disease progression of breast cancer. *IEEE Trans Biomed Eng.* (2020) 99:1. doi: 10.1109/TBME.2020.2993278
10. Zhong X, Luo T, Deng L, Liu P, Hu K, Lu D, et al. Multidimensional Machine Learning Personalized Prognostic Model in an Early Invasive Breast Cancer Population-Based Cohort in China: algorithm Validation Study. *JMIR Medical Informatics.* (2020) 8:e19069. doi: 10.2196/19069
11. Yuan J, Tang Q, Ma Q. Study on the construction of thymus adenoma structured special disease database. *China Medical Devices.* (2019) 34:113–5.
12. IRAC. *Ervik: "CanReg5–The Handbook."* Available online at: http://www.iacr.com.fr/CanReg5/ (accessed June 2021).
13. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (ehr) analysis. *IEEE Journal of Biomedical and Health Informatics.* (2018) 22:1589–604. doi: 10.1109/JBHI.2017.2767063
14. Ai L. *Design and Implementation of Multi-source Heterogeneous Healthcare Big Data Governance Platform*. Jinan: Shandong University (2020).
15. Wen-wen R, Gang W, Qi-li Z. Discussion on value of medical recordsstructured specialized disease database based on artificial intelligence in clinical research. *J Shanghai Jiaotong Univ.* (2020) 40:995. doi: 10.3969/j.issn.1674-8115.2020.07.022.
16. Seneviratne O, Rashid SM, Chari S, McCusker JP, Bennett KP, Hendler JA, et al. Knowledge integration for disease characterization: a breast cancer example. In: *Paper Presented at: International Semantic Web Conference*. (2018). p. 223–38. doi: 10.1007/978-3-030-00668-6_14
17. Wang Q, Mao Z, Wang B, Guo L. Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans Knowl Data Eng.* (2017) 29:2724–43. doi: 10.1109/TKDE.2017.2754499
18. Kannan AV, Fradkin D, Akrotirianakis I, Kulahcioglu T, Canedo A, Roy A, et al. Multimodal knowledge graph for deep learning papers and code. *Paper presented at: Proceedings of the 29th ACM International Conference on Information & Knowledge Management.* (2020). doi: 10.1145/3340531.3417439
19. Yang Q, Liu Y, Cheng Y, Kang Y, Chen T, Yu H. Federated learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning.* (2019) 13:1–207. doi: 10.2200/S00960ED2V01Y201910AIM043
20. Zeng N, Wang Z, Liu W, Zhang H, Hone K, Liu X, et al. A Dynamic neighborhood-based switching particle swarm optimization algorithm. *IEEE Trans Cybern.* (2020) 99:1–12. doi: 10.1109/TCYB.2020.3029748
21. Binder A, Bockmayr M, Hägele M, Wienert S, Heim D, Hellweg K, et al. Morphological and molecular breast cancer profiling through explainable machine learning. *Nature Machine Intelligence.* (2021) 3:355–66. doi: 10.1038/s42256-021-00303-4
22. Parekh VS, Jacobs MA. Integrated radiomic framework for breast cancer and tumor biology using advanced machine learning and multiparametric MRI. *NPJ breast cancer.* (2017) 3:1–9. doi: 10.1038/s41523-017-0045-3
23. Luo H, Xue W, Qiao S. Construction of disease database for scientific research in a hospital under big data environment. *Acad J PLA Med School.* (2019) 40:713–8. doi: 10.3969/j.issn.2095-5227.2019.08.003.

Conflict of Interest: CX is employed by Chengdu Zhixin Electronic Technology Co., Ltd.

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 Computer Science

TECHNOLOGY AND CODE
published: 09 August 2021
doi: 10.3389/fcomp.2021.684521

Multi_Scale_Tools: A Python Library to Exploit Multi-Scale Whole Slide Images

Niccolò Marini[1,2], Sebastian Otálora[1,2], Damian Podareanu[3], Mart van Rijthoven[4], Jeroen van der Laak[4,5], Francesco Ciompi[4], Henning Müller[1,6] and Manfredo Atzori[1,7]*

[1]Information Systems Institute, University of Applied Sciences Western Switzerland (HES-SO Valais), Sierre, Switzerland, [2]Centre Universitaire d'Informatique, University of Geneva, Carouge, Switzerland, [3]SURFsara, Amsterdam, Netherlands, [4]Department of Pathology, Radboud University Medical Center, Nijmegen, Netherlands, [5]Center for Medical Image Science and Visualization, Linkoping University, Linkoping, Sweden, [6]Medical Faculty, University of Geneva, Geneva, Switzerland, [7]Department of Neurosciences, University of Padua, Padua, Italy

Algorithms proposed in computational pathology can allow to automatically analyze digitized tissue samples of histopathological images to help diagnosing diseases. Tissue samples are scanned at a high-resolution and usually saved as images with several magnification levels, namely whole slide images (WSIs). Convolutional neural networks (CNNs) represent the state-of-the-art computer vision methods targeting the analysis of histopathology images, aiming for detection, classification and segmentation. However, the development of CNNs that work with multi-scale images such as WSIs is still an open challenge. The image characteristics and the CNN properties impose architecture designs that are not trivial. Therefore, single scale CNN architectures are still often used. This paper presents Multi_Scale_Tools, a library aiming to facilitate exploiting the multi-scale structure of WSIs. Multi_Scale_Tools currently include four components: a pre-processing component, a scale detector, a multi-scale CNN for classification and a multi-scale CNN for segmentation of the images. The pre-processing component includes methods to extract patches at several magnification levels. The scale detector allows to identify the magnification level of images that do not contain this information, such as images from the scientific literature. The multi-scale CNNs are trained combining features and predictions that originate from different magnification levels. The components are developed using private datasets, including colon and breast cancer tissue samples. They are tested on private and public external data sources, such as The Cancer Genome Atlas (TCGA). The results of the library demonstrate its effectiveness and applicability. The scale detector accurately predicts multiple levels of image magnification and generalizes well to independent external data. The multi-scale CNNs outperform the single-magnification CNN for both classification and segmentation tasks. The code is developed in Python and it will be made publicly available upon publication. It aims to be easy to use and easy to be improved with additional functions.

Keywords: multi-scale approaches, computational pathology, scale detection, classification, segmentation, deep learning

OPEN ACCESS

Edited by:
Nianyin Zeng,
Xiamen University, China

Reviewed by:
Heimo Müller,
Medical University of Graz, Austria
Han Li,
Xiamen University, China

****Correspondence:***
Niccolò Marini
niccolo.marini@hevs.ch

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Computer Science

Received: *23 March 2021*
Accepted: *07 July 2021*
Published: *09 August 2021*

Citation:
Marini N, Otálora S, Podareanu D, van Rijthoven M, van der Laak J, Ciompi F, Müller H and Atzori M (2021) Multi_Scale_Tools: A Python Library to Exploit Multi-Scale Whole Slide Images. Front. Comput. Sci. 3:684521. doi: 10.3389/fcomp.2021.684521

1 INTRODUCTION

The implicit multi-scale structure of digitized histopathological images represents an open challenge in computational pathology. Training machine learning algorithms that can simultaneously learn both microscopic and macroscopic tissue structures comes with technical and computational challenges that are not yet well studied.

As of 2021, histopathology represents the gold standard to diagnose many diseases, including cancer (Aeffner et al., 2017; Rorke, 1997). Histopathology images include several tissue structures, ranging from microscopic entities (such as single cell nuclei) to macroscopic components (such as tumor bulks). Whole Slide Images (WSIs) are digitized histopathology images that are scanned at high-resolution and are stored in a multi-scale (pyramidal) format. WSI resolution is related to the spatial resolution and the optical resolution used to scan the images (Wu et al., 2010). The spatial resolution is the minimum distance that the scanner can capture so that two objects are still distinguished, measured in terms of μm per pixel (Sellaro et al., 2013). The optical resolution (or magnification) is the magnification factor (x) of the lens within the scanner (Sellaro et al., 2013). Currently, the *de facto* standard spatial resolutions adopted to scan tissue samples (for example in The Cancer Genome Atlas) are usually 0.23–0.25 μm (magnification ×40) or 0.46–0.50 μm (magnification ×20). Tissue samples such as surgical resection samples (or specimens) are often approximately 20 mm × 15 mm in size[1], while samples such as biopsies are approximatively 2 mm × 6 mm in size. The size of the samples combined with the spatial resolution of the scanners leads to gigapixel images: image size can reach 200 000 × 200 000 pixels, meaning gigabytes of pixel data. The multi-scale WSI format (**Figure 1**) includes several magnification levels (with a different spatial resolution) of the sample, stored in a pyramid, usually varying between ×1.25 and 40x. The baseline image of the pyramid is the one at the highest resolution. The multi-scale structure of the images allows pathologists to analyze the image from the lowest to the highest magnification level. Pathologists analyze the images by first identifying a few regions of interest and zooming afterwards through them to visualize different details of the tissue (Schmitz et al., 2019). Each magnification level includes different types of information (Molin et al., 2016), since tissue structures appear in different ways according to their magnification level. Therefore, it is essential to detect an abnormality and detect it in a specific range of levels. The characteristics of microscopes and scanners often lead to a scale-dependent analysis. For example, at middle magnification levels (such as 5–10x) it is possible to distinguish between glands, while at the highest ones (such as 20–40x) it is possible to better resolve cells. **Figure 2** includes examples of tissues scanned at different magnification levels.

Computational pathology is the computational analysis of digital images obtained through scanning slides of cells and tissues (van der Laak et al., 2021). Currently, deep Convolutional Neural Networks (CNNs) are the state-of-the-art machine learning algorithms in computational pathology tasks, in particular for classification (del Toro et al., 2017; Arvaniti and Claassen, 2018; Coudray et al., 2018; Komura and Ishikawa, 2018; Ren et al., 2018; Campanella et al., 2019; Roy et al., 2019; Iizuka et al., 2020) and segmentation (Ronneberger et al., 2015; Paramanandam et al., 2016; Naylor et al., 2017; Naylor et al., 2018; Wang et al., 2019) of images. Their success relies on automatically learning the relevant

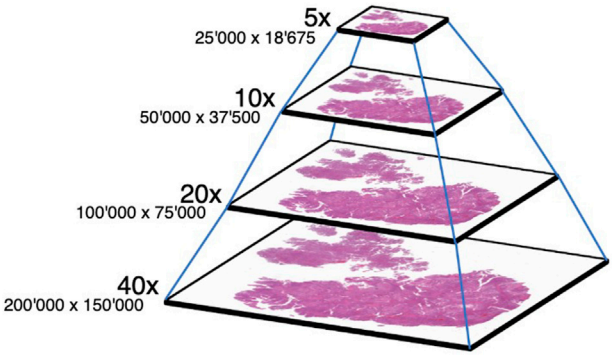


FIGURE 1 | An example of WSI format including multiple magnification levels. The size of each image of the pyramid is reported under the magnification level in terms of pixels.

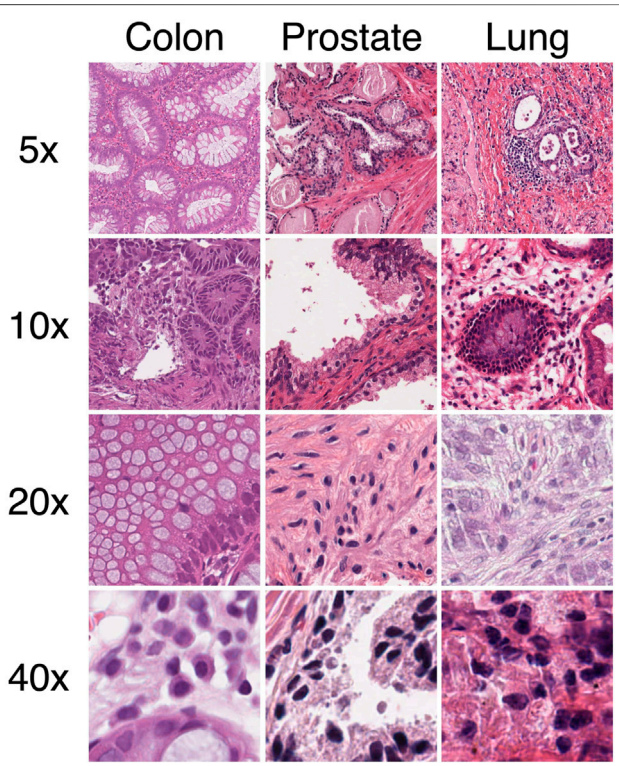


FIGURE 2 | An example of tissue represented at multiple magnification level (5x, 10x, 20x, 40x). The tissues come from colon, prostate and lung cancer images.

[1]http://dicom.nema.org/Dicom/DICOMWSI/. Retrieved 13th of November, 2020

features from the input data. However, usually, CNNs cannot easily handle the multi-scale structure of the images since they are not scale-equivariant by design (Marcos et al., 2018; Zhu et al., 2019) and because of WSI size. The equivariance property of a transformation means that when a transformation is applied, it is possible to predict how the representation will change (Lenc and Vedaldi, 2015; Tensmeyer and Martinez, 2016). This is not normally true for CNNs, because if a scale transformation is applied to the input data, it is usually not possible to predict its effect on the output of the CNN. The knowledge about the scale is essential for the model to identify diseases, since the same tissue structures, represented at different scales, include different information (Janowczyk and Madabhushi, 2016). CNNs can identify abnormalities in tissues, but the information and the features related to the abnormalities are not the same for each scale representation (Jimenez-del Toro et al., 2017). Therefore, the proper scale must be selected to train CNNs (Gecer et al., 2018; Otálora et al., 2018b). Unfortunately, scale information is not always available into images. This is the case, for instance, of pictures taken with standard cameras or processed in compression and resolution, such as images downloaded from the web or images included in scientific articles. Furthermore, modern hardware (Graphic Processing Units, GPUs) cannot easily handle WSIs, due to their large pixel size and the limited video random access memory space that has to temporally store it. The combination of different magnification levels leads to larger images, making it even harder to analyze the images.

The characteristics of the WSIs can lead to modification of CNNs in terms of architecture, both for classification (Jimenez-del Toro et al., 2017; Lai and Deng, 2017; Gecer et al., 2018; Yang et al., 2019; Hashimoto et al., 2020) and segmentation (Ronneberger et al., 2015; Li et al., 2017; Salvi and Molinari, 2018; Schmitz et al., 2019; van Rijthoven et al., 2020), such as multi-brach networks (Yang et al., 2019; Hashimoto et al., 2020; Jain and Massoud, 2020), multiple receptive fields convolutional neural networks (Han et al., 2017; Lai and Deng, 2017; Ullah, 2017; Li et al., 2019; Zhang et al., 2020) and U-Net based networks (Bozkurt et al., 2018; van Rijthoven et al., 2020). The modification of architectures to include multiple scales is prevalent in medical imaging, since it can allow to identify examples of architecture's modifications also from other modalities, such as MRI imaging (Zeng et al., 2021a) and Gold immunochromatographic strip (GIGS) images (Zeng et al., 2019; Zeng et al., 2021b).

The code library (called Multi_Scale_Tools) described in this paper contributes to alleviate the mentioned problems by presenting tools that allow handling and exploiting histopathological images' multi-scale structure end-to-end CNN architectures. The library includes pre-processing tools to extract multi-scale patches, a scale detector, a component to train a multi-scale CNN classifier and a component to train a multi-scale CNN for segmentation. The tools are platform-independent and developed in Python. The code is publicly available at https://github.com/sara-nl/multi-scale-tools. Multi_Scale_Tools is aimed at being easy to use and easy to be improved with additional functions.

2 METHODS

The library includes four components: a pre-processing tool, a scale detector tool, a component to train a multi-scale CNN classifier and a component to train a multi-scale segmentation CNN. Each tool is described in a dedicated subsection as follows:

- Pre-processing component, Sub-**section 2.1**
- Scale detector, Sub-**section 2.2**
- Multi-scale CNN for classification, Sub-**section 2.3**
- Multi-scale CNN for segmentation, Sub-**section 2.4**

2.1 Pre-Processing Component

The pre-processing component allows researchers to generate multi-scale input data. The component includes two parametric and scalable methods to extract patches from the different magnification levels of a WSI: the *grid* extraction and the *multi – center* extraction method. Both methods need a WSI and the corresponding tissue mask as input, and they both produce images and metadata as output. The *grid* extraction methods (Patch_Extractor_Dense_Grid.py, Patch_Extractor_Dense_Grid_Strong_Labels.py), allow to extract patches from one magnification level (**Figure 3**). The tissue mask is split in a grid of patches according to the following parameters: magnification level, mask magnification, patch size, and stride between the patches. The output of the method is a set of patches selected according to the parameters. The *multi – center* extraction methods (Patch_Extractor_Dense_Centroids.py, Patch_Extractor_Dense_Centroids_Strong_Labels.py) allow to extract patches from multiple magnification levels. According to the user's highest magnification level, the tissue mask is split into a grid (as done in the functions previously described). The patches within this grid are called centroids. Each centroid is used to generate the coordinates for a patch at a lower magnification level, so that the latter includes the centroid (the patch at the highest magnification level) in its central section. The method's output is a set of tuples, each one including patches at different magnification levels (**Figure 4**). Compared with other patch extraction methods, such as the one presented in (Lu et al., 2021), this pre-processing component has two main characteristics. The first one is that the component extracts patches from multiple magnification levels of the WSIs, pairing the patches coming from the same region of the image. The second one is that the component allows extracting patches from an arbitrary magnification level, despite the magnification level not being included in the WSI. Usually, patch extractor methods extract patches only from the magnification levels stored in the WSI format (M_a), such as 40x, 20x, 10x, 5x, 2.5x and 1.25x. This process is driven by the input parameters that include both the patch size (P_w) and the magnification wanted (M_w). The method extracts a patch of size P_a from a magnification stored in the WSI and afterwards the patch is resized to P_w.

$$P_w : M_w = P_a : M_a. \tag{1}$$

In both methods, only patches from tissue regions are extracted and saved using tissue masks, distinguishing between

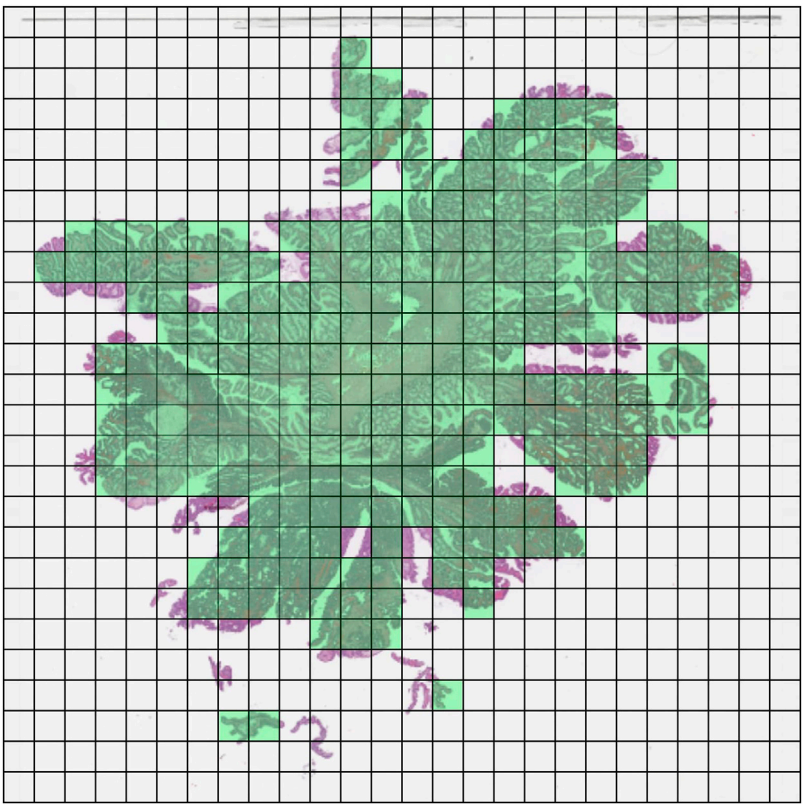


FIGURE 3 | An example of the *grid* extraction method. The patches in green are selected since they contain enough tissue.

patches from tissue regions and patches from the background. The methods are developed to work with masks including tissue and, in case they are available, with pixel-wise annotated masks. In the case of tissue masks, the tissue masks are generated using HistoQC tools (Janowczyk et al., 2019). The HistoQC configuration adopted is reported in the repository. In the case of pixel-wise annotations, the masks must be firstly converted to a RGB image.

Besides the patches, the methods save also metadata file (csv files). The metadata includes information regarding the magnification level where the patches are extracted and the x and y coordinates of the patches' upper left corner. The scripts are developed to be multi-thread, in order to exploit hardware architectures with multiple cores. In the **Supplementary Materials** section, the parameters for the scripts are described in more detail.

2.2 Scale Detector

The scale detector tool is a CNN trained to estimate the magnification level of a given patch or image. This task has been explored in the past Otálora et al. (2018a), Otálora et al. (2018b) in the prostate and breast tissue types. Similar approaches have been recently extended to different organs in the TCGA repository Zaveri et al. (2020). The tool involves the scripts related to the training of the models (the input data generation, the training and testing modules) and a module to use the detector as a standalone component that performs the magnification inference for new images. The models are trained in a fully-supervised fashion. Therefore, the scripts to train them need a set of patches and the corresponding magnification level as input, which are provided into csv files, including the patch path and the corresponding magnification levels. Two scripts are developed to generate the input files, assuming that the patches are previously generated with the pre-processing components, described in the previous section. The first script is made to split the WSIs into partitions (Create_csv_from_partitions.py), which generates three files (the input data for training, validation and testing partitions) starting from three files (previously prepared by the user) including the names of the WSIs. The second script (Create_csv.py) generates an input data csv starting from a list of files. The model is trained (Train_regressor.py) and tested (Test_regressor.py) with several magnification levels that the user can choose (in this paper, 5x, 8x, 10x, 15x, 20x, 30x, 40x were used). Training the model with patches from a discrete and small set of scales can lead to regressors that are precise to estimate the magnifications close to input scales, and less precise when scales are far from them. Therefore, a scale augmentation technique was applied to patches and labels during the training (in addition to more standard augmentation techniques adopted such as rotation, flipping and color augmentation). In order to perform scale augmentation, the image is randomly cropped of a factor and resized to the original patch size. The factor is applied to perturbate also the magnification level. The scale

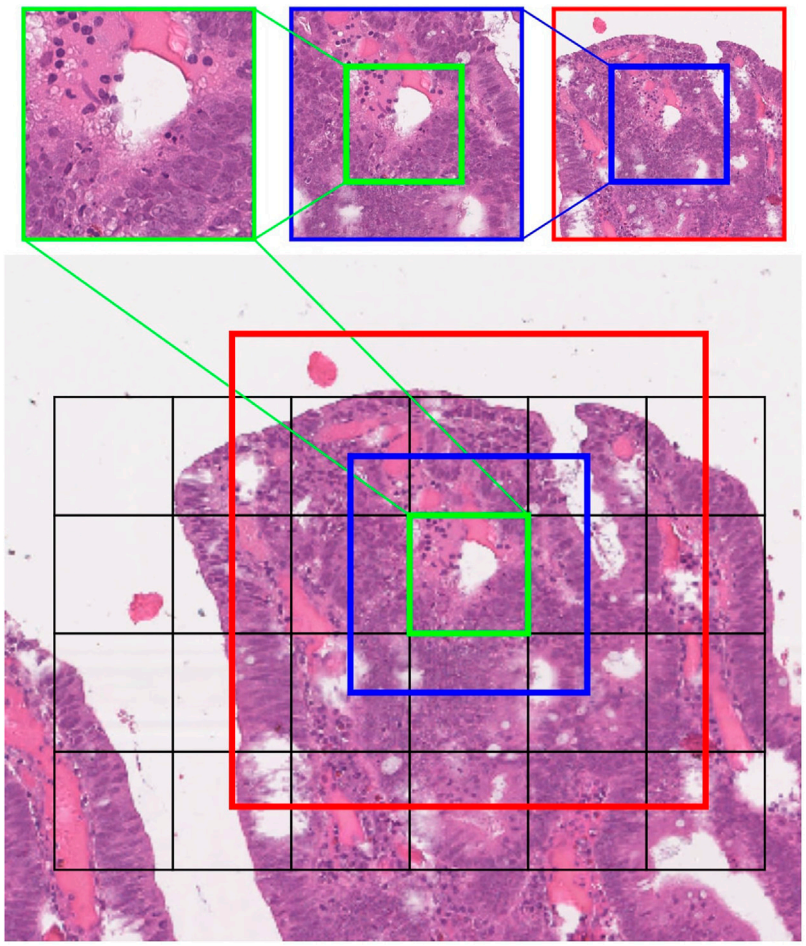


FIGURE 4 | An example of the *multi – center* extraction method. The grid is made according to the highest magnification level selected by the used. The patch is the centroid for patches at lower magnification levels.

detector component includes also a module to import and use the model in the code (regression.py). The component works both as a standalone module (with the required parameters) but it is also possible to load the functions from the python module. The **Supplementary Materials** section includes a more thorough description of the parameters for the scripts.

2.3 Multi-Scale CNN for Classification

The Multi-scale CNN component includes scripts to train a multi-scale CNN for classification, in a fully supervised fashion. Two different multi-scale CNN architectures and two training variants are proposed and compared with a single-scale CNN. The multi-scale CNN architectures are composed of multiple branches (one for each magnification level) trained with patches that come from several magnifications. Each branch is fed with patches from a specific magnification level. The first architecture of multi-scale CNN combines each CNN branch features (the output of the convolutional layers). The scripts developed to train and test the models are Fully_supervised_training_combine_features_multi.py and Fully_supervised_testing_combine_features_multi.py The second architecture of multi-scale CNN combines the classifier predictions (the output of each CNN's fully-connected layers). The scripts developed to train and test the models are Fully_supervised_training_combine_probs_multi.py and Fully_supervised_testing_combine_probs_multi.py Both architectures are presented in two variants, optimizing respectively one and multiple loss functions. In the first variant (one loss function), the input is a set of tuples of patches from several magnification levels (one patch for each level), generated using the *multi – center* extraction tool (presented in **Section 2.1**). The input tuples are generated with a script (Generate_csv_multicenter.py) that exploits the coordinates of the patches (stored in the metadata) to generate the tuples (stored in a csv file). The tuple label corresponds to the class of the centroid patch (the patch from the highest level within the tuple). Therefore, the model outputs only the class of the tuple. Only one loss function is minimized in this variant, i.e. the categorical cross-entropy between the CNN output and the patch ground truth. **Figure 5** summarizes the CNN architecture. In the second variant (multiple loss functions), the input is a set of tuples of patches from several magnification levels (one patch for each level), previously generated using the *grid* extraction method (presented in **Section 2.1**). The input tuples are generated with a script

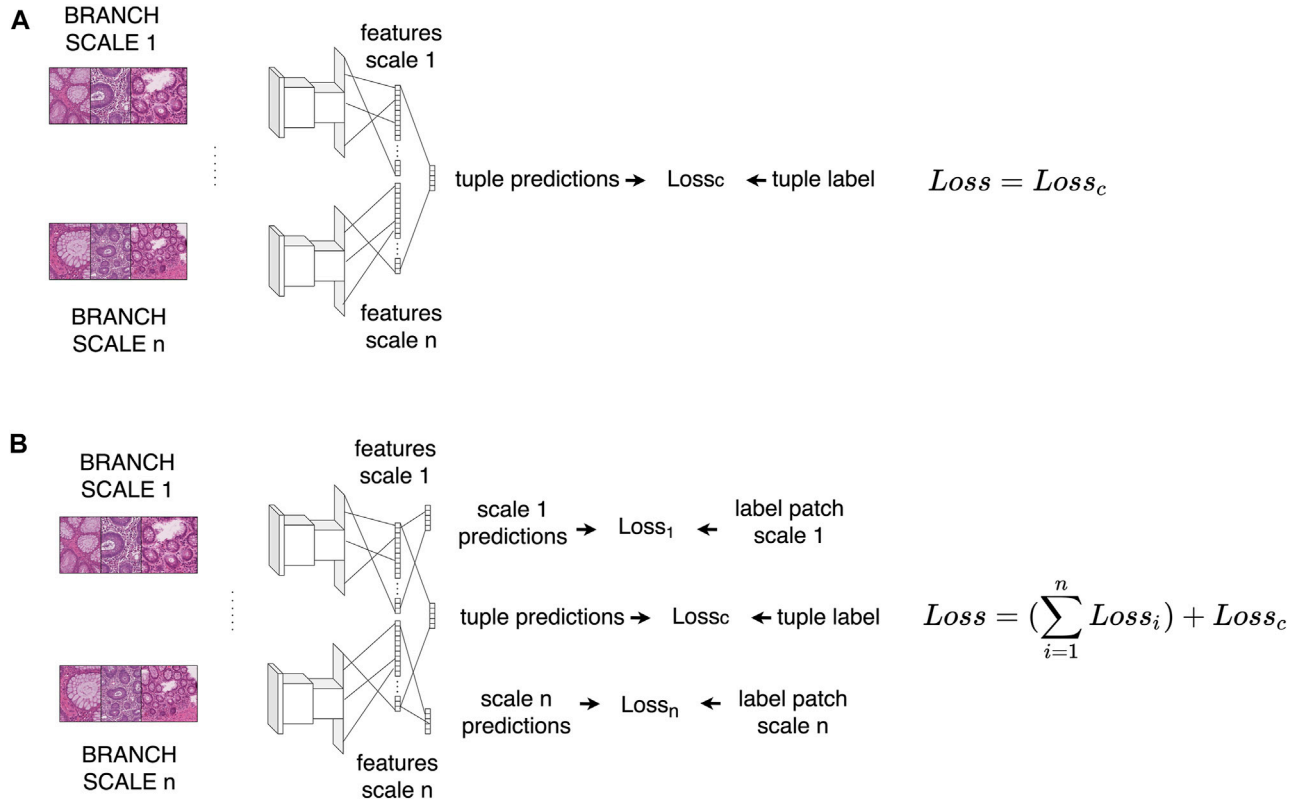


FIGURE 5 | The first multi-scale CNN architecture, in which features are combined from different scale branches, optimizing only one loss function **(A)** and optimizing $n + 1$ loss function **(B)**.

(Generate_csv_upper.py) that exploits the coordinates of the patches (stored in the metadata) to generate the tuples (stored in a csv file). The tuple labels correspond to the classes of the patches. The model has $n + 1$ outputs: the class for each of the n magnification levels and the whole tuple class. In this variant, $n + 1$ loss functions are minimized (n representing the number of magnification levels considered). The n loss functions are the categorical cross-entropy between the output for each of the scale branches and the tuple labels. The other loss term is the categorical cross-entropy between the output of the network (after the combination of the features or the predictions of the single branches) and the tuple labels. **Figure 6** summarizes the CNN architecture. The Supplementary Materials section includes a more thorough description of the parameters.

2.4 Multi-Scale CNN for Segmentation

This component includes HookNet (van Rijthoven et al., 2020), a multi-scale CNN for semantic segmentation. HookNet combines information from low-resolution patches (large field of view) and high-resolution patches (small field of view) to semantically segment the image, using multiple branches. The low-resolution patches come from lower magnification levels and include context information, while the high-resolution patches come from higher magnification levels and include more fine-grained information. The network is composed of two branches of encoder-decoder models, the context branch (fed with low-resolution patches) and the target branch (fed with high-resolution patches). The two branches are fed with concentric multi-field-view multi-resolution (MFMR) patches (284 × 284 pixels in size). Although they have the same architecture, the branches do not share their weights (an encoder-decoder CNN based on U-Net). Hooknet is thoroughly described in a dedicated article (van Rijthoven et al., 2020).

2.5 Datasets

The following datasets are used to develop the Multi_Scale_Tools components:

- Colon dataset, Sub-**section 2.5.1**, used in the Pre-processing component, the Scale detector and the Multi-scale CNN for classification
- Breast dataset, Sub-**section 2.5.2**, used in the Multi-scale CNN for segmentation
- Prostate dataset, Sub-**section 2.5.3**, used in the Scale detector
- Lung dataset, Sub-**section 2.5.4**, used in the Scale detector and the Multi-scale CNN for segmentation

2.5.1 Colon Dataset

The colon dataset is a subset of the ExaMode colon dataset. This subset includes 148 WSIs (provided by the Department of Pathology of Cannizaro Hospital, Catania, Italy), stained with

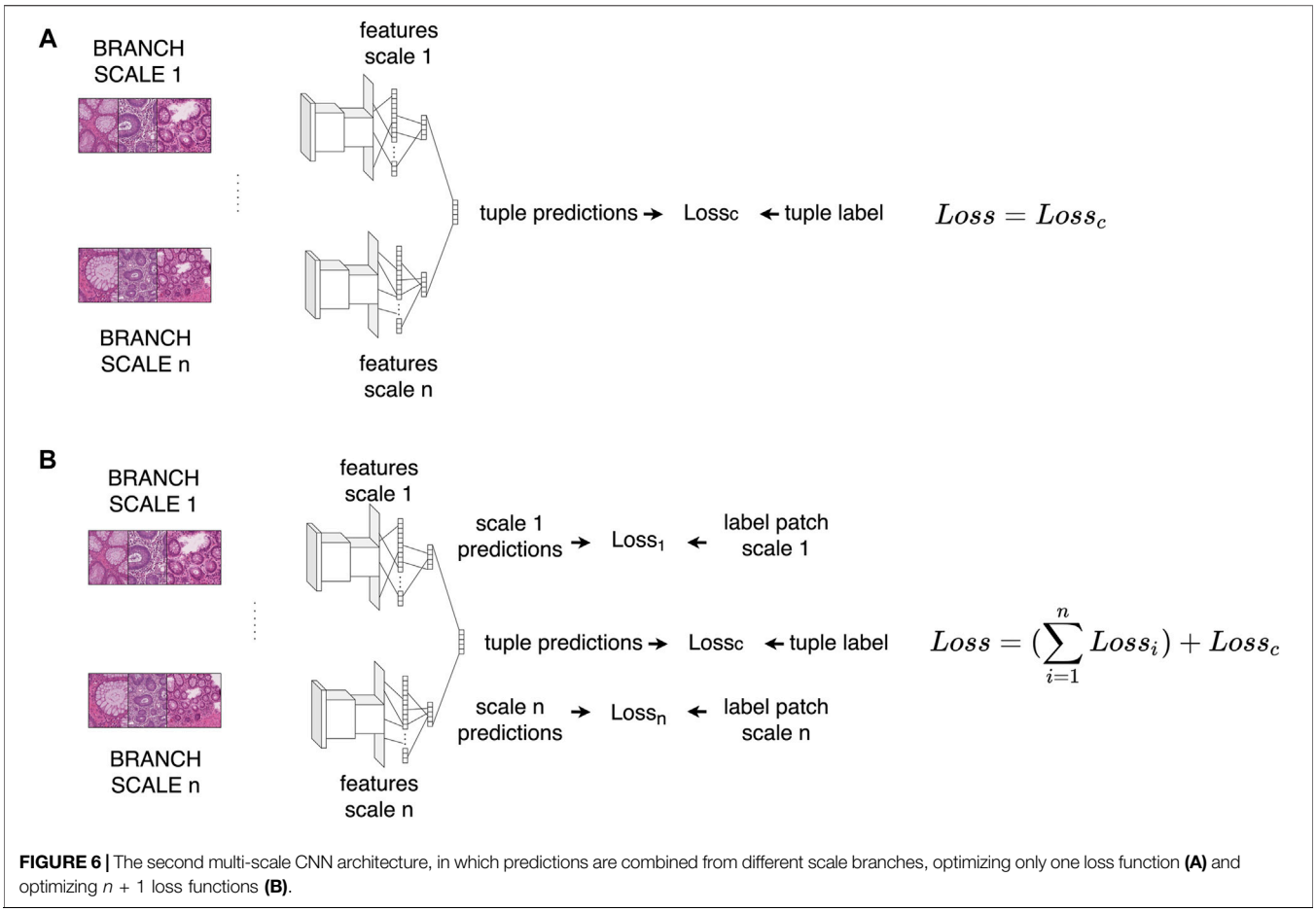


FIGURE 6 | The second multi-scale CNN architecture, in which predictions are combined from different scale branches, optimizing only one loss function **(A)** and optimizing $n + 1$ loss functions **(B)**.

Hematoxylin and Eosin (H&E). The images are digitized with an Aperio scanner: some of the images are scanned with a maximum spatial resolution of 0.50 µm per pixel (20x), while the others are scanned with a spatial resolution of 0.25 µm per pixel (40x). The images are pixel-wise annotated by a pathologist. The annotations include five classes: cancer, high-grade dysplasia, low-grade dysplasia, hyperplastic polyp and non-informative tissue.

2.5.2 Breast Dataset

The breast dataset (provided by Department of Pathology of Radboud University Medical Center, Nijmegen, Netherlands) is a private dataset including 86 WSIs, stained with H&E. The images are digitized with a 3DHistech scanner, with a spatial resolution of 0.25 µm per pixel (40x). The images are pixel-wise annotated by a pathologist. 6,279 regions are annotated, with the following classes: ductal carcinoma *in-situ* (DCIS), invasive ductal carcinoma (IDC), invasive lobular carcinoma (ILC) benign epithelium (BE), other, and fat.

2.5.3 Prostate Dataset

The prostate dataset is a subset of the publicly available database offered by The Cancer Genome Atlas (TCGA-PRAD), that includes 20 WSIs, stained with H&E. The images come from several sources and are digitized with different scanners, with a spatial resolution of 0.25 µm per pixel (40x). The images come without pixel-wise annotations.

2.5.4 Lung Dataset

The Lung dataset is a subset of the public available database offered by The Cancer Genome Atlas Lung Squamous Cell carcinoma dataset (TCGA-LUSC), including 27 WSIs stained with H&E. The images come from several sources and are digitized with different scanners, with a spatial resolution of 0.25 µm per pixel (40x). Initially, the images come without pixel-wise annotation from the repository, but a medical expert from Radboudc Hospital pixel-wise annotated them with four classes: tertiary lymphoid structures (TLS), germinal centers (GC), tumor, and other.

3 EXPERIMENTS AND RESULTS

The Section presents the assessment of the components of the library Multi_Scale_Tools in dedicated subsections as follows:

- Pre-processing component assessment, Sub-**section 3.1**
- Scale detector assessment, Sub-**section 3.2**
- Multi-scale CNN for classification assessment, Sub-**section 3.3**

TABLE 1 | Number of patches extracted with the grid extraction method (above) and with the multi-center method (below), at different magnification level.

Grid					
Dataset/Magnification	5x	10x	20x	40x	Total
Colon dataset (148 WSIs)	15,514	67,592	279,964	1,127,190	1,490,260
Prostate dataset (20 WSIs)	11,468	46,676	187,254	743,583	988,981
Lung dataset (27 WSIs)	22,124	90,307	365,398	886,298	1,364,127

Multicenter			
Dataset/Magnification	5x/10x	5x/10x/20x	Total
Colon dataset	135,184	839,892	975,076
Prostate dataset	93,352	561,762	655,114
Lung dataset	180,614	1,096,194	1,276,808

TABLE 2 | Time needed to extract the patches (in seconds), varying the amount of threads, using the *grid* extraction method (above) and using the *multi – center* method (below). The method is evaluated on colon dataset (148 WSIs). The number of patches extracted from each method is reported in **Table 1**.

Magnification/N_threads	10	20	30	40	50
grid extraction					
5x	408 ± 3	317 ± 7	285 ± 5	255s ± 5	238 ± 10
10x	553 ± 5	429 ± 6	389 ± 5	384 ± 8	371 ± 8
20x	1,295 ± 100	969 ± 113	876 ± 69	872 ± 41	869 ± 15
multi – center extraction					
5x/10x	1,662 ± 30	1,180 ± 119	1,071 ± 50	1,039 ± 18	1,022 ± 14
5x/10x/20x	6,604 ± 104	5,745 ± 45	5,283 ± 161	4,814 ± 137	4,549 ± 82

- Multi-scale CNN for segmentation, Sub-**section 3.4**
- Library organization, Sub-**section 3.5**

3.1 Pre-Processing Tool Assessment

The pre-processing component allows to extract a large amount of patches from multiple magnification levels, guaranteeing scalable performance. The patch extractor components (grid and multi-center methods) are tested on WSIs scanned with Aperio (.svs), 3DStech (.mrxs) and Hamamatsu (.ndpi) scanners, on data coming from different tissues (colon, prostate and lung) and datasets. **Table 1** includes the number of patches extracted. The upper part of the table includes the number of patches extracted with the grid extraction method, considering four different magnification levels (5x, 10x, 20x, 40x). The lower part of the Table includes the number of patches extracted with the multi-center extraction method, considering two possible combinations of magnification levels (5x/10x, 5x/10x/20x). In both cases, patches are extracted with a patch size of 224 × 224 pixels without any stride. Methods performance are evaluated in terms of scalability, since the methods are designed to work on multi-core hardwares. **Table 2** includes the time results obtained with the grid method (upper part) and with the multi-center method (lower part). The evaluation is made considering the amount of time needed to extract the patches from the colon dataset, using several threads. The results show that both the methods benefit from multi-core hardwares, reducing the time needed to pre-process data.

3.2 Scale Detector Assessment

The scale detector shows high performance in estimating the magnification level of patches that come from different tissues. The detector is trained with patches from the colon dataset and it is tested with patches from three different tissues. The performance of the models is assessed with the coefficient of determination (R^2), the Mean Square Error (MSE), the Cohen's κ-score (McHugh, 2012) and the balanced accuracy. While the experimental setup and the metrics descriptions are presented in detail the supplementary material, **Table 3** summarizes the results. The higher performance is reached on the colon test partition, but the scale detector shows high performance also on the other tissues. The scale detector makes almost perfect scale estimations in the colon dataset (data come from the same medical source and include the same tissue type), in both the regression and the classification metrics. The scale detector makes reasonably good scale estimations also on the prostate data, in both the regression and the classification metrics, and in lung dataset, where the performance is the lowest though. The fact that the regressor shows exceptionally high performance in colon data and good performance in other tissues means that it has learnt to distinguish the colon morphology represented at different magnification level very well and that the learnt knowledge can generalize well to other tissues too. Even though tissues from different organs share similar structures (glands, stroma, etc.), the morphology of the structures is different in the organs, such as prostate and colon glands. Training the regressor with patches from several organs may allow to close this gap, guaranteeing extremely high performance for different types of tissue.

TABLE 3 | Performance of the scale detector, evaluated on three different tissue dataset. The scale detector is evaluated in: coefficient of determination (R^2), Mean squared error (MSE), balanced accuracy, Cohen's κ-score.

Dataset/Metric	R^2	MSE	Balanced accuracy	κ-score
Colon dataset	0.9997 ± 0.0001	0.0250 ± 0.0155	0.9859 ± 0.0086	0.9991 ± 0.0004
Prostate dataset	0.8013 ± 0.0798	19.34 ± 7.78	0.9094 ± 0.0268	0.8515 ± 0.0589
Lung dataset	0.6682 ± 0.1549	32.13 ± 15.01	0.7973 ± 0.0458	0.8743 ± 0.0571

TABLE 4 | Performance of the multi-scale CNNs architectures, compared with CNNs trained with patches from only one magnification level, evaluated in κ-score and balanced accuracy. Both the multi-scale architectures are presented (combine features and combine predictions from multi-scale branches) and both the training variants (one loss function and $n + 1$ losses). The values marked in bold highlight the method that reaches the highest performance, respect to the metric.

Magnification/metric	κ-score	Balanced-accuracy
Single scale CNNs		
5x	0.7127 ± 0.0988	0.6558 ± 0.0903
10x	0.6818 ± 0.0940	0.6200 ± 0.0780
20x	0.6005 ± 0.1106	0.5744 ± 0.0804
Multi-scale CNNs (combine features)		
5x/10x (One loss)	0.6955 ± 0.1013	0.6529 ± 0.0859
5x/10x ($n + 1$ losses)	**0.7167 ± 0.1060**	**0.6813 ± 0.0942**
5x/10x/20x (One loss)	0.6630 ± 0.1090	0.6508 ± 0.1089
5x/10x/20x ($n + 1$ losses)	0.6871 ± 0.1110	0.6364 ± 0.1046
Multi-scale CNNs (combine probabilities)		
5x/10x (One loss)	0.6901 ± 0.1136	0.6582 ± 0.0973
5x/10x ($n + 1$ losses)	0.7026 ± 0.0988	0.6626 ± 0.0897
5x/10x/20x (One loss)	0.6678 ± 0.0973	0.6239 ± 0.0860
5x/10x/20x ($n + 1$ losses)	0.6784 ± 0.0995	0.6355 ± 0.0835

3.3 Multi-Scale CNN for Classification Assessment

The multi-scale CNNs show higher performance in the fully supervised classification compared to the single-scale CNNs. Several configurations of the multi-scale CNN architectures are evaluated. They involve variations in optimization strategy (one or multiple loss functions), in the magnification levels (combinations of 5x, 10x, 20x) and in how information from the scales is combined (combining the single-scale predictions or the single-scale features). **Table 4** summarizes the results obtained. The CNNs are trained and tested with the colon dataset, that come with pixel-wise annotations made by a pathologist. The performance of the models is assessed with the Cohen's κ-score and the balanced accuracy. More detailed descriptions of the experimental setup and the metrics adopted are presented in the Supplementary material. In the presented experiment, the best multi-scale CNN architecture is the one that combines features from 5/10x magnification levels and is trained optimizing $n + 1$ loss functions. It outperforms the best single-scale CNN (trained with patches acquired at 5x) in terms of balanced accuracy, while the κ-score of the two architectures is comparable. The characteristics of the classes involved can explain the fact that CNNs trained combining patches from 5/10x reach the highest results. These classes show morphologies including several alterations of the gland structure. Glands can be usually identified at low magnification levels, such as 5/10x, while at 20x the cells are visible. For this reason, the CNNs show high performance with patches from magnification 5/10x, while including patches from 20x decreases the performance. The fact that the discriminant characteristics are identified in a range of scales may explain why the combination of the features shows higher performance than the combination of the predictions.

3.4 Multi-Scale CNN for Segmentation Assessment

The multi-scale CNN (HookNet) shows higher tissue segmentation performance than single-scale CNNs (U-Net). The model is trained and tested with breast and lung datasets, comparing it with models trained with images from a single magnification level. The performance of the models is assessed with the F1 score and the macro F1 score. More detailed descriptions of the experimental setup and the metrics adopted are presented in the **Supplementary Material**. **Table 5** and **Table 6** summarize the results obtained respectively on the breast dataset and on lung dataset. In both the tissues, HookNet shows an higher overall performance, while some of the single scale U-Nets have better performance for some segmentation tasks (such as breast DCIS or lung TLS). This result can be interpreted as a consequence of the characteristics of the task, therefore the user should choose the proper magnification levels to combine, depending of the problem.

3.5 Library Organization

The source code for the library is available on GIT[2], while the HookNet code is available here[3]. The library is available can be deployed as Python package directly from the repository or as Docker container that can be downloaded from[4] (the multiscale folder). Interaction with the library is done through a model class and an Inference class[5]. The model instantion depends on the choice of algorithms. For a more detailed explanation about the hyperparameters and other options please make sure to browse the Readme file[6]. An example can be found here[7]. The Python libraries used to develop Multi_Scale_Tools are reported in **Supplementary Materials**.

[2]https://github.com/sara-nl/multi-scale-tools. Retrieved 11th of January, 2021

[3]https://github.com/DIAGNijmegen/pathology-hooknet. Retrieved 19th of June, 2021

[4]https://surfdrive.surf.nl/files/index.php/s/PBBnjwzwMragAGd. Retrieved 11th of January, 2021

[5]https://github.com/computationalpathologygroup/hooknet/blob/fcba7824ed982f663789f0c617a4ed65bedebb85/source/inference.py#L20. Retrieved 11th of January, 2021

[6]https://github.com/sara-nl/multi-scale-tools/blob/master/README.md. Retrieved 11th of January, 2021

[7]https://github.com/DIAGNijmegen/pathology-hooknet/blob/master/hooknet/apply.py. Retrieved 19th of June, 2021

TABLE 5 | Performance of the U-Net (above) and HookNet (below) on the breast dataset. The architectures are compared on the F1 score, for each tissue type (description of the tissue type in the **Supplementary Material**). The overall macro-F1 score is reported. The values marked in bold highlight the method that reaches the highest performance, respect to the task.

Magnification	DCIS	IDC	ILC	Benign	Other	Fat	Overall
Model: U-Net							
20x	0.47	0.55	0.85	0.75	0.95	0.99	0.76
10x	0.67	0.69	0.79	**0.87**	0.98	1.00	0.83
5x	0.79	0.83	0.79	0.84	0.98	1.00	0.87
2.5x	0.83	0.85	0.63	0.73	0.96	1.00	0.83
1.25x	**0.86**	0.81	0.20	0.66	0.96	1.00	0.75
Model: HookNet							
20x (target)-5x (context)	0.62	0.75	0.82	0.82	0.98	1.00	0.83
20x (target)-1.25x (context)	0.84	**0.89**	**0.91**	0.84	0.98	1.00	**0.91**

TABLE 6 | Performance of the U-Net (above) and HookNet (below) on the lung dataset. The architectures are compared on the F1 score, for each tissue type (description of the tissue type in the **Supplementary Material**). The overall macro-F1 score is reported. The values marked in bold highlight the method that reaches the highest performance, respect to the task.

Magnification	TLS	GC	Tumor	Other	Overal
Model: U-Net					
20x	0.81	0.28	**0.75**	0.87	0.70
10x	**0.86**	0.44	0.71	0.86	0.72
5x	0.84	**0.49**	0.67	0.85	0.71
2.5x	0.80	0.37	0.56	0.80	0.63
1.25x	0.78	0.35	0.39	0.77	0.57
Model: HookNet					
20x (target)-5x (context)	0.84	0.48	0.72	0.87	**0.73**

4 DISCUSSION AND CONCLUSION

Multi_Scale_Tools library aims at facilitating the exploitation of multi-scale structure in WSIs with code that is easy to use and easy to be improved with additional functions. The library currently includes four components. The components are a pre-processing tool to extract multi-scale patches, a scale detector, two multi-scale CNNs for classification and a multi-scale CNN for segmentation. The pre-processing component includes two methods to extract patches from several magnification levels. The methods are designed to be scalable on multi-core hardware. The scale detector component includes a CNN allowing to regress the magnification level of a patch. The CNN obtains high performance in patches that come from the colon (the tissue used to train it) and it reaches good performance on other tissues such as prostate and lung too. Two multi-scale CNN architectures are developed for fully-supervised classification. The first one combines features from multi-scale branches, while the second one combines predictions from multi-scale branches. The first architecture obtains better performance and outperforms the model trained with patches from only one magnification level. The HookNet architecture for multi-scale segmentation is also included into the library, fostering its usage and making the library more complete. The tests show that HookNet outperforms single scale U-Net in the considered tasks. The presented library allows to exploit the multi-scale structure of WSIs efficiently. In any case, the user remains a fundamental part of the system for several components, such as identifying the scale that can be more relevant for a specific problem. The comparison between the single-scale CNNs and the multi-scale CNN is an example of this. The CNN is trained to classify between cancer, dysplasia (both high-grade and low-grade), hyperplastic polyp and non-informative tissue. In the classification task, the highest performance is reached using patches of magnification 5x and 10x, while patches from 20x lead to lower classification performance. This can likely be related to the fact that the main feature related to the considered classes is the structure of the glands, therefore high magnifications (e.g. 20x) limitedly introduce helpful information into the models. The importance of the user to select the proper magnification levels is highlighted even more in the segmentation results. Considering low magnifications, the models show good performance in ductal carcinoma *in-situ* and invasive ductal carcinoma segmentation since these tasks need context about the duct structures in the breast use case. Considering higher magnifications, the models perform well in invasive lobular carcinoma and benign tissue segmentation, where the details are more important. The methods identified to pair images from several magnification levels can pave the way to multi-modal combination of images too. The combination may increase the information included in the single modality, increasing the performance of the CNNs. Some possible applications can be the combination of WSIs stained with different reagents, such H&E and immunohistochemical (IHC) stainings, the application in Raman spectroscopy data, combining information about tissue morphologies and architectures with protein biomarkers, and the combination of patches from different focal planes.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

NM: design of the work, software, analysis, original draft SO: design of the work, revised the work DP: software, revised the

work MR: software, analysis JL: revised the work, approval for publication FC: revised the work, approval for publication HM: revised the work, approval for publication MA: revised the work, approval for publication.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825292 (ExaMode, http://www.examode.eu/).

ACKNOWLEDGMENTS

The authors also thank Nvidia for the Titan Xp GPUs used for some of the weakly supervised experiments. SO thanks to the Colombian science ministry Minciencias for partially funding his Ph.D. studies through the call "756-Doctorados en el exterior".

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2021.684521/full#supplementary-material

REFERENCES

Aeffner, F., Wilson, K., Martin, N. T., Black, J. C., Hendriks, C. L. L., Bolon, B., et al. (2017). The Gold Standard Paradox in Digital Image Analysis: Manual versus Automated Scoring as Ground Truth. *Arch. Pathol. Lab. Med.* 141, 1267–1275. doi:10.5858/arpa.2016-0386-ra

Arvaniti, E., and Claassen, M. (2018). *Coupling Weak and strong Supervision for Classification of Prostate Cancer Histopathology Images*. Medical Imaging meets NIPS Workshop, NIPS 2018. arXiv preprint arXiv:1811.07013.

Bozkurt, A., Kose, K., Alessi-Fox, C., Gill, M., Dy, J., Brooks, D., and Rajadhyaksha, M. (2018). A Multiresolution Convolutional Neural Network with Partial Label Training for Annotating Reflectance Confocal Microscopy Images of Skin. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018 (Springer), 292–299. doi:10.1007/978-3-030-00934-2_33

Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., 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

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

del Toro, O. J., Atzori, M., Otálora, S., Andersson, M., Eurén, K., Hedlund, M., et al. (2017). "Convolutional Neural Networks for an Automatic Classification of Prostate Tissue Slides with High-Grade gleason Score," in *Medical Imaging 2017: Digital Pathology* (Bellingham, WA: International Society for Optics and Photonics), 10140, 101400O. doi:10.1117/12.2255710

Gecer, B., Aksoy, S., Mercan, E., Shapiro, L. G., Weaver, D. L., and Elmore, J. G. (2018). Detection and Classification of Cancer in Whole Slide Breast Histopathology Images Using Deep Convolutional Networks. *Pattern recognition* 84, 345–356. doi:10.1016/j.patcog.2018.07.022

Han, D., Kim, J., and Kim, J. (2017). Deep Pyramidal Residual Networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, July 21-26 2017 (IEEE) 5927–5935. doi:10.1109/cvpr.2017.668

Hashimoto, N., Fukushima, D., Koga, R., Takagi, Y., Ko, K., Kohno, K., et al. (2020). Multi-scale Domain-Adversarial Multiple-Instance Cnn for Cancer Subtype Classification with Unannotated Histopathological Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14–19 June 2020 (IEEE) 3852–3861. doi:10.1109/cvpr42600.2020.00391

Iizuka, O., Kanavati, F., Kato, K., Rambeau, M., Arihiro, K., and Tsuneki, M. (2020). Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours. *Sci. Rep.* 10, 1504–1511. doi:10.1038/s41598-020-58467-9

Jain, M. S., and Massoud, T. F. (2020). Predicting Tumour Mutational burden from Histopathological Images Using Multiscale Deep Learning. *Nat. Mach Intell.* 2, 356–362. doi:10.1038/s42256-020-0190-5

Janowczyk, A., and Madabhushi, A. (2016). Deep Learning for Digital Pathology Image Analysis: A Comprehensive Tutorial with Selected Use Cases. *J. Pathol. Inform.* 7, 29. doi:10.4103/2153-3539.186902

Janowczyk, A., Zuo, R., Gilmore, H., Feldman, M., and Madabhushi, A. (2019). Histoqc: an Open-Source Quality Control Tool for Digital Pathology Slides. *JCO Clin. Cancer Inform.* 3, 1–7. doi:10.1200/cci.18.00157

Jimenez-del-Toro, O., Otálora, S., Andersson, M., Eurén, K., Hedlund, M., Rousson, M., et al. (2017). "Analysis of Histopathology Images," in *Biomedical Texture Analysis* (Cambridge, MA: Elsevier), 281–314. doi:10.1016/b978-0-12-812133-7.00010-7

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

Lai, Z., and Deng, H. (2017). Multiscale High-Level Feature Fusion for Histopathological Image Classification. *Comput. Math. Methods Med.* 2017, 7521846. doi:10.1155/2017/7521846

Lenc, K., and Vedaldi, A. (2015). Understanding Image Representations by Measuring Their Equivariance and Equivalence. In Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, USA, 7–12 June 2015 (IEEE) 991–999. doi:10.1109/cvpr.2015.7298701

Li, J., Sarma, K. V., Chung Ho, K., Gertych, A., Knudsen, B. S., and Arnold, C. W. (2017). A Multi-Scale U-Net for Semantic Segmentation of Histological Images from Radical Prostatectomies, *AMIA Annu. Symp. Proc.*. In AMIA Annual Symposium Proceedings, Washington, DC, 4–8 November 2017 (American Medical Informatics Association), vol. 2017, 1140, 1148.

Li, S., Liu, Y., Sui, X., Chen, C., Tjio, G., Ting, D. S. W., and Goh, R. S. M. (2019). Multi-instance Multi-Scale Cnn for Medical Image Classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019 (Springer), 531–539. doi:10.1007/978-3-030-32251-9_58

Lu, M. Y., Williamson, D. F., Chen, T. Y., Chen, R. J., Barbieri, M., and Mahmood, F. (2021). Data-efficient and Weakly Supervised Computational Pathology on Whole-Slide Images. *Nat. Biomed. Eng.*, 1–16. doi:10.1038/s41551-020-00682-w

Marcos, D., Kellenberger, B., Lobry, S., and Tuia, D. (2018). *Scale Equivariance in Cnns with Vector fields*. ICML/FAIM 2018 workshop on Towards learning with limited labels: Equivariance, Invariance, and Beyond (oral presentation). arXiv preprint arXiv:1807.11783.

McHugh, M. L. (2012). Interrater Reliability: the Kappa Statistic. *Biochem. Med.* 22, 276–282. doi:10.11613/bm.2012.031

Molin, J., Bodén, A., Treanor, D., Fjeld, M., and Lundström, C. (2016). *Scale Stain: Multi-Resolution Feature Enhancement in Pathology Visualization*. arXiv preprint arXiv:1610.04141.

Naylor, P., Laé, M., Reyal, F., and Walter, T. (2018). Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map. *IEEE Trans. Med. Imaging* 38, 448–459. doi:10.1109/TMI.2018.2865709

Naylor, P., Laé, M., Reyal, F., and Walter, T. (2017). Nuclei Segmentation in Histopathology Images Using Deep Neural Networks. In 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017), Melbourne, Australia, 18–21 April 2017 (IEEE), 933–936. doi:10.1109/isbi.2017.7950669

Otálora, S., Atzori, M., Andrearczyk, V., and Müller, H. (2018a). "Image Magnification Regression Using Densenet for Exploiting Histopathology Open Access Content," in *Computational Pathology and Ophthalmic Medical Image Analysis* (New York, USA: Springer), 148–155. doi:10.1007/978-3-030-00949-6_18

Otálora, S., Perdomo, O., Atzori, M., Andersson, M., Jacobsson, L., Hedlund, M., et al. (2018b). Determining the Scale of Image Patches Using a Deep Learning Approach. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, 4–7 Aprile 2018. (IEEE), 843–846. doi:10.1109/isbi.2018.8363703

Paramanandam, M., O'Byrne, M., Ghosh, B., Mammen, J. J., Manipadam, M. T., Thamburaj, R., et al. (2016). Automated Segmentation of Nuclei in Breast Cancer Histopathology Images. *PloS one* 11, e0162053. doi:10.1371/journal.pone.0162053

Ren, J., Hacihaliloglu, I., Singer, E. A., Foran, D. J., and Qi, X. (2018). Adversarial Domain Adaptation for Classification of Prostate Histopathology Whole-Slide Images. International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018 (Springer), 201–209. doi:10.1007/978-3-030-00934-2_23

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical image computing and computer-assisted intervention, Munich, Germany, 5-9 October 2015 (Springer), 234–241. doi:10.1007/978-3-319-24574-4_28

Rorke, L. B. (1997). Pathologic Diagnosis as the Gold Standard. *Cancer* 79, 665–667. doi:10.1002/(sici)1097-0142(19970215)79:4<665::aid-cncr1>3.0.co;2-d

Roy, K., Banik, D., Bhattacharjee, D., and Nasipuri, M. (2019). Patch-based System for Classification of Breast Histology Images Using Deep Learning. *Comput. Med. Imaging Graphics* 71, 90–103. doi:10.1016/j.compmedimag.2018.11.003

Salvi, M., and Molinari, F. (2018). Multi-tissue and Multi-Scale Approach for Nuclei Segmentation in H&E Stained Images. *Biomed. Eng. Online* 17, 89. doi:10.1186/s12938-018-0518-0

Schmitz, R., Madesta, F., Nielsen, M., Krause, J., Werner, R., and Rösch, T. (2019). *Multi-scale Fully Convolutional Neural Networks for Histopathology Image Segmentation: From Nuclear Aberrations to the Global Tissue Architecture*. *Medical Image Analysis* 70, 101996.

Sellaro, T. L., Filkins, R., Hoffman, C., Fine, J. L., Ho, J., Parwani, A. V., et al. (2013). Relationship between Magnification and Resolution in Digital Pathology Systems. *J. Pathol. Inform.* 4, 21. doi:10.4103/2153-3539.116866

Tensmeyer, C., and Martinez, T. (2016). *Improving Invariance and Equivariance Properties of Convolutional Neural Networks* ICLR 2017 conference.

Ullah, I. (2017). *A Pyramidal Approach for Designing Deep Neural Network Architectures* PhD thesis. Available at: https://air.unimi.it/handle/2434/466758#.YQEi7FMzYWo.

van der Laak, J., Litjens, G., and Ciompi, F. (2021). Deep Learning in Histopathology: the Path to the Clinic. *Nat. Med.* 27, 775–784. doi:10.1038/s41591-021-01343-4

van Rijthoven, M., Balkenhol, M., Siliņa, K., van der Laak, J., and Ciompi, F. (2020). Hooknet: Multi-Resolution Convolutional Neural Networks for Semantic Segmentation in Histopathology Whole-Slide Images. *Med. Image Anal.* 68, 101890. doi:10.1016/j.media.2020.101890

Wang, S., Yang, D. M., Rong, R., Zhan, X., and Xiao, G. (2019). Pathology Image Analysis Using Segmentation Deep Learning Algorithms. *Am. J. Pathol.* 189, 1686–1698. doi:10.1016/j.ajpath.2019.05.007

Wu, Q., Merchant, F., and Castleman, K. (2010). *Microscope Image Processing*. New York, USA: Elsevier.

Yang, Z., Ran, L., Zhang, S., Xia, Y., and Zhang, Y. (2019). Ems-net: Ensemble of Multiscale Convolutional Neural Networks for Classification of Breast Cancer Histology Images. *Neurocomputing* 366, 46–53. doi:10.1016/j.neucom.2019.07.080

Zaveri, M., Hemati, S., Shah, S., Damskinos, S., and Tizhoosh, H. (2020). Kimia-5mag–a Dataset for Learning the Magnification in Histopathology Images. In 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), 9–11 November 2020 (IEEE), 363–367.

Zeng, N., Li, H., and Peng, Y. (2021a). A New Deep Belief Network-Based Multi-Task Learning for Diagnosis of Alzheimer's Disease. *Neural Comput. Appl.*, 1–12. doi:10.1007/s00521-021-06149-6

Zeng, N., Li, H., Wang, Z., Liu, W., Liu, S., Alsaadi, F. E., et al. (2021b). Deep-reinforcement-learning-based Images Segmentation for Quantitative Analysis of Gold Immunochromatographic Strip. *Neurocomputing* 425, 173–180. doi:10.1016/j.neucom.2020.04.001

Zeng, N., Wang, Z., Zhang, H., Kim, K.-E., Li, Y., and Liu, X. (2019). An Improved Particle Filter with a Novel Hybrid Proposal Distribution for Quantitative Analysis of Gold Immunochromatographic Strips. *IEEE Trans. Nanotechnology* 18, 819–829. doi:10.1109/tnano.2019.2932271

Zhang, Q., Heldermon, C. D., and Toler-Franklin, C. (2020). Multiscale Detection of Cancerous Tissue in High Resolution Slide Scans. In International Symposium on Visual Computing, San Diego, USA, 5–7 October 2020 (Springer), 139–153. doi:10.1007/978-3-030-64559-5_11

Zhu, W., Qiu, Q., Calderbank, R., Sapiro, G., and Cheng, X. (2019). *Scale-equivariant Neural Networks with Decomposed Convolutional Filters*. arXiv preprint arXiv:1909.11193.

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.

frontiers in Medical Technology

ORIGINAL RESEARCH
published: 12 August 2021
doi: 10.3389/fmedt.2021.666650

FPGA-Based High-Performance Phonocardiography System for Extraction of Cardiac Sound Components Using Inverse Delayed Neuron Model

Madhubabu Anumukonda[1]*, *Prasadraju Lakkamraju*[1] and *Shubhajit Roy Chowdhury*[2]

[1] Center for Very Large Scale Integration and Embedded Systems Technology, International Institute of Information Technology Hyderabad, Hyderabad, India, [2] School for Computing and Electrical Engineering, Indian Institute of Technology Mandi, Suran, India

The study focuses on the extraction of cardiac sound components using a multi-channel micro-electromechanical system (MEMS) microphone-based phonocardiography system. The proposed multi-channel phonocardiography system classifies the cardiac sound components using artificial neural networks (ANNs) and synaptic weights that are calculated using the inverse delayed (ID) function model of the neuron. The proposed ANN model was simulated in MATLAB^R and implemented in a field-programmable gate array (FPGA). The proposed system examined both abnormal and normal samples collected from 30 patients. Experimental results revealed a good sensitivity of 99.1% and an accuracy of 0.9.

Keywords: phonocardiography, cardiac sounds, inverse delayed function model of neuron, artificial neural networks, field programmable gate array

OPEN ACCESS

Edited by:
Nianyin Zeng,
Xiamen University, China

Reviewed by:
Zhe Yang,
Dongguan University of Technology, China
Han Li,
Xiamen University, China

***Correspondence:**
Madhubabu Anumukonda
madhubabu.a@research.iiit.ac.in

Specialty section:
This article was submitted to Medtech Data Analytics, a section of the journal Frontiers in Medical Technology

Received: *10 February 2021*
Accepted: *05 July 2021*
Published: *12 August 2021*

Citation:
Anumukonda M, Lakkamraju P and Chowdhury SR (2021) FPGA-Based High-Performance Phonocardiography System for Extraction of Cardiac Sound Components Using Inverse Delayed Neuron Model. Front. Med. Technol. 3:666650. doi: 10.3389/fmedt.2021.666650

INTRODUCTION

Heart diseases are one of the major causes of human death worldwide. In the last 15 years, heart disease and stroke have been the leading killers and causes of death on a global scale (1). Heart failure has no cure, but early detection of its related symptoms helps in properly diagnosing heart diseases and, thus, reducing the death rate. In the modern technological revolution, many heart diagnosis methods like phonocardiography (PCG), ECG, echocardiogram (Echo), cardiac MRI (CMRI), and CT heart scan are available to detect early heart failure. Each method has its own advantages and disadvantages; for example, ECGs are widely used in diagnosis but have trouble detecting the structural abnormalities of the heart valves, which can be detected through heart murmurs. Furthermore, echo, CMRI, and CT scans provide accurate results but need a lot of pre-evaluation, are high cost, and are not affordable for many people.

Phonocardiography is one of the non-invasive methods for the detection of major heart sounds and murmurs (2). The stethoscope was the primary PCG instrument that was used to detect cardiac auscultation; however, it had limitations in terms of clinical expertise for the analysis of the low-frequency amplitudes formed during heart failures, such as systolic and diastolic murmurs (3, 4). A great deal of research has been conducted in the field of heart sound segmentation and classification in order to detect normal and abnormal components. Many signal processing algorithms are proposed to extract heart sound components for detecting pathological events (5–7). Artificial neural networks (ANNs) are widely used in cardiology (8–10). A neural network

model is a well-known method for separating the normal and abnormal pathological events in heart sounds (11–14). A well-trained neural network model can be used to detect the complex relationship between abnormal and normal cardiac sounds. Deep neural networks extract high-level features from low-level features. The methods proposed (15–17) to play significant roles in solving the non-linear functions in medical applications.

Cardiac auscultations are non-linear in nature and are analyzed using ANNs, which are more useful for the approximation of non-linear functions. The neural networks models are mainly classified by their architecture, activation function, and learning algorithm. Instead of directly emulating biological behavior, the traditional neuron network models translate this behavior in terms of time-averaging techniques (9, 18, 19). The inverse delayed (ID) function model proposed by Nakajima (20–22) is a universal neuron model that includes characteristics of both the Bonhoeffer Van der Pol model (19) and the Hopfield model (23). In addition, the ID function model uses the inverse function of the activation function rather than the traditional activation function and features a finite conversion time from the internal state of the element to the output.

The energy function of the ID function model with symmetric synapse weights is similar to that of the Hopfield model. Through selective destabilization, the negative resistance of the ID model can free the neural network state from such local minima (20). Unlike the chaotic neural network, the ID model does not need to transform the output vector, record the output vector during calculation, or control the dynamics by changing the network parameters (22). We only need to wait for it to enter an inactive state in order to find a solution using the ID neural network and a simple method of implementation. The ID model is capable of resolving combinatorial optimization problems (21). The negative resistance of the ID model can destabilize the stable equilibrium points of a neural network, reducing the possibility of unknown values in suboptimal synaptic weight solutions obtained using an ANN based on a traditional neuron model. The ANN implementation using the ID neuron model needs numerous parallel computations to solve the real-time complex data for the extraction of components. The advances in field-programmable gate arrays (FPGAs) handle these real-time complex computations effectively and improve the performance of the system.

The current study concentrated on the non-invasive detection of cardiac component abnormalities in raw samples collected with micro-electromechanical system (MEMS)-based microphones. Wavelet decomposition algorithms were used to generate the featured set. The ID neuron function was used to create the ANN model, which extracted the cardiac components from the feature set obtained. The ID function model of the neuron was used to optimize the weights of the synapses between the neurons. The entire algorithm was implemented on a Xilinx SoC FPGA XC7Z020CLG400 (Xilinx, USA). The proposed system has been validated using the sensitivity, specificity, and accuracy of the cardiac components, and justified using receiver operating characteristic curve analysis. The study is organized as follows: Section Theory focuses on the theory supporting the proposed method using ID function model of the neuron and illustrates the feature sets for cardiac sound assessment. Section Materials and Methods focuses on the materials and methods for the proposed technology. The results and discussions are presented in Section Experimental Results and Discussion.

THEORY

Inverse Delayed Neuron Function Model of the Neuron

The ANN was realized using the ID function model of the neuron. Nakajima and Hayakawa proposed the ID function model of the neuron by the following set of equations:

$$\tau \frac{\mathbf{du_i}}{\mathbf{dt}} = \sum_{\mathbf{j,j\neq i}} \mathbf{w_{ij}x_j} - \mathbf{a_{ii}x_i} - \mathbf{u_i} \tag{1}$$

$$\tau_{\mathbf{x}} \frac{\mathbf{dx_i}}{\mathbf{dt}} = \mathbf{u_i} - \mathbf{g}(\mathbf{x_i}) \tag{2}$$

$$\mathbf{g}(\mathbf{x_i}) = \mathbf{f^{-1}}(\mathbf{x_i}) - \mathbf{Kx_i} \tag{3}$$

Where u_i is the ith neuron internal state, x_j is the jth neuron output, W_{ij} is the synaptic weight between jth and ith neurons, h_i is the bias input, a_{ii} is the self-connection synaptic weight, τ is the internal state time constant, and τ_x is the neuron output time constant.

From Equation (3), f(x) is the neural network sigmoid function, then $g(x) = f^{-1}(x)$ is the N-shaped inverse output function (24). g(x) can be changed with a positive value of K times output of the neuron. The transition time from u to x is less than τ and thus $\tau_x << \tau$. In general, the transition time should be taken into account if it is significantly less than τ. For the present problem, we used the self-connectionless neurons to devoid the hysteresis effects (24). So $a_{ii} = 0$ in Equation (1).

Differentiating Equation (2) with respect to time t, we get

$$\tau_{\mathbf{x}} \frac{\mathbf{d^2x_i}}{\mathbf{dt^2}} = \frac{\mathbf{du_i}}{\mathbf{dt}} - \frac{\mathbf{dg}(\mathbf{x_i})}{\mathbf{dx_i}} \frac{\mathbf{dx_i}}{\mathbf{dt}} \tag{4}$$

Let us consider

$$\varphi_{\mathbf{i}} = \frac{\mathbf{dg}(\mathbf{x_i})}{\mathbf{dx_i}} + \frac{\tau_{\mathbf{x}}}{\tau} \tag{5}$$

Substituting Equation (5) and (4) gives

$$\tau_{\mathbf{x}} \frac{\mathbf{d^2x_i}}{\mathbf{dt^2}} + \varphi_{\mathbf{i}} \frac{\mathbf{dx_i}}{\mathbf{dt}} - \frac{\tau_{\mathbf{x}}}{\tau} \frac{\mathbf{dx_i}}{\mathbf{dt}} = \frac{\mathbf{du_i}}{\mathbf{dt}} \Leftrightarrow \tau_{\mathbf{x}} \frac{\mathbf{d^2x_i}}{\mathbf{dt^2}} + \varphi_{\mathbf{i}} \frac{\mathbf{dx_i}}{\mathbf{dt}} = \frac{\mathbf{1}}{\tau} \left(\sum_{\mathbf{j}} (\mathbf{w_{ij}x_j} - \mathbf{g}(\mathbf{x_i})) \right) \tag{6}$$

Let, $\frac{\partial U_i}{\partial x_i} = \frac{1}{\tau} \left(g(x_i) - \sum w_{ij} x_j \right)$
Equation (6) becomes

$$\tau_{\mathbf{x}} \frac{\mathbf{d^2x_i}}{\mathbf{dt^2}} + \varphi_{\mathbf{i}} \frac{\mathbf{dx_i}}{\mathbf{dt}} = \frac{\partial \mathbf{U_i}}{\partial \mathbf{x_i}} \tag{7}$$

Where $U_i = \frac{1}{\tau} \left(\int_0^x g(x_i) \, dx_i - x_i \sum_j w_{ij} x_j \right)$

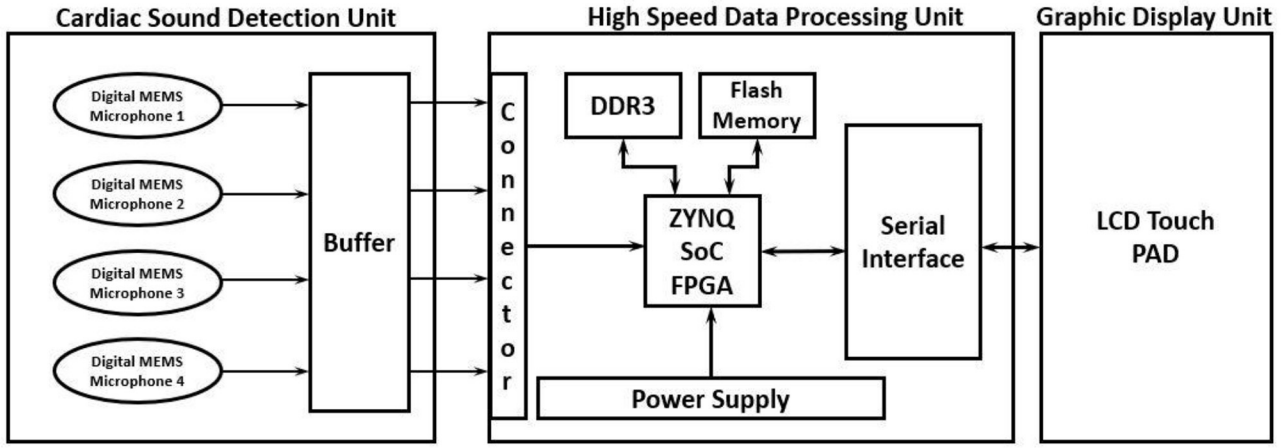


FIGURE 1 | Phonocardiography system block diagram.

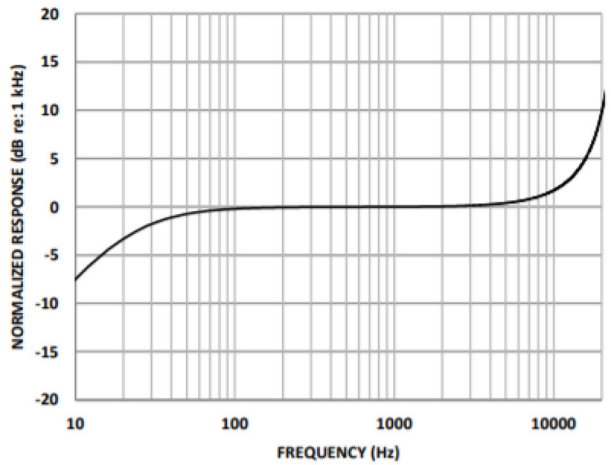


FIGURE 2 | Micro-electromechanical system (MEMS) microphone frequency response.

U_i denotes the potential of the ID function model of the neuron. In Equation (7), the first term denotes the inertia and the second term denotes the friction. If $g(x_i)$ is an N-shaped function, then the area where $\frac{dg(x)}{dx_i}$ is less than $\frac{-\tau_x}{\tau}$ for specific values of x_i is called the negative resistance region.

From the Lyapunov function, the energy of the ID function model is

$$E = -\frac{1}{2\tau}\sum_i\sum_j w_{ij}x_ix_j + \frac{1}{\tau}\sum_i\int_0 g(x_i)\,dx_i + \frac{\tau_x}{2}\sum_i\left(\frac{dx_i}{dt}\right)^2 \quad (8)$$

Since the proposed neuron network has self-connectionless neurons, the self-connections between neurons are ignored. The last term in Equation (8) shows the time delay in the ID function model.

Differentiating both sides of Equation (8) with respect to time t, we get

$$\frac{dE}{dt} = -\sum_i \frac{dx_i}{dt}\left\{\frac{1}{\tau}w_{ij}x_j - \frac{1}{\tau}g(x_i) - \tau_x\frac{d^2x_i}{dt^2}\right\} \quad (9)$$

$$\frac{dE}{dt} = -\sum_i\left(\frac{dg(x_i)}{dx_i} + \frac{\tau_x}{\tau}\right)\left(\frac{dx_i}{dt}\right)^2 \quad (10)$$

$$\frac{dE}{dt} = -\sum_i \varphi_i\left(\frac{dx_i}{dt}\right)^2 \quad (11)$$

From Equation (11), the energy (E) of the ID model, similar to that of the Hopfield model, decreases with time if the network state is in the positive resistance region ($\phi i > 0$). However, in the negative resistance region ($\phi i < 0$), the energy (E) increases with time; thus, even if the state is in the minima region, it quickly exits this region. It is necessary to have this feature in order to avoid local minima. As a result, if the network is an ID function model, the likelihood of escaping the local minima is expected to increase.

Feature Sets for Cardiac Sound Assessment

The assessment of the cardiac sound components involved different parameters that were deferred by a set of features to sort out the components from the heart sounds. In the current research, the Springer segmentation algorithm by Springer et al. (25) was used to differentiate the heart sounds using timing intervals of S1, S2, systole, and diastole. The following parameters were used as feature sets to identify the low-frequency abnormal components and normal heart sound components.

TABLE 1 | Cardiac sounds feature sets.

S. no	Feature	Description
1	F1	Mean of the Systolic to diastolic time interval ratio of each heart sound.
2	F2	Mean of the S1, S2 intervals ratio.
3	F3	Mean of the heart sound peak energy in systolic cycle to total cardiac cycle energy of each heart beat
4	F4	Mean of the heart sound peak energy in diastolic cycle to total cardiac cycle energy of each heartbeat.
5	F5	Mean of spectral frequencies from 10 to 900 Hz with window resolution of 10 Hz in systole cycle of each heartbeat.
6	F6	Mean of spectral frequencies from 10 to 900 Hz with window resolution of 10 Hz in diastolic cycle of each heartbeat.

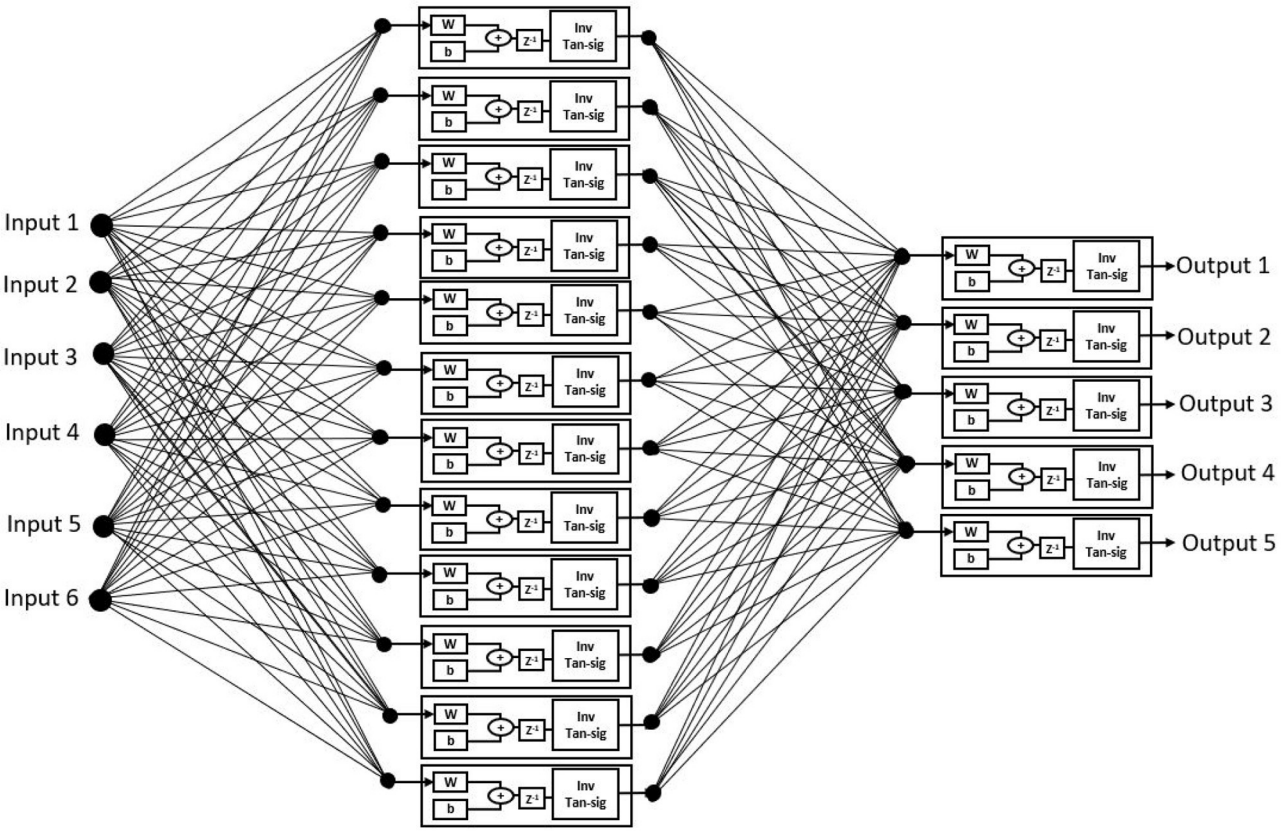


FIGURE 3 | Artificial neural networks (ANNs) using the ID function neuron model.

MATERIALS AND METHODS

High-Performance Phonocardiography System Hardware

The proposed high-performance phonocardiography system was developed based on the advanced MEMS microphone to capture the low-frequency components and analyze the captured data using a proposed algorithm based on the inverse delayed neuron model. The proposed algorithm was implemented on the Xilinx Zynq-7 System on-chip FPGA, which has a dual-core ARM cortex-A9 for application software and programmable logic for algorithm complex computations. The detailed block diagram for the high-performance phonocardiography system is as shown in **Figure 1**.

A cardiac sound detection unit consists of a MEMS microphone, which is a tiny integrated circuit with a sound transducer, an analog front end, and a signal conditioning circuit (7). The MEMS microphone has a high Signal-to-Noise Ratio (SNR) at 70 dB and a good frequency response from 10 Hz to 10 kHz as shown in **Figure 2**. Due to its flat response and high SNR in the lower region, a cardiac sound detection (CSD) is more suitable for the detection of the third and fourth heart sounds and murmurs. The CSD module has four microphones, the placements of which were on the basis of sound source localization to cover the four heart valves (aortic valve, tricuspid valve, mitral valve, and pulmonary valve). These valves are the origins of cardiac sounds.

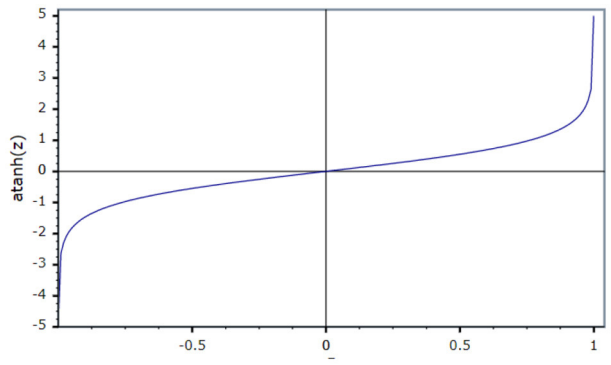


FIGURE 4 | Transfer function.

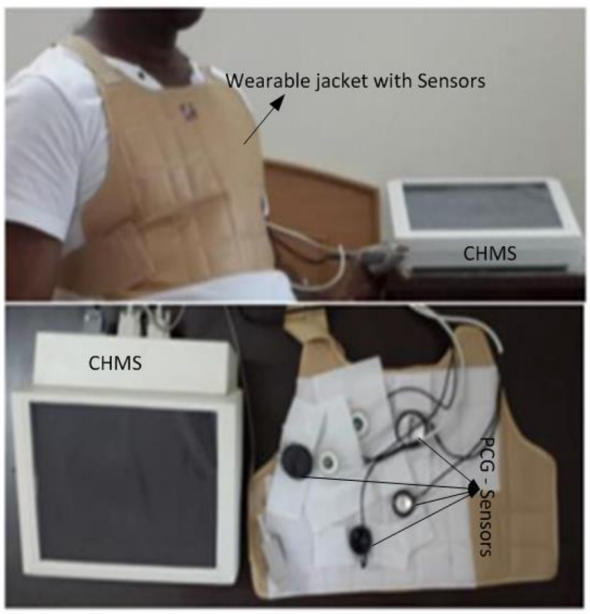


FIGURE 5 | Real-time experimental setup with proposed hardware.

A high-speed data processing unit consists of a Xilinx Zynq-7 System on-chip FPGA, which has a dual-core ARM cortex-A9 processor for the application software and programmable logic for algorithm complex computations. The high-speed data processing (HSDP) module is responsible for the separation of cardiac sounds based on the frequencies and processes used in the proposed ANN-based ID model neuron. The Graphical Display Recorder (GDR) unit consists of an LCD touchpad for parameter configuration and a display for analyzing the results for further diagnosis. The application software was developed on an ARM cortex-A9 processor and interfaced to the GDR module.

The current system implemented was in two phases. In the first phase, the proposed algorithm was modeled using the MATLAB, simulated with different test parameters, and baselined as a golden reference for further hardware system development. The proposed neural network model was implemented on FPGA in the second phase to achieve performance comparable to the MATLAB model. The cardiac sounds detected by the CSD unit were then passed to the HSDP unit for the extraction of the cardiac feature set mentioned in **Table 1**. Afterward, these feature sets were given as inputs to the hidden layer and known spectrograms, which would train the network and predict the error deviation to assess the features of the cardiac sounds.

Prediction Model Using ID Function Model of the Neuron

The proposed ANN algorithm is based on a feedforward network with three layers as shown in **Figure 3**. The first layer has six inputs for the six feature sets mentioned in **Table 1**. The second layer is a hidden layer that consists of 12 hidden neurons that compute the delayed weighted sum of inputs and the inverse tangent sigmoid non-linear function for the feature extraction. The third layer is an output layer that consists of five output neurons. It is a logical net to reduce the error in extraction and sends the output based on the input from the hidden layer.

Realization of Inverse Activation Function

An activation function is used to present non-linearity into the output of the neuron. For the current work, an inverse tan hyperbolic function is taken as the activation function.

$$\mathbf{f}(\mathbf{x}) = \mathbf{tan}^{-1}\,\mathrm{h}(\mathbf{x}) = \frac{1}{2}\ln\left(\frac{1+\mathbf{x}}{1-\mathbf{x}}\right) \tag{12}$$

$$\mathbf{f}(\mathbf{x}) = \frac{1}{2}\left(\ln(1+\mathbf{x}) - \ln(1-\mathbf{x})\right) \tag{13}$$

$$\mathbf{f}(\mathbf{x}) = \frac{1}{2}\left(2\mathbf{x} + \frac{2}{3}\mathbf{x}^3 + \frac{2}{5}\mathbf{x}^5 + \ldots\right) \tag{14}$$

$$\mathbf{f}(\mathbf{x}) = \mathbf{x} + \frac{\mathbf{x}^3}{3} + \frac{\mathbf{x}^5}{5} + \ldots \tag{15}$$

Neglecting the higher terms f(x) becomes

$$\mathbf{f}(\mathbf{x}) = \mathbf{x} + \frac{\mathbf{x}^3}{3} \tag{16}$$

Equation (16) is realized using a constant (1/3), adder, and multiplier. The activation limits the output in the range of (1,−1).

Figure 4 shows that the activation function in the ID model is an N-shaped transfer function.

The ID network, which consisted of 12 hidden neurons in the hidden layer, was trained using MATLAB^{R}. Learning was accomplished through the use of the Levenberg–Marquardt backpropagation algorithm. Backpropagation was used to obtain the input and layer weight matrices by incorporating the derivatives of the inverse functions. These matrices were used to replicate the ID network onto the FPGA. The neural network was trained using the ID function model. The result of the regression coefficient "*R*" is shown in **Figure 13**.

System-Level Implementation

The experimental setup with PCG sensors and the Cardiac Health Monitoring System (CHMS) is shown in **Figure 5**. The proposed

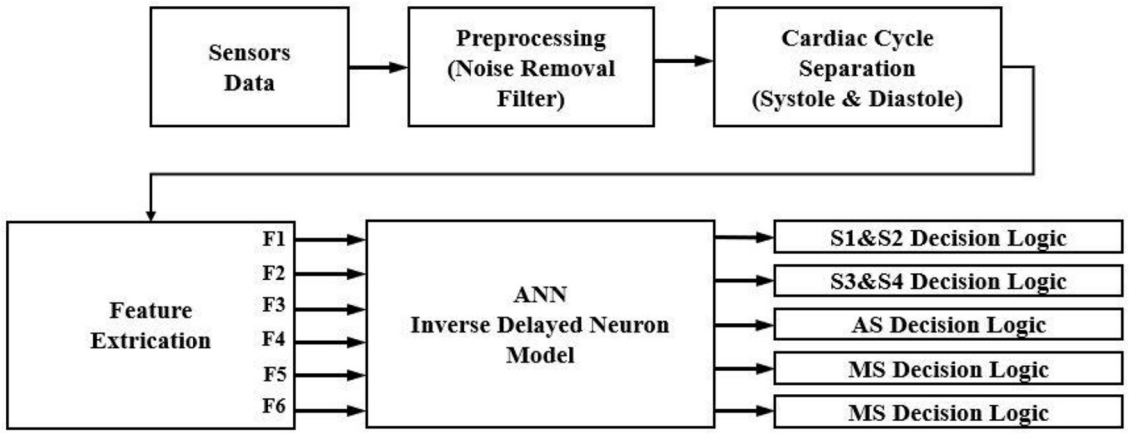


FIGURE 6 | Flow diagram for extraction of cardiac sound components.

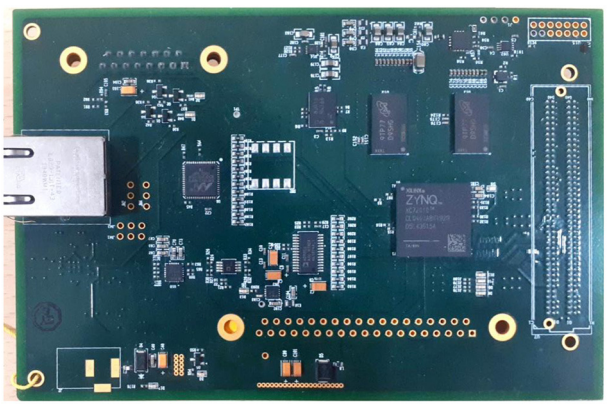


FIGURE 7 | Field-programmable gate array (FPGA) hardware for the proposed system.

algorithm implementation in FPGA is shown in **Figure 6**. The prediction model, which was based on the ID function neuron model, was developed in the system generator tool, and the Verilog netlist for RTL integration was generated. The VIVADO design suite 2018.2 was used for simulation and synthesis. The model was successfully dumped onto the Zynq7 – XC7Z010-1CLG400 device FPGA board. The PCG sensor data captured in FPGA internal Block RAM memory were passed through the pre-processing module to remove the unwanted noise. The cleaned data were then processed through the cardiac cycle separation engine to differentiate the systole and diastole cycles. The feature extraction module extracted the features mentioned in **Table 1** from systole and diastole. The extracted six featured sets were passed through the ANN model to classify them into five heart sound component groups. The decision logic outputted the true-negative (TN), true-positive (TP), false-negative (FN), and false-positive (FP) for corresponding heart sound components. **Figure 7** shows the FPGA hardware board developed for the proposed system.

The ANN with the inverse delayed neuron model was implemented in MATLAB^R using the neural network toolbox and the fixed-point toolbox as discussed in previous sections. The model was simulated to check for functionality and then used to calculate the synaptic weights and bias required for the hidden and output layers. The simulated fixed-point MATLAB model was taken as a golden reference for further hardware realization using FPGA.

The Xilinx System generator tool was used for the implementation and generation of the Verilog code for the system integration. As discussed in earlier sections, the ANN model was realized in three stages. For the first stage, in the input layer, the inputs were scaled with weights and passed to hidden neurons. For the second stage, the bias was added to the summed weights by the hidden layer, which then passed through the delayed activation function. For the third stage, the output layer computed the output value from all hidden neurons, the output bias, and the activation function. The sigmoid activation function was a building block that was used in both the hidden and output layers (26). Equation (16) was used to implement the inverse activation function, which consists of an adder, multiplier, and constant value. As shown in **Figure 8**, the neuron model was realized using a Mult-Add block and a constant block for bias, an activation function.

Figure 9 depicts the proposed ANN model implementation. The estimated sigmoid value is closer to the real sigmoid value obtained from MATLAB, allowing the approximation effect to be reduced for improved accuracy when implemented in hardware.

EXPERIMENTAL RESULTS AND DISCUSSION

The extraction of cardiac sounds from the MEMS-based high-performance phonocardiography system using a neural network was based on the ID function model of the neuron. The neural networks based on the ID function model and the

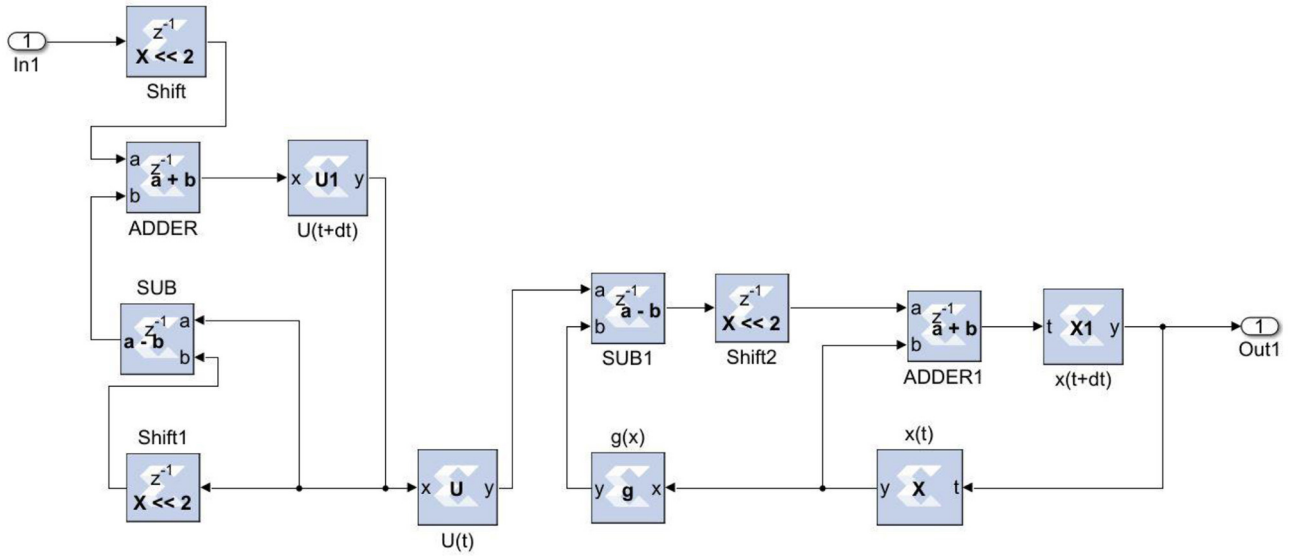


FIGURE 8 | Neuron model.

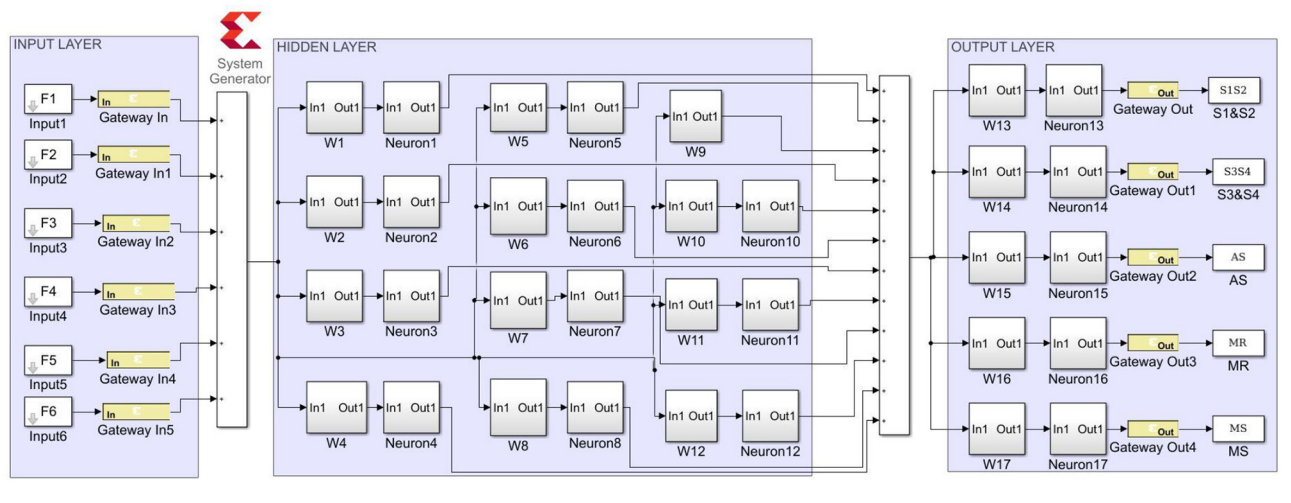


FIGURE 9 | System generator implementation of the ANN model.

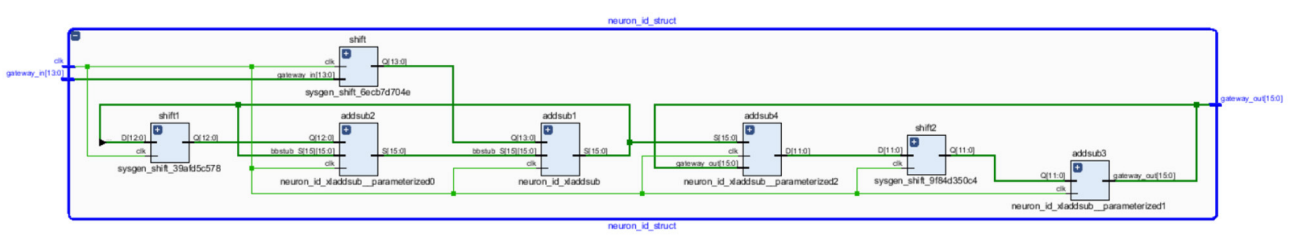


FIGURE 10 | RTL Schematic—Neuron Model.

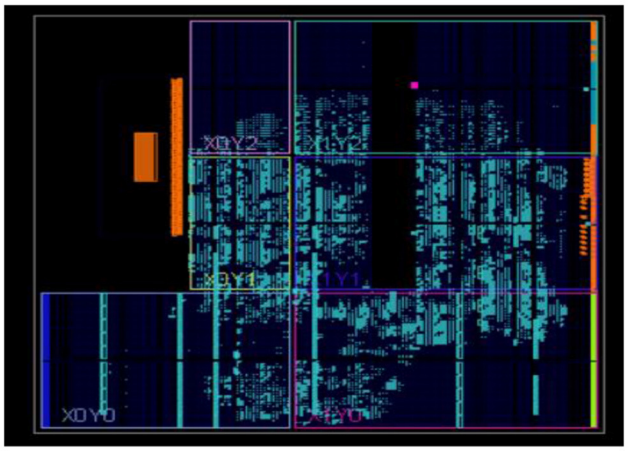


FIGURE 11 | Physical layout for the proposed system.

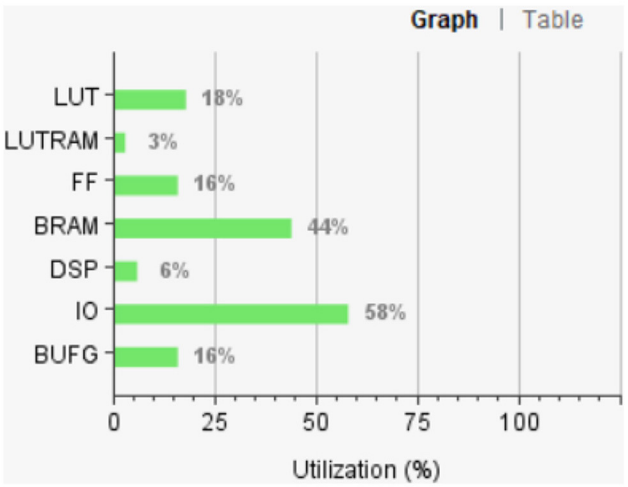


FIGURE 12 | Resource utilization of the proposed system.

conventional neuron models were realized on FPGA, and the hardware requirements and performance of the two models were compared. The proposed system was validated using the data of 30 patients in accordance with the Declaration of Helsinki. After obtaining informed consent, a total of 60 patients were made available for evaluation, with the data of 30 patients being used for training the neural networks and the data of the remaining 30 patients used for testing the proposed system.

The neural networks were trained offline using data from the 30 patients in MATLAB^R, and then implemented on FPGA to reduce design circuitry. The MATLAB^R model weights and biases were used as hard-coded values in the FPGA ANN model to reduce computational cycles and achieve the accuracy obtained in the simulation. The FPGA implementation of the inverse tangent sigmoid function, which requires the realization of an N-shaped activation function, involved multiplication but not division. The multiplier was all that was required for the functional units of the ID model. This greatly reduced hardware complexity. The system generator FPGA netlist files were used to run the synthesis, and implement and generate the bit file needed to program the FPGA. **Figure 10** depicts the RTL schematic for the neuron model following RTL synthesis. The physical layout of the proposed system is depicted in **Figure 11**.

Figure 12 shows the FPGA resource utilization of the proposed layout after place and route.

The mean square error in the extraction of cardiac sound components detection rate using neural networks with 12 neurons in the hidden layer was .9. Regression analysis was performed on the input and target datasets, and the mean square error was found to be 4.4×10^{-5}. **Figure 13** depicts the regression analysis of the training and validation of the network. In the regression analysis, the parameter "R" = 0.99 represents the correlation between extracted cardiac components and actual cardiac components.

The regression analysis showing the training and validation of the network is depicted in **Figure 13**. The parameter "R," which is equal to 0.99 in the regression analysis, signifies the correlation between extracted cardiac components with actual cardiac components. **Figure 14** depicts the training state analysis of a neural network based on the ID function model of a neuron. **Figure 15** depicts a neural network for performance analysis. Training analysis was performed for epoch 11, with a gradient factor of 9.0661×10^{-5} and validation checks equal to 6.

In accordance with the Declaration of Helsinki, clinical trials were conducted on 30 patients using the proposed high-performance phonocardiography system, and the results were compared to the known test results from the medical practitioner. The prevalence of disease in the tested population, the outcome of the diagnostic test, and the sensitivity and specificity of the test all had an impact on the reliability of any diagnostic test result. The sensitivity, the specificity rate, and the accuracy can be computed as follows

$$\textbf{Sensitivity} = \frac{\textbf{TP}}{\textbf{TP} + \textbf{FN}} \times \textbf{100\%} \tag{17}$$

$$\textbf{Specificity} = \frac{\textbf{TN}}{\textbf{TN} + \textbf{FP}} \times \textbf{100\%} \tag{18}$$

$$\textbf{Accuracy} = \frac{\textbf{TP} + \textbf{TN}}{\textbf{TP} + \textbf{TN} + \textbf{FP} + \textbf{FN}} \times \textbf{100\%} \tag{19}$$

where true-negative (TN) represents the number of correct heart sound components rejected, true-positive (TP) represents the number of correct heart sound components detected by the proposed system, false-negative (FN) represents the number of incorrect heart sound components rejected, and false-positive (FP) represents the number of incorrect heart sound components detected by the proposed system. **Table 2** displays the sensitivity, specificity, and accuracy of cardiac component detection for a healthy individual under specific disease conditions.

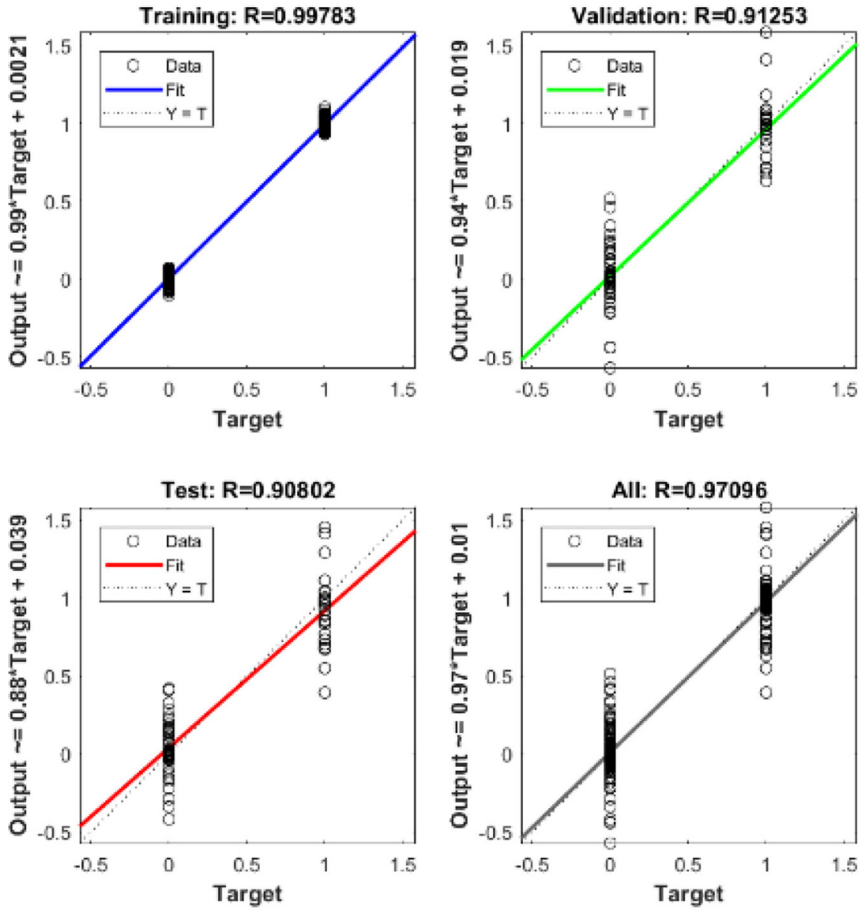


FIGURE 13 | Regression analysis of neural network based on the ID function model of the neuron.

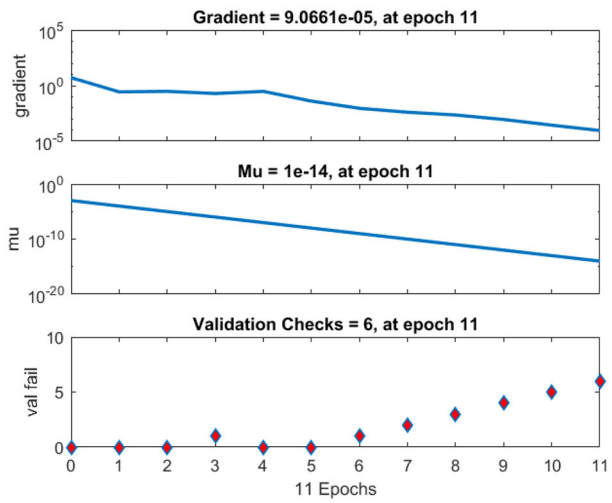


FIGURE 14 | Training state analysis of neural network based on the ID function model of the neuron.

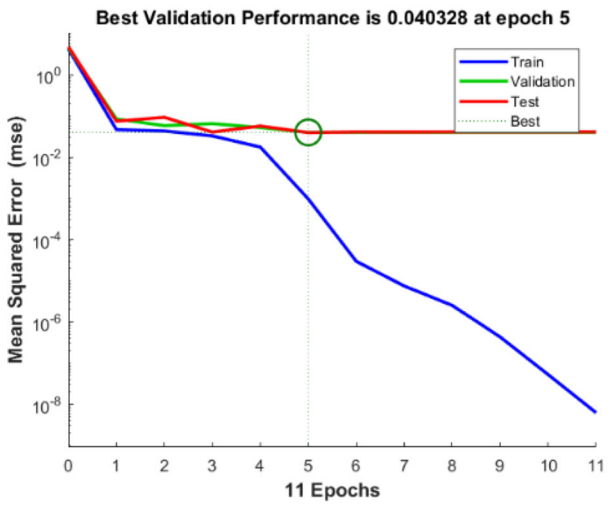


FIGURE 15 | Performance analysis of neural network based on the ID function model of the neuron.

The performance of the proposed method was evaluated using the receiver operating characteristic (ROC) curve area under curve (AUC) value. This validated the extraction of cardiac components from the captured data using the proposed algorithm. The ROC curve for the extraction of cardiac component accuracy for the proposed ID neuron model system is shown in **Figure 16**. The accuracy of S1 and S2, S3 and S4, aortic stenosis, mitral Stenosis, and mitral regurgitation is 99.3, 98.6, 98.7, 98.7, and 98.6%, respectively, based on the AUC values in **Figures 16A–E**.

TABLE 2 | Accuracy of proposed system for different cardiac components in heart sounds.

Heart sound components	Sensitivity (%)	Specificity (%)	Accuracy
S1 and S2	99.1	99.3	0.99
S3 and S4	98.1	98.6	0.98
Aortic stenosis	98.3	98.7	0.98
Mitral stenosis	98.5	98.7	0.98
Mitral regurgitation	98.2	98.4	0.98

CONCLUSION

The current research focused on the development of multi-channel MEMS-based phonocardiography system to capture heart sounds and process the acquired sample to remove unwanted noise and derive a feature set using wavelet transforms. Thereafter, the low frequency cardiac sounds were extracted using the ANN based on the ID function of the neuron model. The neural network was trained using real, known data, and the proposed system was tested using patient test data. The developed ANN-based phonocardiography system was useful to the physician for recognizing abnormal, low-frequency heart sounds with a simple diagnosis setup similar to the stethoscope and visualizing graphical data for better medical diagnoses. The performance of the phonocardiography system was evaluated using 2,150 cardiac cycles of PCG from a cohort of 30 patients with different pathophysiological conditions, resulting in a sensitivity of 99% and an accuracy of 0.9.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of Non-disclosure Agreement with the patients.

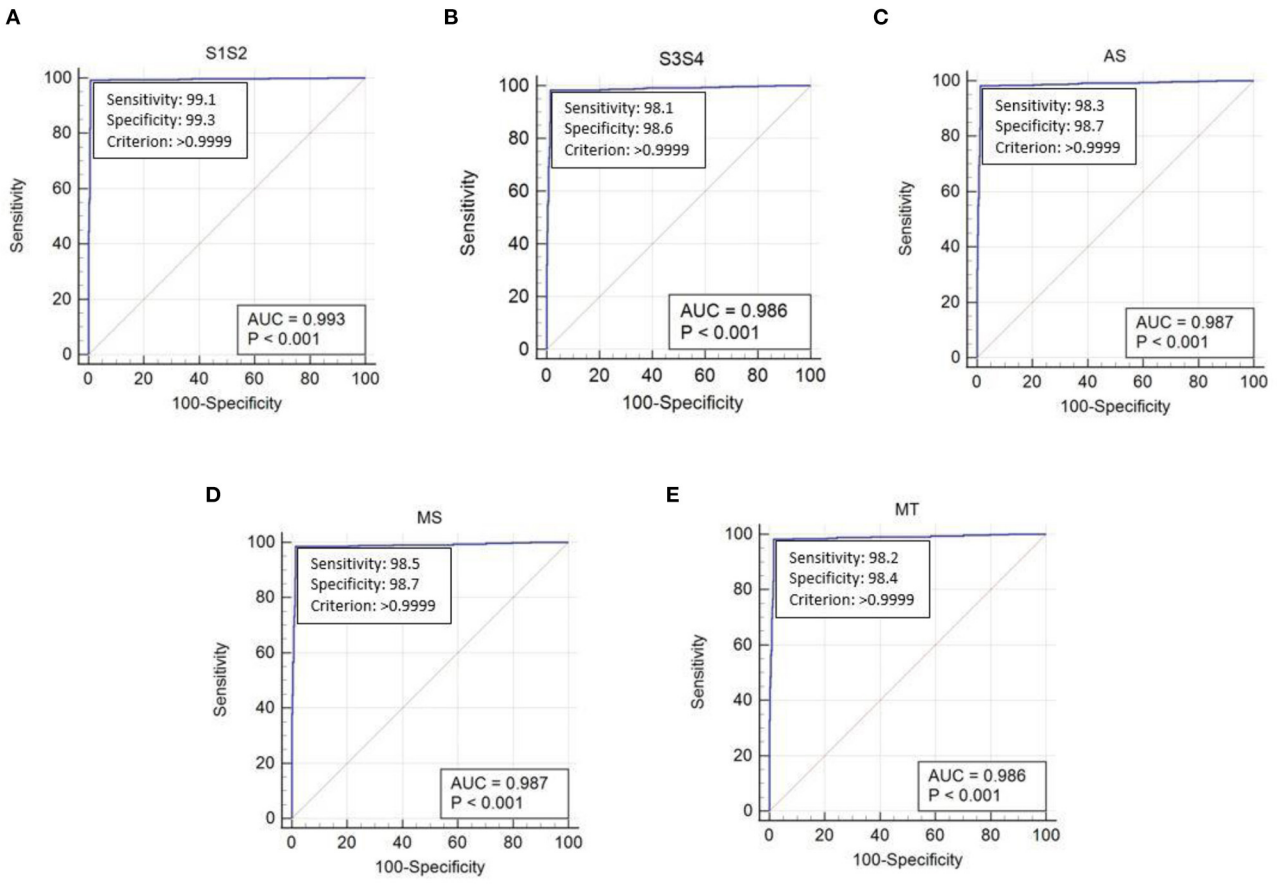


FIGURE 16 | The receiver operating characteristic (ROC) curve analysis of heart sound components with ID function of the neuron model **(A)** first and second heart sound, **(B)** third and fourth heart sound, **(C)** aortic stenosis, **(D)** mitral stenosis, and **(E)** mitral regurgitation.

Requests to access the datasets should be directed to Madhubabu, mail.madhubabu@gmail.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Citizens Hospitals Hyderabad. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed in developing the proposed system, methodology, analyze the results, and manuscript editing.

ACKNOWLEDGMENTS

The author would like to thank the Faculty of Center for VLSI and Embedded systems, IIIT-Hyderabad and SCEE, and IIT-Mandi for their valuable support during this research.

REFERENCES

1. Benjamin E, Muntner P, Alonso A, Bittencourt M, Callaway C, Carson A, et al. Heart disease and stroke statistics-2019 update: a report from the american heart association. *Circulation*. (2019) 139:659. doi: 10.1161/CIR.0000000000000659
2. Karnath B, ThorntonW. *Auscultation of the Heart*. Hospital Physician (2002). p. 39–43.
3. Rangayyan RM, Lehner RJ. Phonocardiogram signal analysis: a review. *Crit Rev Biomed Eng*. (1987) 15:211–36.
4. Mangione S. Cardiac auscultatory skills of physiciansin-training: a comparison of three english-speaking countries. *Am J Med*. (2001) 110:210–6. doi: 10.1016/S0002-9343(00)00673-2
5. Koutsiana E, Hadjileontiadis LJ, Chouvarda I, Khandoker AH. Fetal heart sounds detection using wavelet transform and fractal dimension. *Front Bioeng Biotechnol*. (2017) 5:e00049. doi: 10.3389/fbioe.2017.00049
6. Anumukonda M, Lakkamraju PR, Chowdhury SR. Classification of abnormal and normal heart sounds using the MEMS based high performance phonocardiography system. In: *International Conference on Artificial Intelligence and Signal Processing (AISP).* Amaravati (2020). p. 1–7.
7. Durand LG, Pibarot P. Digital signal processing of the phonocardiogram: review of the most recent advancements. *CRC Crit Rev Biomed Eng*. (1995) 23:163–219. doi: 10.1615/CritRevBiomedEng.v23.i3-4.10
8. Wu M, Lu Y, Yang W, Wong SY. A study on arrhythmia *via* ECG signal classification using the convolutional neural network. *Front Comput Neurosci*. (2021) 14:564015. doi: 10.3389/fncom.2020.564015
9. Aydin RC, Braeu FA, Cyron CJ. General multi-fidelity framework for training artificial neural networks with computational models. *Front Mater*. (2019) 6:e00061. doi: 10.3389/fmats.2019.00061
10. Akay M. Noninvasive diagnosis of coronary artery disease using a neural network algorithm. *Biol Cybern*. (1992) 67:361–7. doi: 10.1007/BF02414891
11. Messner E, Zöhrer M, Pernkopf F. Heart sound segmentation-an event detection approach using deep recurrent neural networks. *IEEE Trans Biomed Eng*. (2018) 65:1964–74. doi: 10.1109/TBME.2018.2843258
12. Andrisevic N, Ejaz K, Gutierrez FR, Flores RA. Detection of heart murmurs using wavelet analysis and artificial neural networks. *J Biomech Eng*. (2005) 127:899–904. doi: 10.1115/1.2049327
13. Deperlioglu O. Classification of phonocardiograms with convolutional neural networks. *Broad Res Artif Intell Neurosci*. (2018) 9:22–33. Available at: https://lumenpublishing.com/journals/index.php/brain/article/view/2030 (accessed July 27, 2021).
14. Anderson JA. *An Introduction to Neural Networks*. Boston: MIT Press (1995).
15. Zeng N, Li H, Peng Y. A New Deep Belief Network-Based Multi-Task Learning for Diagnosis of Alzheimer's Disease. *Neural Comput Appl*. doi: 10.1007/s00521-021-06149-6
16. Zeng N, Li H, Wang Z, Liu W, Liu S, Alsaadi FE, et al. Deep-reinforcement-learning-based images segmentation for quantitative analysis of gold immunochromatographic strip. *Neurocomputing*. (2021) 425:173–80. doi: 10.1016/j.neucom.2020.04.001
17. Zeng N, Li H, Li Y, Luo X. Quantitative analysis of immunochromatographic strip based on convolutional neural network. *IEEE Access*. (2019) 7:16257–63. doi: 10.1109/ACCESS.2019.2893927
18. Li H, Ren Y, Zhang G, Wang R, Cui J, Zhang W. Detection and classification of abnormities of first heart sound using empirical wavelet transform. *IEEE Access*. (2019) 7:139643–52. doi: 10.1109/ACCESS.2019.2943705
19. FitzHugh R. Impulses and physiological states in theoretical models of nerve membrane. *Biophys J*. (1961) 1:445–66. doi: 10.1016/S0006-3495(61)86902-6
20. Hayakawa Y, Denda T, Nakajima K. Inverse function delayed model for optimization problems. In: Negoita MG, Howlett RJ, Jain LC, editors. *Knowledge-Based Intelligent Information and Engineering Systems. KES 2004. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer (2004).
21. HayakawaY, Nakajima K. Design of the inverse function delayed neural network for solving combinatorial optimization problems. *IEEE Trans Neural Networks*. (2010) 21:224–37. doi: 10.1109/TNN.2009.2035618
22. Nakajima K, Hayakawa Y. Characteristics of inverse delayed model for neural computation. In: *Proceedings 2002 International Symposium on Nonlinear Theory and Its Applications*. (2002). p. 861–4.
23. Hopfield JJ, Tank DW. Neural computational of decisions in optimization problems. *Biol Cybern*. (1985) 52:141–52.
24. Yanai H, Sawada Y. Associative memory network composed of neurons with hysteresis property. *Neural Netw*. (1990) 3:223–8. doi: 10.1016/0893-6080(90)90091-X
25. Chengyu L, Springer D, Qiao L, Benjamin M, Juan R, Chorro F, et al. An open access database for the evaluation of heart sound algorithms. *Physiol Meas*. (2016) 37:2181. doi: 10.1088/0967-3334/37/12/2181
26. Zhang M, Vassiliadis S, Delgado-Frias JG. Sigmoid generators for neural computing using piecewise approximations. *IEEE Trans Comput*. (1996) 45:1045–9. doi: 10.1109/12.537127

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.

frontiers in Public Health

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fpubh.2021.726140

Identifying the Phenotypic and Temporal Heterogeneity of Knee Osteoarthritis: Data From the Osteoarthritis Initiative

Mengjiao Li[1*†], *Lan Lan*[1†], *Jiawei Luo*[1], *Li Peng*[1], *Xiaolong Li*[2] *and Xiaobo Zhou*[3]

[1] West China Biomedical Big Data Center, West China Hospital/West China School of Medicine, Sichuan University, Chengdu, China, [2] Surgery, Civil Aviation Medical Center Chengdu, Chengdu, China, [3] School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States

OPEN ACCESS

Edited by:
Yonghong Peng,
Manchester Metropolitan University,
United Kingdom

Reviewed by:
Haiyan Yu,
Chongqing University of Posts and Telecommunications, China
Yifan Wu,
Guangdong Provincial Hospital of Chinese Medicine, China

***Correspondence:**
Mengjiao Li
limengchiao91@sina.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 16 June 2021
Accepted: 16 July 2021
Published: 19 August 2021

Citation:
Li M, Lan L, Luo J, Peng L, Li X and Zhou X (2021) Identifying the Phenotypic and Temporal Heterogeneity of Knee Osteoarthritis: Data From the Osteoarthritis Initiative. Front. Public Health 9:726140. doi: 10.3389/fpubh.2021.726140

Objective: Previous studies discussing phenotypic and temporal heterogeneity of knee osteoarthritis (KOA) separately have fatal limitations that either clustering patients with similar severity or assuming all knees have a single common progression pattern, which are unreliable. This study tried to uncover more reliable information on phenotypic and temporal heterogeneity of KOA.

Design: Data were from Osteoarthritis Initiative database. Six hundred and seventy-eight unilateral knees that have greater Kellgren and Lawrence (KL) grade than the contralateral knees at baseline and in all follow-up 48 months were included. Measurements of biomarkers at baseline were chosen. Subtype and Stage Inference model (SuStaIn) was applied as a subtype-progression model to identify subtypes, subtype biomarker progress sequences and stages of KOA.

Results: This study identified three subtypes which account for 15, 61, and 24% of knees, respectively. Each subtype has distinct subtype biomarker progress sequence. For knees with KL grade 0/1, 2, 3, and 4, they have different distributions on stage and 26, 53, 89, and 95% of them are strongly assigned to subtypes. When assessing whether a knee has KL (grade $\geq$ 2), subtypes and stages from subtypes-progression model (SuStaIn) are significantly better fitting than those from subtypes-only (mixture of Gaussians) (likelihood ratio = 105.59, $p = 2.2 \times 10^{-16}$) or stages-only (SuStaIn where setting $c = 1$) (likelihood ratio = 58.04, $p = 2.57 \times 10^{-14}$) model. Stages in subtypes-progression model has greater β than stages-only model. Subtypes from subtypes-progression model have no statistical significance.

Conclusions: For subtypes-progression model, stages contain more complete temporal information and subtypes are closer to real OA subtypes.

Keywords: osteoarthritis, phenotype, subtype, progression, trajectories

INTRODUCTION

Knee osteoarthritis (KOA) is recognized as a complex condition with different clinical characteristics (1–3). Most of previous studies only discussed phenotypic or temporal heterogeneity of KOA, which referred as a subtypes-only models or stages-only models.

Subtypes-only models cluster knees together into subtypes based on the similarity of the biomarker measurements (1, 4). The limitation is that those models could result in clusters of patients with similar osteoarthritis (OA) severity, which would not represent true OA subtypes. Stages-only models are always built based on regression model (5–8). The inherent assumption is that all knees have a same single common progression pattern. But disease progressions in most cases are complex and knees have phenotypic heterogeneity. Therefore, stages identified based on the above assumption have limited reliability. Some researchers tried to investigate subtypes and temporal heterogeneity together (2, 9). However, they discussed the distinct subtypes and OA severity scores separately. The above-mentioned limitations could not be avoided in previous studies.

KOA is a chronic progressive disease and has a long course. The ideally long-term frequent follow-up data are difficult to obtain. In this study, we use cross-sectional data to research the characteristics of KOA progression and phenotypic heterogeneity. There are three basic assumptions for this study. (1) Cross-sectional data contain a certain amount of temporal information. Knees have different changes in biomarker measurement, which implies the disease stage that they belong to. (2) Cross-sectional data consists of subjects at all the periods of the whole disease course. It can be roughly affirmed by the knees distributed in all the KL grades. The first two assumptions make it possible that researchers can reconstruct the trajectory of disease progression with cross-sectional data. (3) Knees from different subtypes have different trajectories of biomarker progression. Thus, the optimal number of biomarkers progression trajectories that maximizes the data likelihood represents the optimal number of subtypes.

New machine learning and deep learning methods (10–12) are brought into medical research. The recently introduced Subtype and Stage Inference model (SuStaIn) is an unsupervised machine-learning technique (13) and learns distributions of biomarker values from the data. SuStaIn calculates the optimal subtype biomarker progress sequences and the optimal number of sequences to maximize the data likelihood. The number of subtypes is represented by the optimal number of sequences. The subtype biomarker progress sequence stands for the order of the biomarker changed as disease progresses for a particular group of knees. With the subtype biomarker progress sequences, knees can be assigned into a certain subtype and progression stage. Therefore, we used SuStaIn as the subtypes-progression model to uncover phenotypic and temporal heterogeneity of KOA simultaneously. And finally, we identified 3 KOA subtypes, rebuilt the subtype biomarker progress sequences and assign each individual to a most probable subtype and stage.

MATERIALS AND METHODS

Data Description

Study Population

The data were from Osteoarthritis Initiative (OAI) database (https://nda.nih.gov/oai). OAI is a large multi-center, 10-year prospective observational cohort study. The original OAI participant recruitment and data collection process have obtained ethical approval and informed consent. No specific ethical approval was required for this study.

Previous study shows that risk of KOA increased with the incidence of contralateral knee OA (3). For the two knees of each subject which afflicted with KOA earlier and later, there may be different risk factors and disease progression patterns. As limited by available number of knees, we decided to discuss a single condition that the unilateral knees have greater Kellgren and Lawrence (KL) grade than the contralateral knees at baseline and in all the follow-up 48 months. The exclusion criteria are knees (1) having no KL grade assessment at baseline or in any follow-up visit, (2) having no complete data of radiograph and MRI image assessment at baseline. Finally, our study population includes 678 eligible knees with different KOA severity.

Obviously, to construct the subtype biomarker progress sequences, the data should cover the whole disease course of KOA. Ideally, it should contain complete biomarkers' measurements in the whole disease course of patients. However, the course of KOA, just like other chronic diseases, takes a period of decades. Following up for the whole disease course is impossible. Study the disease trajectories with only cross-sectional data is necessary. Then, we only chose the measurements of biomarkers at baseline for each knee. Because knees' KOA severity increases over time and the number of knees with mild severity reduces at follow-up time points. KOA severity of study group ranges more widely at baseline than any follow-up time points.

Biomarkers

Candidate biomarkers used in this study were the OA symptoms which were obtained through questionnaires [Western Ontario and McMaster Universities (WOMAC) pain score], quantitative radiographic readings [Joint Space Width (JSW)], quantitative MRI measures of cartilage thickness, and semi-quantitative radiographic readings (osteophytes and sclerosis, per anatomical compartment for the tibia and femur).

To simplify the model, increase the clinical utility and improve the generalization ability the most, we used the backward deletion to select the fewest biomarkers, which maximized the data likelihood (14). All the biomarkers were used to reconstruct KOA subtypes progression models shown in **Table 1** and **Figure 2A**.

Study Roadmap

The study roadmap is shown in **Figure 1**. We included 678 knees in the study population. Measurements of biomarkers at baseline were chosen. Three models (SuStaIn model, mixture of Gaussians model and SuStaIn model where setting $c = 1$) were fitted to the study population and formulated the

TABLE 1 | Detailing the biomarkers used in the study.

No.	Biomarkers	Number of z-scores events	Maximum z-score
1	**WOMAC pain score**	3	5
2	**Medial minimum JSW**	3	5
Osteophytes			
3	Tibia medial subregion	3	3
4	Tibia lateral subregion	2	3
5	Femur lateral subregion	3	3
6	Femur medial subregion	2	3
Sclerosis			
7	Femur medial subregion	2	3
8	Tibia medial subregion	1	3
Mean cartilage thickness			
9	Central medial femur condyle (external)	3	5
10	Central medial femur condyle (center)	3	5
11	Central medial femur condyle (internal)	2	2
12	Central lateral femur condyle (internal)	2	3
13	Central lateral femur condyle (external)	3	3

*The order numbers of biomarker in **Table 1** match the numbers in **Figure 2A**.*

subtypes-progression model, subtypes-only model, and stages-only model, respectively. The subtypes, subtype biomarker progress sequences and stages from the subtypes-progression model were described and assessed. Finally, we compared the subtypes and stages among the three models.

Subtypes-Progression Model

KOA Subtypes, Subtype Biomarker Progress Sequences, Stages, and SuStaIn Model

In this study, we used SuStaIn as the subtypes-progression model to identify KOA subtypes, rebuilt the subtype biomarker progress sequences and assigned an individual to a most possible subtype and stage. SuStaIn defined four important concepts. (1) Biomarker event. A biomarker event is a new change in symptom or structural lesion as KOA progression advances. Each biomarker event corresponds to a switching from a z-score to another. (2) Subtype. knees from different subtypes have different trajectories of biomarker progression. The number of biomarkers progression trajectories can be discovered in a disease represents the number of all subtypes it contains. (3) Subtype biomarker progress sequence. Each disease subtype has a particular progress course. The course of each disease subtype progression can be depicted by the order in which the biomarker events occur as each disease subtype progresses, which is called as subtype biomarker progress sequence. (4) Stage. Each disease subtype progression advances from 0-th stage to S-th stage. S is the number of all biomarker events. The i-th stage that a knee belongs to is defined as a specific state that the previous i events of the sequence have occurred. Occurrence of an event indicates that the disease advances in a biomarker and disease progression switches from a stage to the next.

SuStaIn is an unsupervised machine-learning technique and doesn't rely on a priori staging or subtype. SuStaIn is a mixture of linear z-score models. It describes the subtype biomarker progress sequence as a linear z-score model, which is an improved model of the original event-based model (EBM) (15, 16). In EBM, each event represents the switch of a biomarker from a normal to an abnormal level. SuStaIn reformulates the events to make them correspond to the continuous linear accumulation and each event of a biomarker represents change from one z-score to another. SuStaIn simultaneously optimizes the number of subtypes, subtype biomarker progress sequence, and the posterior distributions of both. What's more, SuStaIn estimates the probability of assignment to a most probable subtype and stage, respectively, for each knee. The most likely biomarker event sequences are ones that maximizes the data likelihood. The optimal number of subtype biomarker progress sequences that maximize the data likelihood represent the optimal number of subtypes. We fitted SuStaIn with python (version 3.7). The source code of SuStaIn can be acquired on https://github.com/EuroPOND/pySuStaIn.

Data Pre-processing

Every biomarker measurement was expressed as a z-score relative to the control group. Since this, the corresponding z-score can describe the abnormal degree of each biomarker measurement from study population relative to control group. Inclusion criterion for the OAI control group are (1) no pain, aching or stiffness in both knee in the past year; (2) no radiograph OA; (3) no eligibility risk factors; (4) age $\leq$ 70 years. This study had an additional exclusion criterion: incomplete data on radiographic and MRI image assessment at baseline.

With KOA progress, medial minimum JSW and mean cartilage thickness decrease and their z-scores became negative. We took the negative value of the z-scores for convenience, so that all the z-scores would increase as KOA severity increasing.

Input

A biomarker event is a new change in symptom or structural lesion as KOA progression advances. Each biomarker event corresponds to a switching from a z-score to another. The z-score events of each biomarker were the most important input. For each biomarker, the z-score events initially include z-scores of 1, 2, 3, and 5. However, some z-score events had low occurrence frequency in the disease progression, thus fewer than 10 knees had greater than that z-score. To simplify the model, we excluded those z-score events. Finally, 32 z-score events were included from the 13 biomarkers.

The maximum z-score means the final stage of the progression. If the maximum z-score event was 1, 2, 3, and 5, the input "maximum z-score" was set to be 2, 3, 5, and 7, respectively. Detail of z-score of each biomarker is shown in **Table 1**.

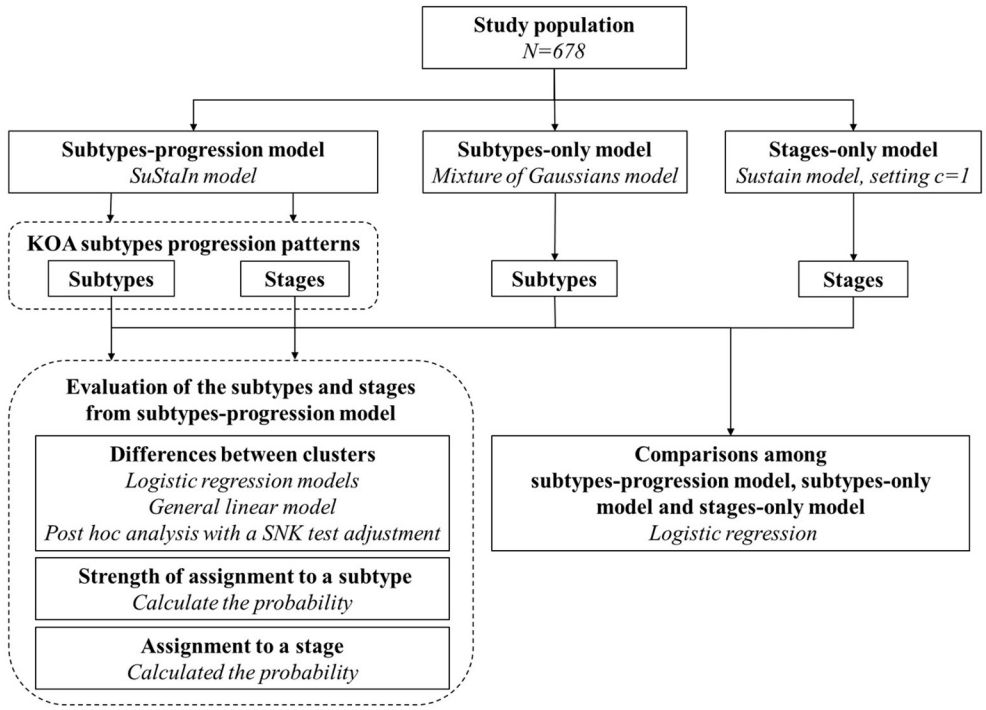


FIGURE 1 | Study roadmap for our study.

Evaluation of Subtypes and Stages From Subtype Biomarker Progress Sequences

We used SuStaIn as the subtypes-progression model to identify the KOA subtypes, subtype biomarker progress sequences and the probability of assignment to a most probable subtype and stage, respectively, for each knee.

We used 10-fold cross-validation to gain the optimal number of subtypes by the Cross-Validation Information Criterion (CVIC) (16). The CVIC was defined as CVIC $= -2 \times \log [P(X|M)]$, where $P(X|M)$ is the probability of the data X for a particular subtypes-progression model M.

We tested the differences between the subtypes for a specific biomarker with R (version 3.6.3). The logistic regression was used for binary measurements and a general linear model for continuous or ordered categorical measurements. Then we used *post-hoc* analysis with a SNK test adjustment to test for which subtypes the measurements were different.

We measured the strength of assignment to one of the subtypes in KOA subtypes progression. A strong assignment was defined as that the maximum probability of assigning to a particular subtype is 1.5 times greater than any other two subtypes.

We surveyed the consistency of KL grade and stages from KOA subtypes progression model. KL grade represented the temporal state of knees with KOA roughly. We grouped the knees by KL grade and estimated the probability of assignment to a stage.

Comparisons Among Subtypes-Progression Model, Subtypes-Only Model, and Stages-Only Model

Subtypes-Only Model and Stages-Only Model

We used SuStaIn as the subtypes-progression model to identify the KOA subtypes, subtype biomarker progress sequences, the probability of assignment to a most probable subtype, and stage, respectively, for each knee. The subtypes-only model and stages-only model can also assign knees to a subtype or stage, respectively. The subtype or stage that a knee is assigned reflects the phenotypic or temporal information from our models.

We formulated subtypes-only and stages-only models as close as possible to the subtypes-progression model. Thus, subtypes and stages from the three models (subtypes-only model and stages-only model, subtypes-progression model) are comparable.

In this study, the subtypes-only model was fitted to OA (KL grade $\geq$2) with a mixture of Gaussians model. The stages-only model was formulated by SuStaIn model, setting the subtype number to 1.

Comparisons Among Subtypes and Stages From Three Models

To compare the subtypes and stages from the three models, we put forward a task that separates the knees with/without doubtful KOA (KL grade 0/1) from those with KOA (KL grade $\geq$2). This task can test the ability of the stages from the three models that

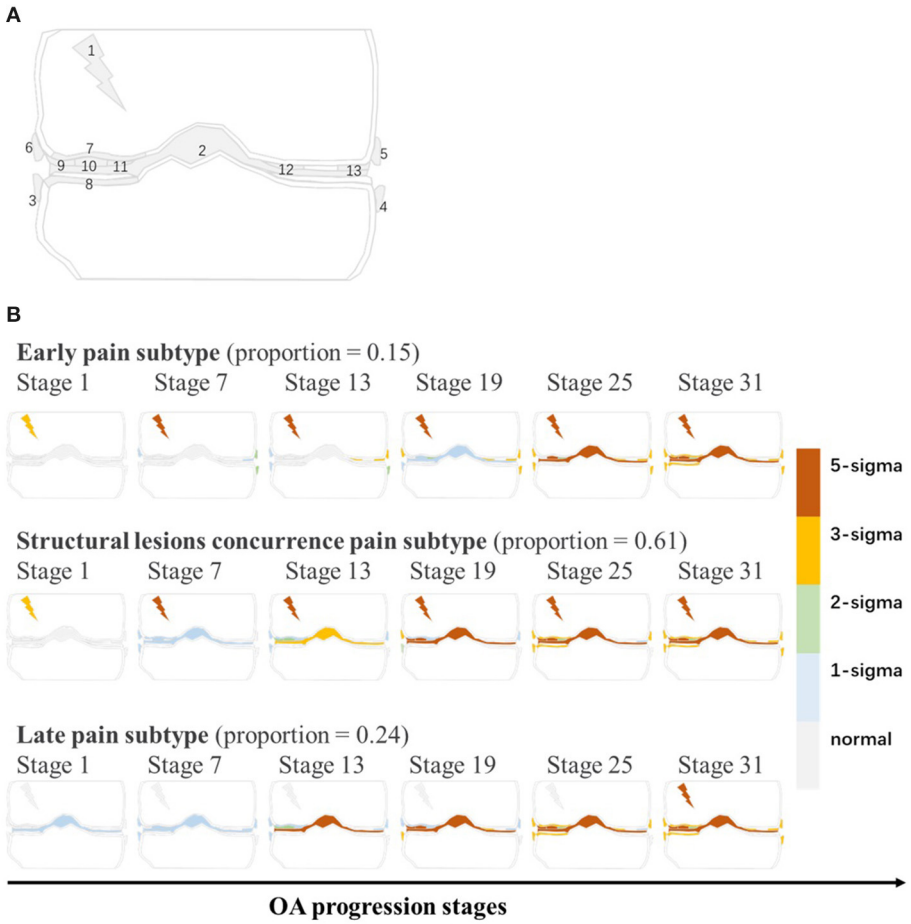


FIGURE 2 | (A) The graphical representation of all biomarkers. Gray line is coronal sketch map of knee joint. Numbers in this figure match the order numbers of biomarkers in **Table 1**. **(B)** KOA subtypes and subtype biomarker progress sequences identified by subtypes-progression model. Proportion of knees assigned to each subtype are shown. Each row represents an biomarkers changes order as KOA stages advance. Biomarkers from different stages reach different z-scores relative to control group. At each stage color in each region indicates level of severity of pain or lesions: gray is unaffected; blue is mildly affected (z-score of 1), and so on.

separates the knees in time-perspective. The more benefit the stages contribute to the task, the more temporal information are included in the stages. However, if the subtypes have more contributions to the task, it has negative meaning. It indicates that those subtypes include temporal heterogeneity and tend to cluster the knees with similar KOA severity and don't represent true KOA subtypes.

We used logistic regression to compare the subtypes and stages from the three models by the task of separating no or doubtful KOA (KL grade 0/1) from KOA (KL grade $\geq$2). The input variables were stages and subtypes from the three models and demographic factors: gender, age, BMI and injury. Likelihood ratio comparison between two models was used to assess the goodness of fit of them (17). Statistical significance was set as $P < 0.05$. All were analyzed with R (version 3.6.3).

RESULT

There were 678 unilateral knees included, which are from subjects with average age 62.15 years old, 55.01% female, average BMI of 23.1 and 20.65% injury in this study. KL grade 0/1, 2, 3, and 4 accounted for 7.47, 34.22, 40.56, and 17.85%, respectively. 140 knees have injury which was defined as ever injured badly enough to limit ability to walk for at least 2 days.

KOA Subtypes and Subtype Biomarker Progress Sequences

As is shown in **Figure 2**, the subtypes-progression model identified three subtypes. Each subtype had different clinical symptoms and structural lesions at different progression stages. We termed three subtypes as early pain, structural lesions concurrence pain and late pain, which account for 15, 61, and 24% of the knees. Early pain subtype could be described as a mild subtype. Serious pain occurs at first stages and osteophytes follows. Not until the latter half stages, other structural lesions occur. In structural lesions concurrence pain, serious pain also occurs at first stages. but all the structural lesions appear and progress soon and almost distribute in almost all the rest stages. Late pain subtype is very similar to the former, but the pain hides till the last stages and structural lesions occur at very early stages.

TABLE 2 | Baseline demographics and risk factors for three subtypes and differences between the measurements of all biomarkers used for subtype definitions.

	Stages *P*-value	Subtypes *P*-value	Subtype[a]		
			Early pain ($n = 78$)	Structural lesions concurrence pain ($n = 419$)	Late pain ($n = 181$)
KL grade, *n* (%)					
0/1			0 (0)^{S,L}	36 (8.59)^{E,L}	33 (18.23)^{E,S}
2			17 (21.80)^{S,L}	149 (35.56)^{E}	79 (43.65)^{E}
3			40 (51.28)^{L}	175 (42.72)^{E,L}	58 (32.04)^{E}
4			21 (26.92)^{E,S}	59 (14.08)^{E,L}	11 (6.08)^{E,S}
Injury, yes, *n* (%)	0.824	0.633	14 (17.95)	88 (21.00)	38 (20.99)
Gender, female, *n* (%)	0.294	0.419	50 (64.10)	220 (52.51)	103 (56.91)
Age, years, mean (standard deviation)	<0.001	0.023	62.46 (8.67)	61.75 (9.20)	62.96 (9.54)
BMI, kg/M^2, mean (standard deviation)	0.393	−0.004	29.36 (4.79)^{L}	29.90 (4.79)^{L}	28.02 (4.13)^{E,S}
WOMAC pain score	<0.001	<0.001	4.96 (3.52)^{L}	4.42 (3.27)^{L}	0.02 (0.30)^{E,S}
Medial minimum JSW	<0.001	<0.001	5.53 (1.24)^{S,L}	3.1 (1.59)^{E,L}	3.9 (1.30)^{E,S}
Osteophytes					
Tibia medial	<0.001	<0.001	0.73 (0.77)^{S,L}	1.21 (0.86)^{E}	0.85 (0.71)^{E}
Tibia lateral	<0.001	<0.001	2.17 (0.90)^{S,L}	0.62 (0.79)^{E}	0.49 (0.78)^{E}
Femur medial	<0.001	0.778	1.49 (1.29)^{S,L}	1.26 (1.13)^{E,L}	0.77 (1.02)^{E,S}
Femur lateral	<0.001	<0.001	2.40 (0.76)^{S,L}	0.69 (0.90)^{E}	0.59 (0.93)^{E}
Sclerosis					
Femur medial	<0.001	<0.001	0.10 (0.38)^{S,L}	1.12 (0.94)^{E,L}	0.57 (0.76)^{E,S}
Tibia medial	<0.001	<0.001	0.15 (0.40)^{S,L}	1.12 (0.95)^{E,L}	0.53 (0.80)^{E,S}
Mean cartilage thickness					
Central medial femur (external)	<0.001	<0.001	1.60 (0.42)^{S,L}	1.08 (0.49)^{E,L}	1.25 (0.35)^{E,S}
Central medial femur (center)	<0.001	<0.001	2.39 (0.55)^{S,L}	1.62 (0.75)^{E,L}	1.96 (0.54)^{E,S}
Central medial femur (internal)	<0.001	<0.001	2.10 (0.50)^{S,L}	1.80 (0.47)^{E}	1.93 (0.38)^{E}
Central lateral femur (internal)	<0.001	0.0522	1.70 (0.52)^{S,L}	1.86 (0.35)^{E}	1.80 (0.36)^{E}
Central lateral femur (external)	0.002	<0.001	1.09 (0.59)^{S,L}	1.65 (0.35)^{E}	1.53 (0.38)^{E}

[a]Indices is the index number of the subtypes. The subtypes with corner mark [a] indicate that the specific subtype is significantly different from the subtypes represented by index numbers ($P < 0.05$). E, early pain; S, structural lesions concurrence pain; L, late pain.

Evaluation of Subtypes and Stages From Subtype Biomarker Progress Sequences

Demographic and Risk Factors Differences Between Subtypes

All subtypes contained knees of the entire range of the KL grade, except for the early pain subtype on KL grade 0/1 (**Table 2**). **Table 2** shows that early pain subtype had the highest proportion of female (65%) than other two subtypes. BMI in late pain subtype was significantly lowest. Age, injury proportion and gender proportion had no significant difference between the three subtypes.

After correction for stages, all biomarkers except osteophytes femur medial compartment showed significant differences between the subtypes. *Post-hoc* analysis showed that early pain subtype was mainly different with others. Early pain subtype had more changes in WOMAC pain score. What's more, structural lesions between the lateral and medial compartment were different. More severe lesions occurred in osteophytes from tibia lateral compartment and mean cartilage thickness in lateral femur compartment. Less changes occurred in medial minimum JSW, sclerosis and mean cartilage thickness in central medial subregion. Further, the characteristics between the other subtypes were different. Late pain subtype had slighter change in most structural lesions (WOMAC pain score, medial minimum JSW, osteophytes, sclerosis, and in mean cartilage thickness of central lateral femur subregion).

Strength of Assignment to a Subtype

As is shown in **Figure 3**, for OA knees with KL grade 0/1, 2, 3, and 4, the strong assignment to subtypes were 26, 53, 89, and 95%. With disease progress, the strength of assignment in KOA subtypes increased. Moreover, even at early disease stages (KL grade 0/1), 26% of the knees were strongly assigned to a subtype.

Probability of Assignment to a Stage

To evaluate the assignment to a particular stage, we estimated the probability knees from each KL grade belonging to each of the stages. As is shown in **Figure 3**, the knees from different KL grade had different distributions in the stages.

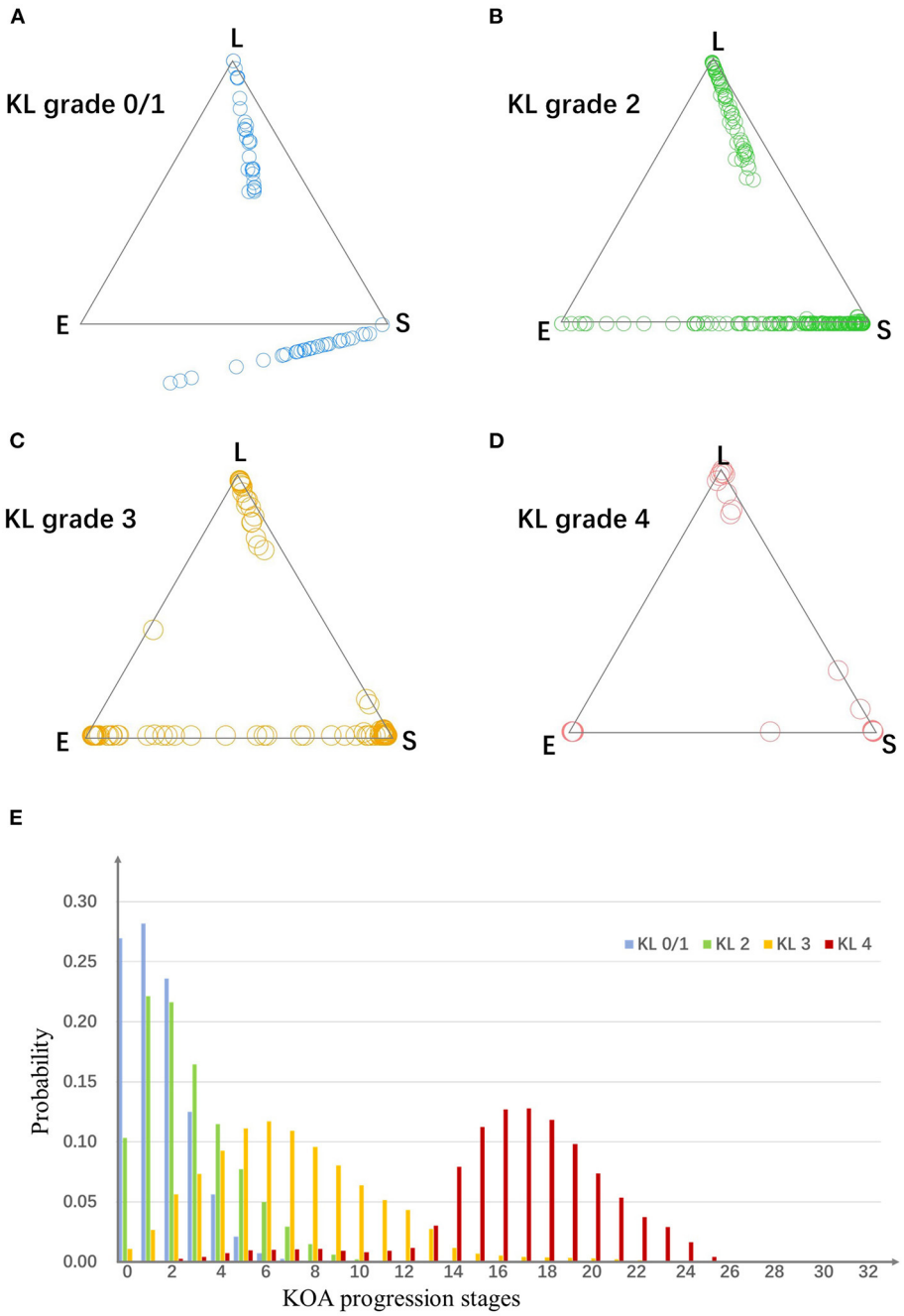


FIGURE 3 | (A–D) The assign ability of subtypes and stages. Plots in four scatter plots indicate knees from four KL grade groups. For a triangle scatter plot, each corner indicates a probability of 1 of assigning to a particular subtype, and 0 for the other two subtypes; the center point of the triangle indicates a probability of 1/3 of assigning to each subtype. **(E)** The probability of knees from each KL grade group belong to each stage. E, early pain; S, structural lesions concurrence pain; L, late pain.

Comparisons With the Stages-Only Model and Subtypes-Only Model

We used logistic regression model to separate knees with/without doubtful KOA (KL grade 0/1) from those with KOA (KL grade ≥2). **Table 3** shows that stages ($p = 1.08 \times 10^{-8}$) and subtypes ($p = 0.332$) from subtypes-progression model had significant hazards.

We compared subtypes-progression model, subtypes-only and stages-only models with likelihood ratio tests. Subtypes-progression model was significantly better fit than

TABLE 3 | Results for comparing the subtypes and stages from subtypes-progression model with those from subtypes-only and stages-only with logistic regression model by separating knees with no or doubtful KOA (KL grade 0/1) from those with KOA (KL grade ≥2).

	Subtypes-progression model		Stages-only model		Subtypes-only model	
	β (95% CI)	Adjusted *P*-value	β (95% CI)	Adjusted *P*-value	β (95% CI)	Adjusted *P*-value
Intercept	2.82 (−6.26, 0.56)	0.0974	−2.79 (−5.48, −0.14)	0.040	−0.89 (−3.75, 2.13)	0.544
Stages	0.70 (0.49, 0.98)	<0.001	0.32 (0.22, 0.43)	<0.001	—	—
Subtypes	0.29 (−0.29, 0.89)	0.314	—	—	−2.00 (−3.82, −0.81)	0.006
Injury	0.33 (−0.39, 1.12)	0.400	0.43 (−0.25, 1.20)	0.247	0.48 (−0.12, 1.23)	0.181
Gender	0.40 (−0.98, 0.17)	0.158	−0.04 (−0.58, 0.49)	0.872	−0.10 (−0.62, 0.42)	0.729
Age	0.01 (−0.02, 0.04)	0.476	0.03 (0.00, 0.06)	0.029	0.04 (0.01, 0.06)	0.141
BMI	0.05 (−0.01, 0.13)	0.086	0.05 (−0.01, 0.12)	0.089	0.10 (0.04, 0.16)	<0.001

CI, confidence intervals.

subtypes-only (likelihood ratio $= 105.59$, $p = 2.2 \times 10^{-16}$) and stages-only (likelihood ratio $= 58.04$, $p = 2.57 \times 10^{-14}$) models. A likelihood ratio of above 1 shows that, for distinguish knees with no or doubtful KOA (KL grade 0/1) from those with OA (KL grade ≥2), the subtypes and stages of KOA subtypes progression model provided a significantly better fit than a subtypes-only or stages-only model.

Table 3 shows that the stages in subtypes-progression model had greater β than stages-only model. Rather than subtypes-only model, the subtypes from subtypes-progression model had no statistical significance.

DISCUSSION

Some studies tried to describe the temporal or phenotypic heterogeneity for KOA. However, those studies that only explain the temporal progression based on the assumption that all the knees came from a single disease progress sequence (1, 4). As KOA is a disease with complex clinical characteristics, this assumption may not hold. Some studies only discussed about phenotypic heterogeneity and don't account for temporal heterogeneity (5–8). Subtypes identified by those studies tend to cluster the knees with similar OA severity, which don't represent true OA subtypes. In our study, we used SuStaIn as a subtypes-progression model to study the temporal or phenotypic heterogeneity simultaneously and constructed a reliable picture of how the lesions spread from a distinct region over the rest of the knee in each subtype.

Because of the limitation of available number of knees, we only discuss a single condition that the unilateral knees have greater KL grade than the contralateral knees at both baseline and all the follow-up 48 months. We fitted the SuStaIn model to 678 knees to identify KOA subtypes and subtype biomarker progress sequences. The biomarkers included the WOMAC pain score, medial minimum JSW, quantitative MRI measures of cartilage thickness, and semi-quantitative radiographic readings (osteophytes and sclerosis, per compartment for the tibia and femur). The subtypes-progression model identified three subtypes and each subtype had its distinct biomarker progress sequence. Three subtypes had different characteristics and were termed as early pain, structural lesions concurrence pain and late pain. The severity of all biomarker increased with greater stages.

Our results are agreed with previous studies, such as bigger pain sore is always associated with severer osteophytes (18–23), and narrower medial minimum JSW is always associated with thinner mean cartilage thickness (24–28).

Between lateral and medial subregion of each subtype, the changes in mean cartilage thickness existed inconsistency. Early pain subtype had significantly slightest changes in medial subregion and significantly greatest changes in lateral subregion. Structural lesions concurrence pain subtype had opposite changes. Late pain subtype had medium changes in both subregion. Some studies found that greater BMI is associated with incident medial tibiofemoral OA (3) and more serious changes in lateral subregion in mean cartilage thickness, which is consistent with our results.

The existence of phenotypic heterogeneity of KOA is proved by our study. When we measured the strength of a knee's assignment to a given subtype, it showed strong assignment of KOA patients to the subtypes. Therefore, explaining heterogeneity in this study about KOA progression is very necessary.

The subtypes and stages we identified have power to separate the knees with phenotypic or temporal heterogeneity. At no or doubtful KOA (KL grade 0/1), many knees gather around the vertices of the triangles. It shows the subtypes are so effective that have the ability of identifying knees even in very early stages. Besides, the stages are certified to have the power to separate the knees with different disease severity. The probabilities of knees with each KL grade belonging to each of the stages show that the distribution of the stages differ between KL grades.

The subtypes and stages from subtypes-progression model performed significantly better than subtypes-only and stages-only models. We used logistic regression to compare the subtypes and stages from the three models by the task of separating no or doubtful KOA (KL grade 0/1) from KOA (KL grade ≥2). With the temporal task of separating early disease stages (KL grade 0/1) from OA (KL grade ≥2), it shows that subtypes and stages from subtypes-progression model are close to true KOA subtypes and stages. Bigger regression coefficient of stages from subtypes-progression model shows that they contain more

complete temporal information. Furthermore, rather than those from subtypes-only model, subtypes from subtypes-progression model have no contribution to this temporal task. It suggests that the subtypes from subtypes-progression model contain hardly any temporal heterogeneity and are close to true OA subtypes.

The subtypes and stages identified by this study have clinical practice value. First, with our subtypes-progression model, each knee can be assigned into certain subtype, suggesting particular disease characteristics. So that, we can help doctor to identify the significant features of each knee and make more proper treatment plan. Second, whole KOA progress course is divided into more detailed stages. They offer a mini scale for doctor to learn disease progress statue of a knee. What's more, the biomarker progress sequence can show the progress pathway for each knee. The doctors and patients can foresee their subsequent biomarker changes and estimate if a knee maintain same stage over a period of time. It has a significant meaning in chronic disease management. And finally, the study method can also be expanded to carry out the research of other chronic disease.

There are some limitations in this study. As being confined to the quantity of the knees available, the study population only included the knee that afflicted with KOA earlier and had greater KL grade for every knee. So our results may only act as a reference for the knee with more serious OA condition of the subject. In addition, the biomarkers only contain the pain score and image assessment data. In future work, we can study other groups of knees that afflicted with KOA later and remain smaller KL grade or alternately having smaller KL grade, and try to analyze more categories of biomarkers, e.g., biochemical biomarker measurements from serum and urine samples, so that we can learn the KOA progression in wider range.

CONCLUSION

The subtypes-progression model identifies three subtypes and each subtype has its distinct biomarker progress sequence. There exists phenotypic heterogeneity of KOA, bigger pain sore is always associated with severer osteophytes and the changes in mean cartilage thickness exist inconsistency between lateral and medial subregions of each subtype. The subtypes and stages from subtypes-progression model have power to separate the knees with phenotypic or temporal heterogeneity and perform significantly better than those from stages-only model and subtypes-only model. In a word, with subtypes-progression model, stages contain more complete temporal information and subtypes are close to real OA subtypes.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://nda.nih.gov/oai/.

ETHICS STATEMENT

The original OAI participant recruitment and data collection process have obtained ethical approval and informed consent. No specific ethical approval was required for this study.

AUTHOR CONTRIBUTIONS

ML participated in the design of the study, analysis and interpretation of the data, and drafting of the article. LL participated in analysis of data, manuscript preparation, and critical revision of the article for important intellectual content. JL participated in the design of the study and analysis of the data. LP participated in analysis and interpretation of the data. XZ participated in design of the study and critical revision of the article for important content. XL participated in interpretation of the data and drafting of the article. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Center of Excellence-International Collaboration Initiative Grant (No. 139170052) and the 1·3·5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (No. ZYJC18010). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the OAI for providing data for this work.

REFERENCES

1. Waarsing JH, Bierma-Zeinstra SM, Weinans H. Latent class cluster analysis shows four distinct subtypes of knee OA: data from the osteoarthritis initiative. *Osteoarthritis Cartilage.* (2012) 20:S177–8. doi: 10.1016/j.joca.2012.02.272
2. Waarsing JH, Bierma-Zeinstra SM, Weinans H. Distinct subtypes of knee osteoarthritis: data from the osteoarthritis initiative. *Rheumatology (Oxford).* (2015) 54:1650–8. doi: 10.1093/rheumatology/kev100
3. Wei J, Gross D, Lane NE, Lu N, Wang M, Zeng C, et al. Risk factor heterogeneity for medial and lateral compartment knee osteoarthritis: analysis of two prospective cohorts. *Osteoarthritis Cartilage.* (2019) 27:603–10. doi: 10.1016/j.joca.2018.12.013
4. Kittelson AJ, Stevens-Lapsley JE, Schmiege SJ. Determination of pain phenotypes in knee osteoarthritis: a latent class analysis using data from the osteoarthritis initiative. *Arthritis Care Res (Hoboken).* (2016) 68:612–20. doi: 10.1002/acr.22734
5. Cibere J, Sayre EC, Guermazi A, Nicolaou S, Kopec JA, Esdaile JM, et al. Natural history of cartilage damage and osteoarthritis progression on magnetic resonance imaging in a population-based cohort with knee pain. *Osteoarthritis Cartilage.* (2011) 19:683–8. doi: 10.1016/j.joca.2011.02.008
6. Collins JE, Losina E, Nevitt MC, Roemer FW, Guermazi A, Lynch JA, et al. Semiquantitative imaging biomarkers of knee osteoarthritis

progression: data from the foundation for the national institutes of health osteoarthritis biomarkers consortium. *Arthritis Rheumatol.* (2016) 68:2422–31. doi: 10.1002/art.39731
7. de Lange-Brokaar BJ, Bijsterbosch J, Kornaat PR, Yusuf E, Ioan-Facsinay A, Zuurmond AM, et al. Radiographic progression of knee osteoarthritis is associated with MRI abnormalities in both the patellofemoral and tibiofemoral joint. *Osteoarthritis Cartilage.* (2016) 24:473–79. doi: 10.1016/j.joca.2015.09.021
8. Yoo JJ, Kim DH, Kim HA. Risk factors for progression of radiographic knee osteoarthritis in elderly community residents in Korea. *BMC Musculoskelet Disord*. (2018) 19:80. doi: 10.1186/s12891-018-1999-5
9. Halilaj E, Le Y, Hicks JL, Hastie TJ, Delp SL. Modeling and predicting osteoarthritis progression: data from the osteoarthritis initiative. *Osteoarthritis Cartilage*. (2018) 26:1643–50. doi: 10.1016/j.joca.2018.08.003
10. Zeng N, Wang Z, Liu W, Zhang H, Hone K, Liu X. A dynamic neighborhood-based switching particle swarm optimization algorithm. *IEEE Trans Cybern*. (2020) 1:1–12. doi: 10.1109/TCYB.2020.3029748
11. Zeng N, Wang Z, Zineddin B, Li Y, Du M, Xiao L, et al. Image-based quantitative analysis of gold immunochromatographic strip *via* cellular neural network approach. *Med Imaging*. (2014) 33:1129–36. doi: 10.1109/TMI.2014.2305394
12. Zeng N, Wang Z, Zhang H, Kim K-E, Li Y, Liu X. An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips. *IEEE Trans Nanotechnol*. (2019) 18:819–29. doi: 10.1109/TNANO.2019.2932271
13. Young AL, Marinescu RV, Oxtoby NP, Bocchetta M, Yong K, Firth NC, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat Commun*. (2018) 9:4273. doi: 10.1038/s41467-018-05892-0
14. Abe S. *Modified Backward Feature Selection by Cross Validation*. Bruges: European Symposium on Esann (2005). p. 163–8.
15. Fonteijn HM, Modat M, Clarkson MJ, Barnes J, Lehmann M, Hobbs NZ, et al. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage*. (2012) 60:1880–9. doi: 10.1016/j.neuroimage.2012.01.062
16. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comput*. (2013) 24:997–1016. doi: 10.1007/s11222-013-9416-2
17. Royle JA, Dorazio RM. 2 - Essentials of statistical inference. In: Royle JA, Dorazio RM, editors. *Hierarchical Modeling and Inference in Ecology*. Academic Press: San Diego (2009). p. 27–82.
18. Aylward S, Hadjiiski LM, Galván-Tejada JI, Celaya-Padilla JM, Martínez-Torteya A, Rodriguez-Rojas J, et al. Wide association study of radiological features that predict future knee OA pain: data from the OAI. In: *Medical Imaging 2014: Computer-Aided Diagnosis*. (2014). p. 93539.
19. Javaid MK, Kiran A, Guermazi A, Kwoh CK, Zaim S, Carbone L, et al. Individual magnetic resonance imaging and radiographic features of knee osteoarthritis in subjects with unilateral knee pain: the health, aging, and body composition study. *Arthritis Rheum*. (2012) 64:3246–55. doi: 10.1002/art.34594
20. Javaid MK, Lynch JA, Tolstykh I, Guermazi A, Roemer F, Aliabadi P, et al. Pre-radiographic MRI findings are associated with onset of knee symptoms: the most study. *Osteoarthritis Cartilage*. (2010) 18:323–8. doi: 10.1016/j.joca.2009.11.002
21. Sayre EC, Guermazi A, Esdaile JM, Kopec JA, Singer J, Thorne A, et al. Associations between MRI features versus knee pain severity and progression: data from the Vancouver longitudinal study of early knee osteoarthritis. *PLoS ONE*. (2017) 12:e0176833. doi: 10.1371/journal.pone.0176833
22. Niethammer TR, Valentin S, Gülecyüz MF, Roßbach BP, Ficklscherer A, Pietschmann MF, et al. Bone marrow edema in the knee and its influence on clinical outcome after matrix-based autologous chondrocyte implantation. *Am J Sports Med*. (2015) 43:1172–9. doi: 10.1177/0363546515573935
23. Zhai G, Cicuttini F, Ding C, Scott F, Garnero P, Jones G. Correlates of knee pain in younger subjects. *Clin Rheumatol*. (2007) 26:75–80. doi: 10.1007/s10067-006-0248-8
24. Eckstein F, Maschek S, Roemer FW, Duda GN, Sharma L, Wirth W. Cartilage loss in radiographically normal knees depends on radiographic status of the contralateral knee - data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage*. (2019) 27:273–7. doi: 10.1016/j.joca.2018.10.006
25. Eckstein F, Wirth W, Hunter DJ, Guermazi A, Kwoh CK, Nelson DR, et al. Magnitude and regional distribution of cartilage loss associated with grades of joint space narrowing in radiographic osteoarthritis - data from the Osteoarthritis Initiative (OAI). *Osteoarthritis Cartilage*. (2010) 18:760–8. doi: 10.1016/j.joca.2009.12.009
26. Pelletier J-P, Raynauld J-P, Berthiaume M-J, Abram F, Choquette D, Haraoui B, et al. Risk factors associated with the loss of cartilage volume on weight-bearing areas in knee osteoarthritis patients assessed by quantitative magnetic resonance imaging: a longitudinal study. *Arthritis Res Ther*. (2007) 9:R74. doi: 10.1186/ar2272
27. Wirth W, Buck R, Nevitt M, Le Graverand MPH, Benichou O, Dreher D, et al. MRI-based extended ordered values more efficiently differentiate cartilage loss in knees with and without joint space narrowing than region-specific approaches using MRI or radiography - data from the OA initiative. *Osteoarthritis Cartilage*. (2011) 19:689–99. doi: 10.1016/j.joca.2011.02.011
28. Wirth W, Maschek S, Roemer FW, Sharma L, Duda GN, Eckstein F. Radiographically normal knees with contralateral joint space narrowing display greater change in cartilage transverse relaxation time than those with normal contralateral knees: a model of early OA? - data from the Osteoarthritis Initiative (OAI). *Osteoarthritis Cartilage*. (2019) 27:1663–8. doi: 10.1016/j.joca.2019.06.013

Conflict of Interest: The authors declare that this study received funding from the Center of Excellence-International Collaboration Initiative Grant (No. 139170052) and the 1.3.5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (No. ZYJC18010). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

frontiers in Medical Technology

ORIGINAL RESEARCH
published: 25 August 2021
doi: 10.3389/fmedt.2021.666671

Improvements in Medical System Safety Analytics for Authentic Measure of Vital Signs Using Fault-Tolerant Design Approach

Prasadraju Lakkamraju[1]*, *Madhu Anumukonda*[1] and *Shubhajit Roy Chowdhury*[2]

[1] Center for Very Large Scale Integration and Embedded Systems Technology, International Institute of Information Technology (IIIT), Hyderabad, India, [2] School for Computing and Electrical Engineering, Indian Institute of Technology (IIT), Mandi, India

The study presents a novel design method that improves system availability using fault-tolerant features in a non-invasive medical diagnostic system. This approach addresses the effective detection of functional faults, improves the uninterruptible system operating period with reduced false alarms, and provides an authentic measure of vital cardiac signs using diverse multimodal sensing elements like the photoplethysmogram (PPG) and the ECG. Most systems rely on a 1oo1 (one-out-of-one) design method, which inherently limits accuracy in existing practice. In this proposed approach, the quality of segregated authentic vital sign measured values could tremendously benefit the performance of resourceful nursing with negligible alarm fatigue and predict illness more accurately. The system builds upon the selected 2oo2 (two-out-of-two) safety-related design architecture and is evaluated with implemented functions like the fault detection and identification logic, the correlation coefficient-based safety function, and the fault-tolerant safe degradation switching mechanism for accurate measurements. The system was tested on 50 adults of various age groups. The analyzed captured data showed highly accurate vital sign data in this fault-tolerant approach with reduced false alarms. The proposed design method evaluated safety-related mechanisms along with a combination of the same and diverse sensors in a medical monitoring device, showing more reliable functioning of the system and authentic data for better nursing. This design approach showed a 45–55% increased improvement in system availability, thus allowing for accurate and uninterruptable tracking of vital signs for better nursing during critical times in the ICU.

Keywords: correlation, electrocardiography, fault-tolerant systems, fault detection, field programmable gate array, multimodal sensor systems, photoplethysmography, safety analytics

OPEN ACCESS

Edited by:
Yonghong Peng, Manchester Metropolitan University, United Kingdom

Reviewed by:
Jorge Igual, Universitat Politècnica de València, Spain
Guanghao Sun, The University of Electro-Communications, Japan

***Correspondence:**
Prasadraju Lakkamraju
prasadaraju.lvr@research.iiit.ac.in

Specialty section:
This article was submitted to Medtech Data Analytics, a section of the journal Frontiers in Medical Technology

Received: *10 February 2021*
Accepted: *28 July 2021*
Published: *25 August 2021*

Citation:
Lakkamraju P, Anumukonda M and Chowdhury SR (2021) Improvements in Medical System Safety Analytics for Authentic Measure of Vital Signs Using Fault-Tolerant Design Approach. Front. Med. Technol. 3:666671. doi: 10.3389/fmedt.2021.666671

INTRODUCTION

Advanced smart medical system engineering for safety-critical medical applications requires the consideration of several dependable features, such as the functional safety, availability, and reliability of the engineered system in signal processing algorithms (1, 2). With this consideration, the advancement in high-performance electronics and sensors can become affordable and, thus, increase the usage of these smart medical systems in clinical environments, such as

performing critical patient treatments like robotic surgeries and medication supervisions through the continuous monitoring of patients (3–5). These smart, intelligent multimodal computational electronics, along with system portability, come with a significant increase in system complexity and bring in major challenges like functional safety, reliability of measurements, and patient safety. This type of safety-critical smart medical device is often subject to an insignificant number of failures with potentially catastrophic impacts on patients. A previous study (6) on medical device recalls between 2006 and 2011 shows a 69.8% increase in the number of products being recalled and a 103.3% increase in the number of adverse events to patients like incorrect medications and deaths, where the majority of the recalls were due to software faults. In a recent report (7, 8) for the year 2018, a significant spike of 126% in product recalls informed the U.S. Food and Drug Administration (FDA) that most of the causes were due to software faults. These detected spurious faults always interrupt the intended system functionality and may generate safe alarms during critical periods. However, the cause of numerous spurious alarms, incorrect measurements of vital signs, single point of failure (SPOF), and undetected faults may cause dangerous effects on patients in nursing care. In the past, some of the medical system designs were adopted with safety feature improvements like data fusion and voting techniques (9–14). However, all these techniques have their limitations and lack the mitigation of challenges, such as the reduction of false alarms by effectively detecting the actual faults, improvements in fault tolerability having built-in safe degradable mechanisms, and providing authentic measurements through the safe computation of vital signs without interruption to system operability.

This study focused on presenting a safety-related, fault-tolerant design approach to improve the safety features in addressing the challenges related to the detection of software faults in a non-invasive medical device. The safety features include: (1) effective fault detection function and (2) fault-tolerability with a safe degradable function. These two functions were implemented and evaluated using the proposed conceptualized 2oo2 (two-out-of-two) safety-related design architecture. This approach provided uninterruptable operability of the system by removing its faults. A detailed framework proposed with five configurable conceptualized safety design architectures based on composite fail-safety techniques were realized and evaluated. A safe degradable function was implemented using Karl Pearson's coefficient of correlation method. Thus, the results were analyzed and evaluated toward the aims of reducing alarms for better nursing with reduced alarm fatigue and providing improved authentic vital signs measured values for better predictability of illnesses and proper assessment of the condition of patients.

The systematic design assurance guidelines (15–17) were followed in the implementation of the proposed architectures. Detailed experimental research activity was performed to evaluate the results and analyzed in detail for safety improvements. The conceptual safety-related 2oo2 design architecture was reused from previous studies (1) in implementing this proposed conceptual fault-tolerant safety architecture for system functional safety evaluation. This study focused on evaluating vital cardiac signs, such as heart rate (HR), using the proposed design approach. The assessment included the continuity of the measurement of HR-values and the authenticity of the measured values. Two diverse sensors were selected to detect biomedical signals, with these signals being based on the physical media of light and electric potential. These sensors, which consist of light-emitting diodes (LEDs) and an optical detector, were used to detect the photoplethysmogram (PPG) biomedical signal, and an electric-potential integrated circuit (EPIC) sensor was used to detect the ECG biomedical signal. Two diverse independent algorithms were used to measure the HR in two independent channels in a 2oo2 safety-related design architecture. Analytics were performed on the captured data to detect and identify the potential systemic faults during vital sign parameter measurement at each independent channel.

A set of measured parametric data was collected periodically, such as HR, from two independent channels and correlated to check for any computational faults. Karl Pearson's coefficient of correlation method was used (1) for safe voting logic and implemented to detect the computational faults of the two independent channels. A built-in test (BIT) fault-tolerant safe degradation function was implemented to identify and isolate systemic and computational faults. A set of sequence operations were performed for any fault detected, such as (1) the system operated in safe mode with safe degradation switching from 2oo2 to 1oo2 or 1oo1 (one-out-of-one) and vice versa toward the isolation of the fault and, thus, provided authentic data; (2) the system used negation error codes for each fault category, thus generating the related alarm for each significant detected fault, and subsequently logging the results. Similarly, the experiment was repeated on the proposed conceptualized five design configurations and evaluated for system resilience. The proposed design configurations are the following: (1) using diverse ECG and PPG sensors and algorithms at each independent channel to measure the HR, (2) using different ECG sensors at each independent channel with diverse algorithms to measure the HR, (3) using different PPG sensors at each independent channel with diverse algorithms to measure the HR, (4) using a single ECG sensor and measuring the HR with diverse algorithms at each independent channel, and (5) using a single PPG sensor and measuring the HR with diverse algorithms at each independent channel. A detailed analysis was carried out for each implemented conceptual architecture, and the rightness of the cardiac vital sign heartrate measurements and the coefficient of correlative results of vital sign measurements between two channels were analyzed and presented using Bland–Altman and correlative plots (18, 19). The recorded data, failure detected signal, and vital sign HR measurements at each channel output and the safe function output results were analyzed for the effective functioning of fault isolation and reducing of a SPOF.

The contributions of this research analysis to this configured medical system prototype, with the analytics on data collected using safety-related design approaches having interfaces with diverse PPG and ECG sensors signified:

1. Improvements in the elimination of PPG and ECG sensor-related problems in bio-signal detection and identification of the root causes for removing the deficiencies in signal processing techniques to extract the authentic vital sign signal information;
2. Efficient predictability in the estimation of illness using accurate data, which eventually improves proper nursing;
3. Improvement in system operability with significantly less insignificant alarms, thus improving the fault detection mechanism, fault identification mechanism, fault tolerability, and safety integrity level of the system for usage in safety-critical applications.

Section Theoretical Framework provides the details about the framework to address the specified challenges in the software faults. This framework includes detailed requirements about the conceptualized architectures and their approaches toward the analysis of the challenges. Section Methodology provides a detailed methodology about the system overview and its realization of two independent channels for experimental studies. Section Experimental Results and Discussion provides the experimental results and analysis about the measured HR-values evaluated along with the fault alarms between the diverse channels and calculated correlation coefficient values for the safety logic. This section tabulated a cause-and-effect system analysis, emphasizing authentic measures, and safety improvements in the medical system.

THEORETICAL FRAMEWORK

Work in safety-related electronic systems design and development is multi-dimensional, which means several safety aspects need to be considered in all phases of the product development life-cycle (PDLC) (15–17). These related systems should adhere to the standard development processes and guidelines and need compliance to respective standards like IEC, ANSI, AAMI, CENELEC, RTCA-DO (15, 17) to domains such as the automotive, medical, railways, and aerospace domains. The objective of any safety-related electronic system should efficiently detect faults. It shall drive the system into a fail-safe mode of system operation based on the severity level of the fault that occurs. The systems where faults and failures like random failures, systematic failures, hardware failures, software failures, or any unknown erroneous errors lead to hazardous situations like death, injury, or environmental damage. However, how effective a system is engineered for a particular application and a signal processing algorithm plays a critical role, as it extracts and provides vital information to make decisions. Generally, for this computation development process, defining the system requirements always challenges dependencies and limitations in sensors, hardware, and other operational environmental factors. In particular, electronic medical systems, sensors, and related computational systems provide important vital sign information regarding the nursing of a subject by monitoring and controlling treatment without any faulty or insignificant information. Thus, the system should not go into a non-operational mode during critical surgical procedures in the ICU.

Study of Potential Faults With PPG and ECG Sensor Systems

Recent studies have shown significant differences in the many brands of patient monitoring systems, particularly regarding their bias and precision, even though they use identical hardware. These variances are almost certainly due to the different algorithms used in processing the PPG signal (20, 21). Equally, different algorithms were used (21) in processing the EEG signal to mitigate differences in the sensing of materials used and the artifact of interest in the signal to be measured. In-depth studies show that potential faults can notice if anything is compromised or missed in defining the system requirements.

In defining the signal processing requirements, the consideration of a few challenging areas (20) in the PPG signal processing include (1) the selection of LEDs and detectors frequency ranges, (2) the placement of the sensor probes at the fingertips, ear, nose, or forehead, (3) non-invasive probing mechanism, either through light radiation transmission or reflection, (4) other considerable effects like changes in saturation, signal quality, effects of dyshemoglobins, dyes, other pigments, and extraneous factors, and (5) subject physical motion and environmental effects. Equally, in the processing of the ECG signal, the challenging areas (21–24) include: (1) the sensing materials, (2) noise removal and signal quality, (3) non-invasive probing mechanisms using or not using dielectric mediums between the probes and subjects, and (4) other considerable effects like subject physical motion and environmental effects. However, considering these defined requirements and the implementation of the signal processing algorithms software, there is any inconsistency between the algorithms to extract the same vital sign parameter due to faults or limitations at a certain level. In the recent past, few experimental studies with varied sample parameter voting and data fusion methods were used (9, 12–14, 25–29) to better the safety of these devices. Additionally, few studies (14, 21, 22) reported that the same vital parametric data, like HR, can be realized with the different mediums of the sensor.

The present research focused on the fault-tolerant design approach and a safety aspect in the effective detection of the software computing faults in an algorithmic function, which used a defined framework and provided accurate measured parameter data with reduced alarms. The presented framework provided the fault detection and analysis approach and the implementation results using the configurable safety-related 2oo2 design architecture.

Proposed Conceptual Safety Architectures

The proposed 2oo2 evaluated safety-related design approach (1) was used for further experimental investigations and performed functional safety assessments to validate the implemented bio-signal processing functions in safety-critical medical applications of the biomedical systems. The proposed concept evaluated a set of configurable diverse medical sensors and different signal processing algorithms to measure selected vital parameters. This approach of processing the same parameter in a diverse method, with correlative analytics, offered a scope to improve

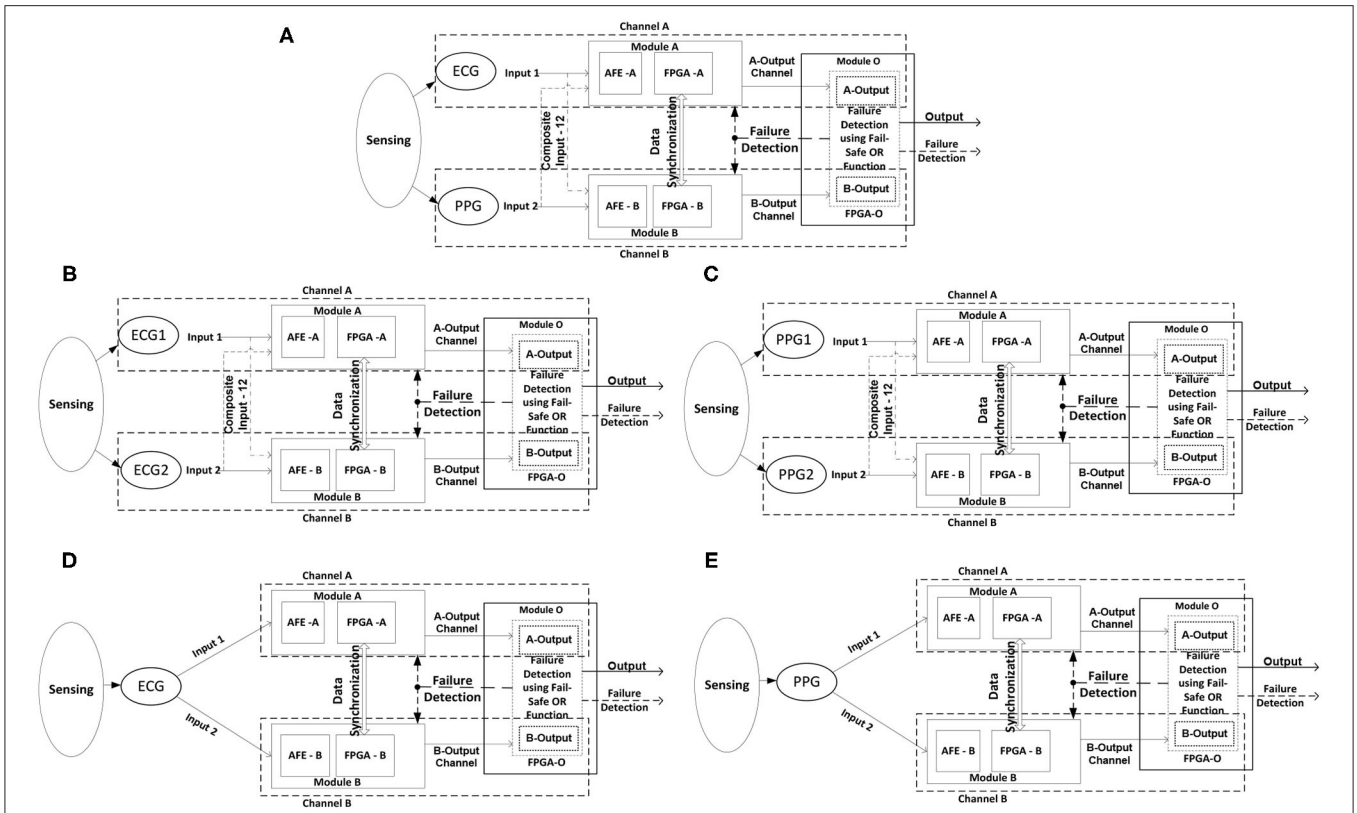


FIGURE 1 | Configurable 2oo2 system. **(A)** ECG and PPG sensor and diverse algorithm. **(B)** ECG1 and ECG2 sensor and the same algorithm. **(C)** PPG1 and PPG2 sensor and the same algorithm. **(D)** Single ECG sensor and the same algorithm. **(E)** Single PPG sensor and the same algorithm.

a useful technique in detecting faults. It also provided options to implement fail-safe degradation mechanisms, and the related usage of safety-related design architecture provided redundancy and availability.

This study considered five different concepts for analysis by configuring the system with diverse ECG and PPG sensors and diverse algorithmic computing software. A correlative analysis was performed and logged the results for each configuration. It tabulated all analysis inferences for each detected abnormality in the functional software and the results with analysis of cause and effects toward improving functional safety. The functional description of the 2oo2 architecture including hardware, software, sensors, and its algorithms used is detailed (1). The set of five selected configurations for framework analysis is described below.

Multimodal Sensor Configurable 2oo2 System

The 2oo2 System in **Figure 1A** is configured in 2oo2 with diverse sensor inputs, i.e., the ECG sensor interfaced to the analog front-end (AFE) device in channel-A and the PPG sensor interfaced to the AFE integrated chip (IC) in channel-B. Both channels are configured with selected diverse algorithms (1) to process the HR parameter. The outputs of modules A and B from each channel correlated with a "safe correlative-bounded configurable limits" function are implemented in Module-O to detect faults and generate the failure detection signal. An output signal consisted of accurate parametric data along with processed alarm signal data generated by the fault-tolerant safe degradation function. The safe function implementation included Karl Pearson's coefficient of correlation method, which used (1) a time series sliding window technique between the data of both channels and a fault-tolerant safe degradation voting logic function for reliable switching between the computed output of the channels.

As part of the framework analysis, the recorded output data were analyzed toward the reliability and efficiency of the implemented algorithmic software function and provided the inferences on improvements in the system level requirements. Thus, a practical HR computing function was realized in both channels with authentic HR output with reduced alarms. Similarly, this analysis was further carried out on selected concepts to improve the safety check, like:

a) System configured with diverse ECG or PPG sensors with the same algorithms in both channels, as shown in **Figures 1B,C**, providing the opportunity to investigate the sensor sensitivity and related deviances in the functional requirement for measuring the same desired parameter.
b) A system configured with one sensor interfacing with both channels, with the same algorithmic function for computation, as shown in **Figures 1D,E**. This provided the

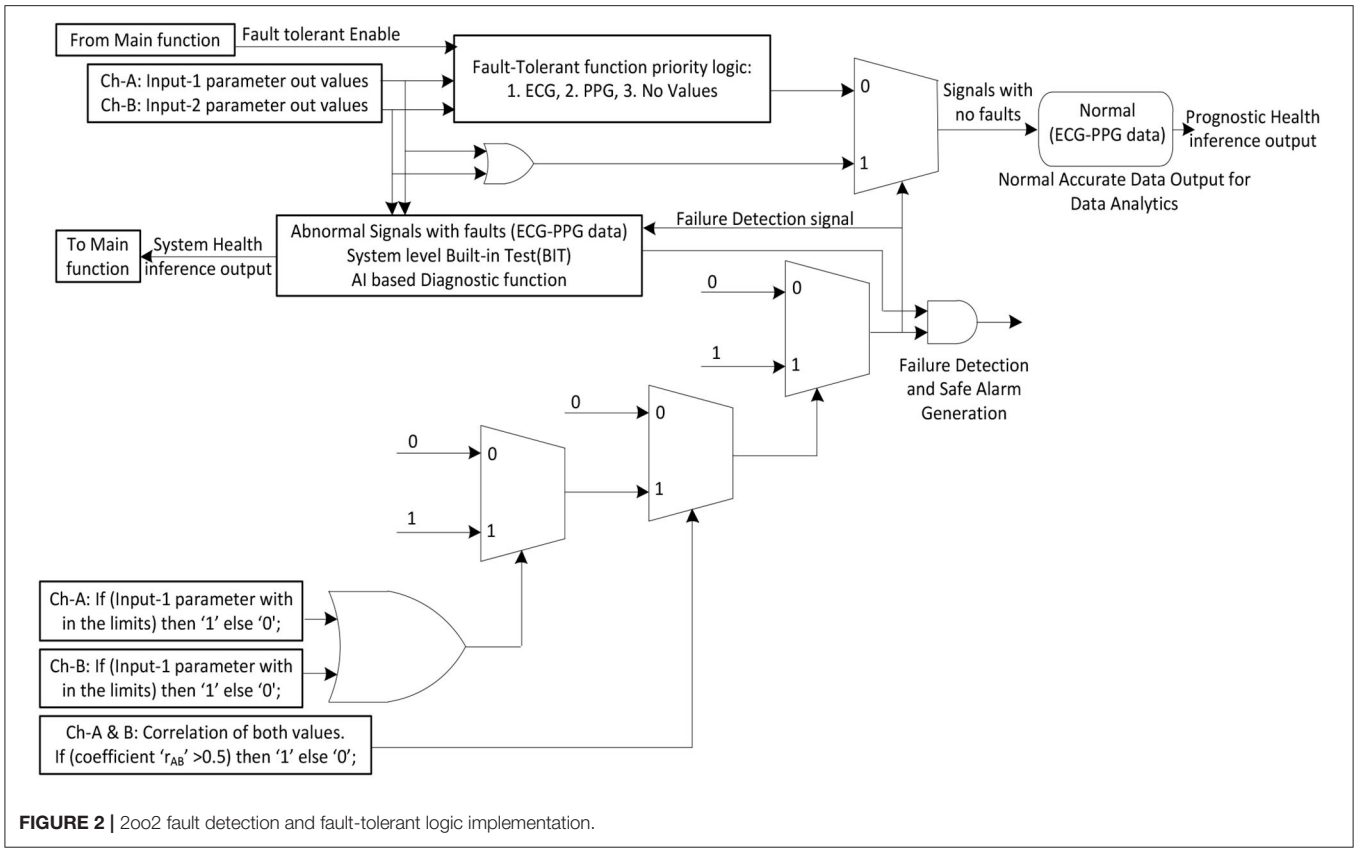


FIGURE 2 | 2oo2 fault detection and fault-tolerant logic implementation.

opportunity to investigate the response of the safety functions to implemented improvements in defining the requirements. Furthermore, it helps in the evaluation of system-level hardware failure analysis.

Evaluation Framework

Safety compliance is a systematic process followed in every phase of the product development life cycle. Every safety-related electronic system aims to detect faults when the system is in an active state and shall drive the system into a fail-safe state. The safe state shall be defined based on the mode of system operation in the active state. During the system operation at a specific mode, the cause of faults and failures is categorized as negation codes. It needs to be defined by labeling and the severity level of the fault that occurs.

In this study, we focused on the functional safety validity of a specific signal processing algorithmic software function and explained the implemented fault detection logic in the 2oo2 approach, its theoretical fault identification mechanism, and its further evaluation framework.

2oo2-Fault Detection Logic and Analysis

The function 2oo2 fault detection logic received the parameter input data from channel-A or channel-B and conditionally checked between the threshold limits of that particular parameter. The generated outputs were logically ORed as a parallel circuit, and the hardcoded safe-output was selected using the correlative coefficient condition (>0.5), which has a moderate- to a very high-state relationship with parameter "*r*" (1), as shown in **Figure 2**, to generate the fault detection signal. Furthermore, this fault detection signal was driven back as a feedback input to the primary function, which triggered the diagnostic BIT functions to generate the output enable signal. As shown, the primary function triggered a fault-tolerant enable signal for every authentic fault detected. Thus, a predefined priority-based fault-tolerant safe degradation sequence was initiated to select the apt output to pass onto the display. However, if there is no fault detected, both channel output parametric values were in a good positive relationship. Thus, a selected ORed output was released to the display and AI-based computed prognostic health inference outputs.

In the process of the fault identification mechanism, a defined (1) positive correlation coefficient constant value was compared and continuously monitored to a correlation coefficient value measured between two independently received output-parameter values. Thus, a signal was generated when there was deviance in the relationship between them. In theory, as per the 2oo2 safety-related voting approach, both channels needed to fail to initiate the system to fail-safe mode. Thus, a failure detection signal would then be generated when both the relationship signal and ORed output signal failed. All identified cases were analyzed as detailed in **Figure 3** from one to eight cases to determine the actual fault. As part of the evaluation framework, all these cases are tabulated in **Tables 1**, **3** to

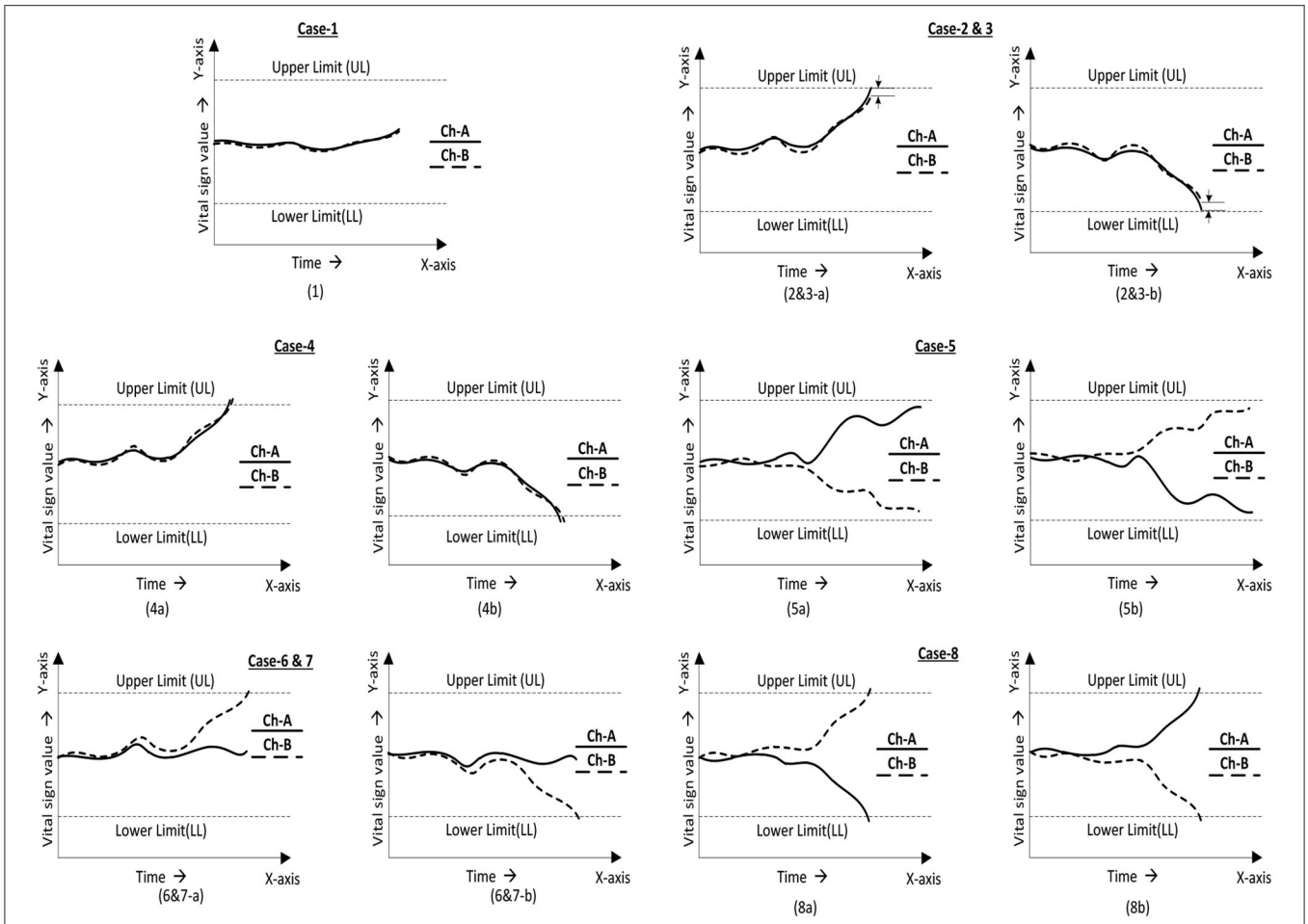


FIGURE 3 | 2oo2 voting logic cases for assessment. (1) Values are in a relationship and are in the normal range. (2 and 3) Values in relationship with one signal drift toward the limit. (4) Values in relationship with both signals drift toward the limit. (5) Values out of a relationship and in the normal range. (6 and 7) Values out of a relationship with one signal drift toward the limit. (8) Values out of a relationship with both signals drift toward limits.

TABLE 1 | 2oo2-based safe-voting logical truth table for fault detection using a correlative technique.

Evaluation cases	Ch_A	Ch_B	r_{AB} (>0.5) (Correlation coefficient)	Output signal (2oo2 fault-tolerant) (O_{AB})	Fault detection (Diagnostics) Yes/No (Fd)
Case-1	True	True	True	No alarm	No
Case-2	False	True	True	No alarm	Yes
Case-3	True	False	True	No alarm	Yes
Case-4	False	False	True	Alarm	No
Case-5	True	True	False	Alarm	Yes
Case-6	False	True	False	Alarm	Yes
Case-7	True	False	False	Alarm	Yes
Case-8	False	False	False	Alarm	Yes

analyze a selected vital parameter and were evaluated by checking the validity of the related implemented function for safety.

Following standard compliance, best development practices, and analysis framework approaches in implementing safety-critical medical systems significantly addressed the identified challenges and improved the safety features of the medical system. As part of the evaluation framework, a detailed analysis was carried out for a specific vital parameter. Preliminary analytical research results showed improvements in functional safety in the reduction of spurious alarms, effective detection of functional faults, and improvement of the uninterruptable function of the system by providing authentic measurements.

METHODOLOGY

Experimental research activities of this part of the project were divided into three main phases. The first phase was dedicated to the study and evaluation of the aptness for use of the safety-related architectures in the targeted non-invasive diagnostic medical monitoring and control systems to measure basic vital parameters and identify suitable sensors for sensing biomedical signals like sensing through electric potential, sound, and light.

The second phase was dedicated to the design and realization of the three-independent modular designed channels. Each channel was interfaced with a suitable sensor and each channel module based on a field-programmable gate array (FPGA) design with a modular integrated system interface research prototype was built for experimental research studies. The final phase was the validation of the experimental research platform. The system was evaluated with varied configurations with sensors, configurable safety-related architectures, and FPGA circuits. Lab and field trials were performed to assess vital parameters like HR, address the mentioned challenges, reduce fault alarms, identify algorithm limitations, and improve uninterruptable functionality with safe degradation mechanisms.

The present study focused on the final phase of the research activities to improve fault detection mechanisms and, subsequently, address the mentioned challenges. Using configurable safety-related architectures with a combination of ECG and PPG sensor interfaces, the details of the first and second

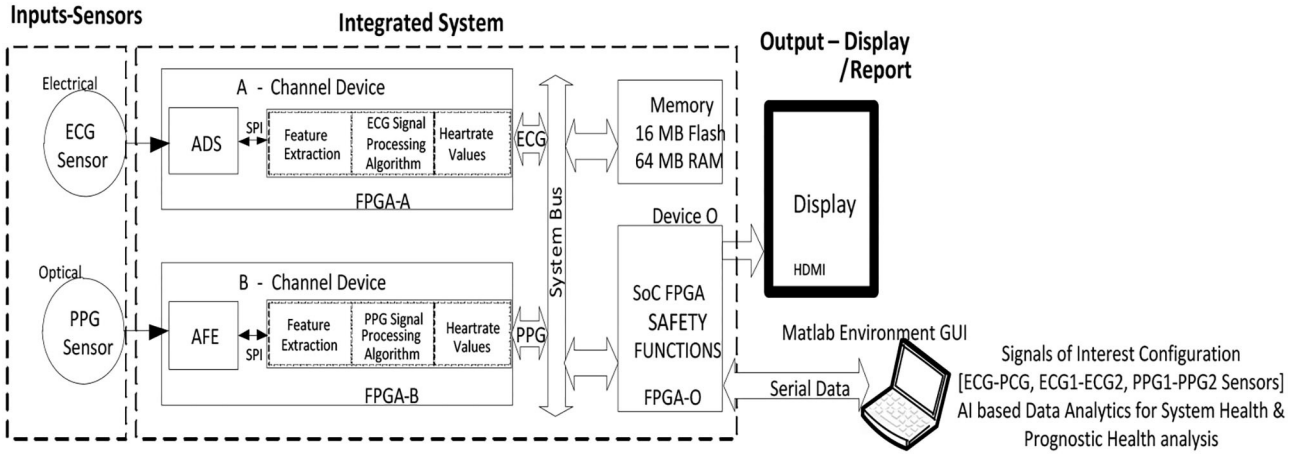
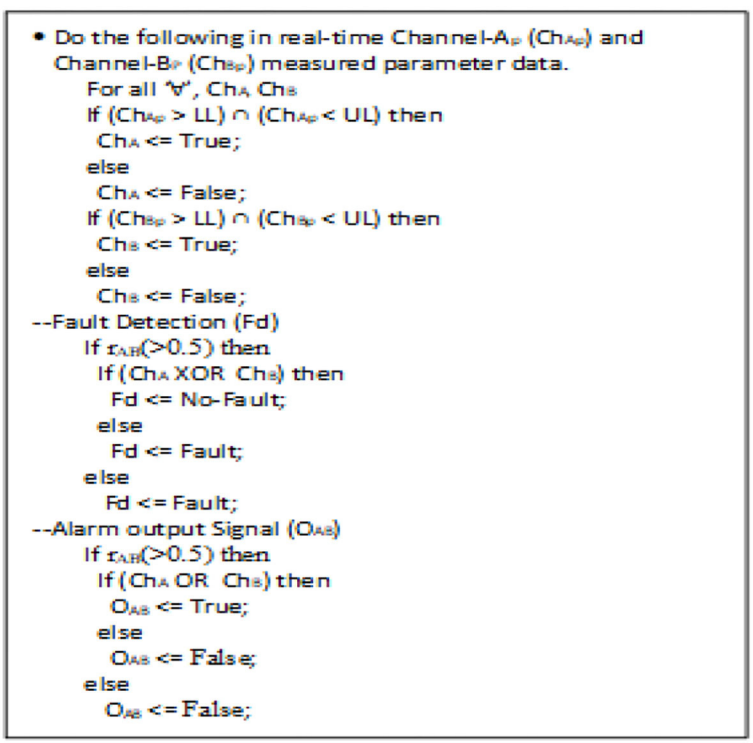


FIGURE 4 | 2oo2 fault-tolerant cardiac health monitoring system block diagram.



```
• Do the following in real-time Channel-AP (ChAP) and
  Channel-BP (ChBP) measured parameter data.
    For all 'P', ChA ChB
    If (ChAP > LL) ∩ (ChAP < UL) then
     ChA <= True;
    else
     ChA <= False;
    If (ChBP > LL) ∩ (ChBP < UL) then
     ChB <= True;
    else
     ChB <= False;
--Fault Detection (Fd)
    If rAB(>0.5) then
     If (ChA XOR ChB) then
      Fd <= No-Fault;
     else
      Fd <= Fault;
    else
      Fd <= Fault;
--Alarm output Signal (OAB)
    If rAB(>0.5) then
     If (ChA OR ChB) then
      OAB <= True;
     else
      OAB <= False;
    else
      OAB <= False;
```

FIGURE 5 | Pseudocode for fault detection analytics and alarm output signal.

part of the research prototype were used (1) to build activities, design evaluation procedures, and application protocols.

System Overview

The cardiac health monitoring system (CHMS) is modular and configured to a 2oo2 safety-related computing platform. The system consists of two independent operating channels, as shown in **Figure 4** and explained in (1) detail with CHMS Graphical User Interface (GUI) interfaces for test evaluations.

Sensors—PPG, ECG, and Its System Configurations

The set of PPG sensors (LEDs and a detector), ECG sensors (EPIC and Ag-AgCl electrodes), and its interfaces used in this system configuration are detailed (1). These combinations of sensors bonded into a single module for convenience, and each module was placed on a subject at the prescribed location for better measurements. These sensor modules were packaged in modular combinations like ECG–PPG, ECG–ECG, and PPG–PPG, and each module was interfaced with the integrated computing system, as shown in **Figure 4**, configured, and evaluated for each conceptual design.

2oo2 Based Safe-Voting Method Using Correlation Analytics

The prototype system needed to be configured with 2oo2 with a safe-voting mechanism for measuring a particular identified

TABLE 2 | 2oo2 configured-system safe degradation scheme.

Vital sign	Fault-tolerant system	Degradation level			System configuration selection scheme for high availability or safety mode	
		1	2	3	High availability and fault-tolerant	High safety and fault-tolerant
Heart rate or pulse	2oo2	1oo2	1oo1	Shut down/Safe mode	2oo2 → 1oo1 → Safe mode	1oo2 → 1oo1 → Safe mode
Respiratory rate	2oo2	1oo2	1oo1	Shut down/Safe mode		
Blood pressure	2oo2	1oo2	1oo1	Shut down/Safe mode		
Body temperature	2oo2	1oo2	1oo1	Shut down/Safe mode		
Pulse oximetry	2oo2	1oo2	1oo1	Shut down/Safe mode		

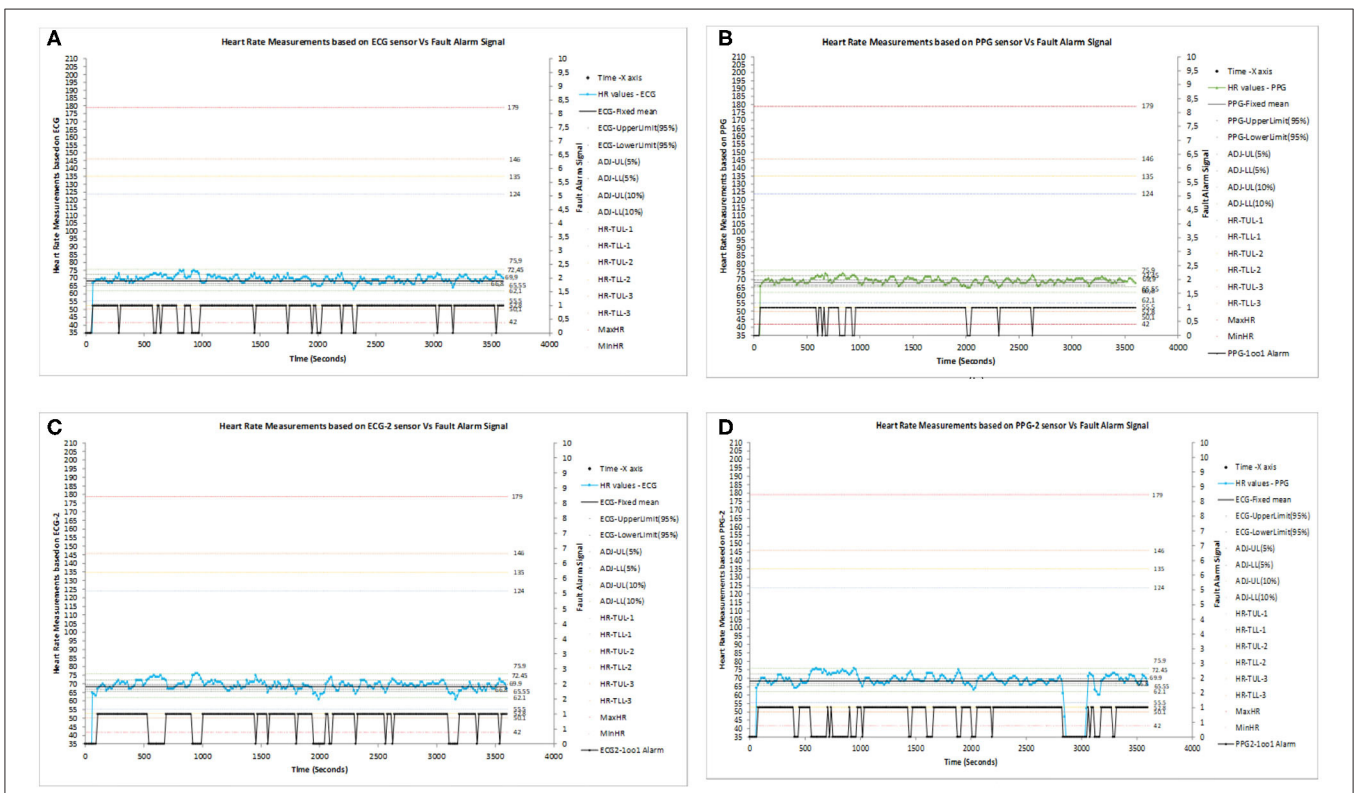


FIGURE 6 | (A) Heart rate (HR) measurement based on the ECG sensor vs. the fault alarm signal in 1oo1 configuration. **(B)** HR measurement based on the PPG sensor vs. the fault alarm signal in 1oo1 configuration. **(C)** HR Measurement based on the ECG-2 sensor vs. the fault alarm signal in 1oo1 configuration. **(D)** HR Measurement based on the PPG-2 sensor vs. the fault alarm signal in 1oo1 configuration.

vital parametric function. In this system operation approach, both channels A and B received the sample data and computed the vital parameter independently. These measured values were inputs to the safety function, where it correlated and voted to generate the alarm output signal and fault-detection signal, which is as per **Table 1** and **Figure 5**. Furthermore, the generated signals triggered the related diagnostic functions to process the data to select the authentic vital output for display and the fault alarm signal. Thus, this mechanism of configuring and configured safe-voting computation was performed for each essential parameter measured by the medical system.

2oo2 Safety-Related Degradation Mechanism

Based on the selected configured parameters in an active system, the system went into a predefined sequence of degradation for any fault detected. These sequences of degradation depended on the configuration of the system and its limitations, such as sensors and availability of the independent processing channels for the same selected parameter. A few identified vital parameters and their feasible degradation scheme are shown in **Table 2**. Since the system is a modular interface and a provision for high reconfigurability was provided during its initial start phase, it must either have a high-availability system or high safety. Thus, the scheme from 2oo2 to 1oo1 can be altered for high availability with fault-tolerability (or) 2oo2 to 1oo2 to 1oo1 for high-safety with fault-tolerability.

$$Availability = \frac{Uptime}{(Uptime + Downtime)} \quad (1)$$

Experimental Setup and System Evaluation

The experimental setup and its application protocol detailed in (1) were effectively reused in assessing the proposed concepts. The MATLAB-based CHMS GUI tool was used to configure and capture the resultant data to Plot.

EXPERIMENTAL RESULTS AND DISCUSSION

The system configured with 1oo1 could capture HR data values vs. fault alarm signals individually as ECG−1oo1, PPG−1oo1, ECG2–1oo1, and PPG2–1oo1, which are presented in **Figures 6A–D** from a single subject. The captured alarm data were analyzed to set configurable adjustable upper and lower limits (ADJ-UL and LL) (1).

The configured 2oo2 outputs of ECG–PPG, ECG1–ECG2, and PPG1–PPG2 correlation signal vs. the fault alarm signal are shown in **Figures 7A–C**. The analysis showed that the reduction of alarm readings from all 50 subjects was similar in meeting the objectives. **Figures 8A–C** present the configured 2oo2 results

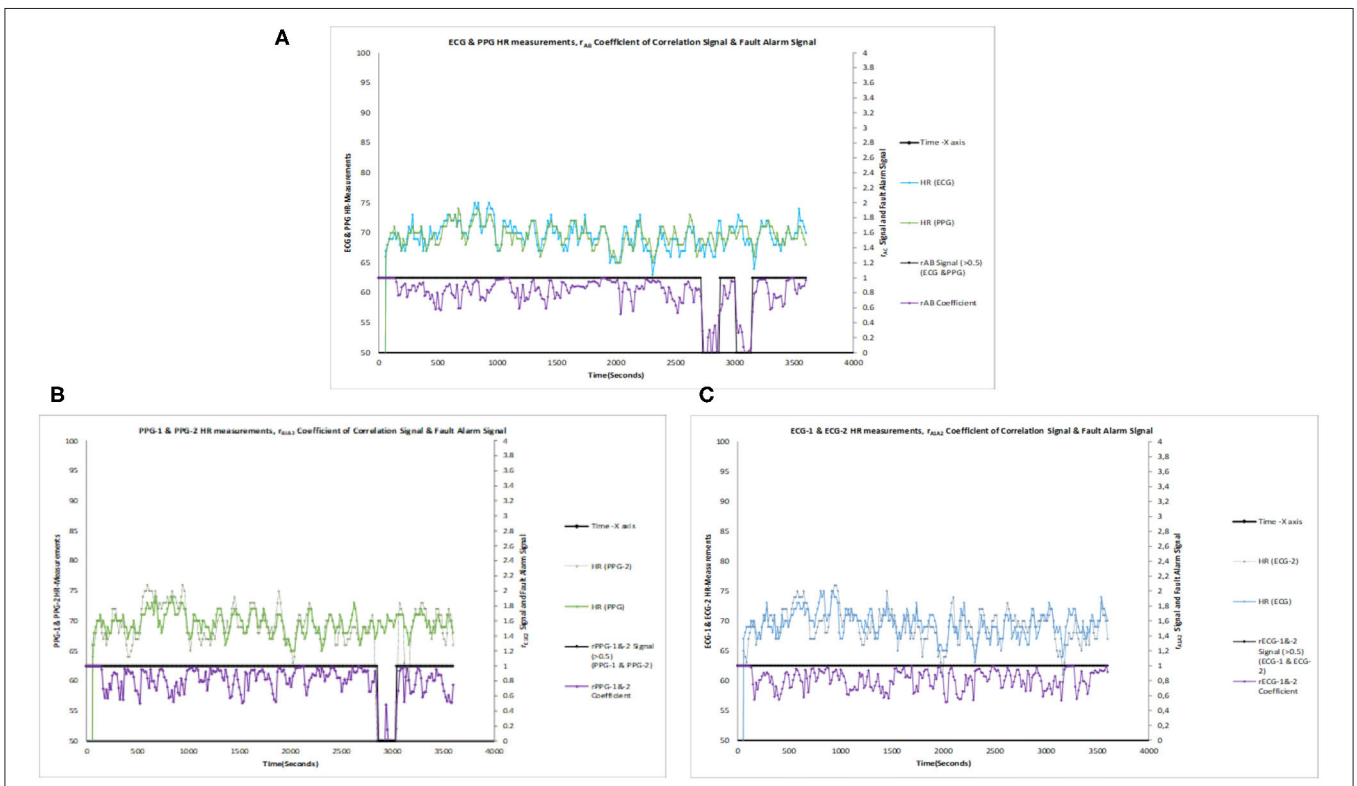


FIGURE 7 | (A) ECG and PPG HR measurements, r_{AB} coefficient of correlation signal vs. fault alarm signal. **(B)** ECG-1 and ECG-2 HR measurements, r_{A1A2} coefficient of correlation signal vs. fault alarm signal. **(C)** PPG-1 and PPG-2 HR Measurements, r_{B1B2} coefficient of correlation signal vs. fault alarm signal.

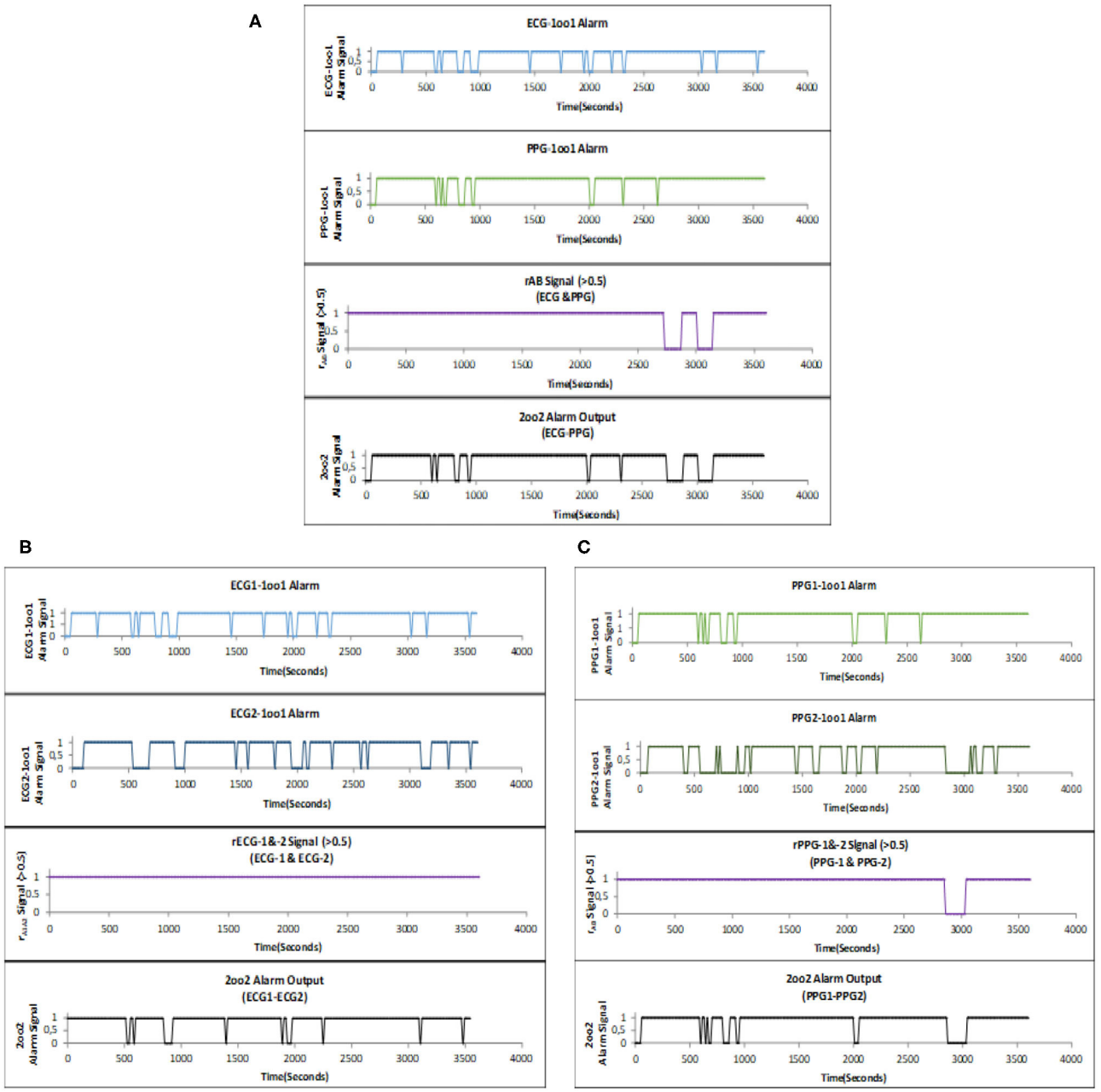


FIGURE 8 | (A) ECG–PPG recorded 2oo2 fault alarm output during HR monitoring within a 1-h period. **(B)** ECG1–ECG2 recorded 2oo2 fault alarm output during HR monitoring within a 1-h period. **(C)** PPG1–PPG2 recorded 2oo2 fault alarm output during HR monitoring within a 1-h period.

of a single subject, which show the reduction in alarms and its related cause-and-effect evaluated analysis presented in **Table 3** with inferences.

The sensors and ECG and PPG processed signal data were captured and used to performed analytics with the MATLAB tool by configuring the system in 2oo2 configuration. The tool computed system uptime and downtime by separating the normal and abnormal signal data as shown in **Figure 9**. Similarly, the data captured for ECG1–ECG2 and PPG1–PPG2 configurations and results were recorded in **Table 4** for a single subject. Following Helsinki's declaration and consent, the monitoring system evaluated 50 subjects of various age groups and recorded the uptime and downtime of the system during evaluation, with an average of 50 h for the total operating time of the system and calculated system availability in percentile as per Equation (1). **Table 4** provides these results along with system health inferences computed per negation codes specified in **Table 3** and assesses in the three configurations that the

TABLE 3 | 2oo2-based cause and effect—fault detection and evaluation analysis and feasible mitigation solution for the identified cause.

Evaluation cases	Channel-A (ECG) (parameter within set limits is true else false)	Channel-B (PPG) (parameter within set limits is true else false)	ECG–PPG *r* (>0.5) (correlation coefficient)	Output signal (2oo2 fault-tolerant) (true—no alarm; false—alarm)	Fault diagnostics required for detected failure yes/no	Software functional fault/ negation codes [Sw, software; Ag, algorithm; F, function; (four-digit)—Code]	Fault severity analysis (no-fault/minor/major/ critical) (algorithms function analysis)	Fault description and probable cause	Mitigation solution for the identified cause
Case-1	True	True	True	True	No	SwAgF4001	No-Fault	Data authentic	No action
Case-2	False	True	True	True	Yes	SwAgF4002	Minor	• Data authentic • Software sync/delay issue between channels (or) • Negligible higher pulse count detected in ECG	Perform analytics for consistency (or) repetitive higher pulse count on ecg signal and rectify the issue
Case-3	True	False	True	True	Yes	SwAgF4003	Minor	Similarly, as above for PPG	Similarly, as above for PPG
Case-4	False	False	True	False	No	SwAgF4004	No-Fault	Data authentic	No action
Case-5	True	True	False	False	Yes	SwAgF4005	Major	Possible software sync/Delay issue between channels	Perform analytics to check for the consistent difference between channels to rectify the issue
Case-6	False	True	False	False	Yes	SwAgF4006	Major	Possible cause in software or hardware issue	Perform built-in-test (BIT) and analytics between channels to rectify the issue
Case-7	True	False	False	False	Yes	SwAgF4007	Major	Refer above comment	Refer above comment
Case-8	False	False	False	False	Yes	SwAgF4008	Critical	Possible cause in software or hardware issue	Perform BIT and analytics between channels to rectify the issue

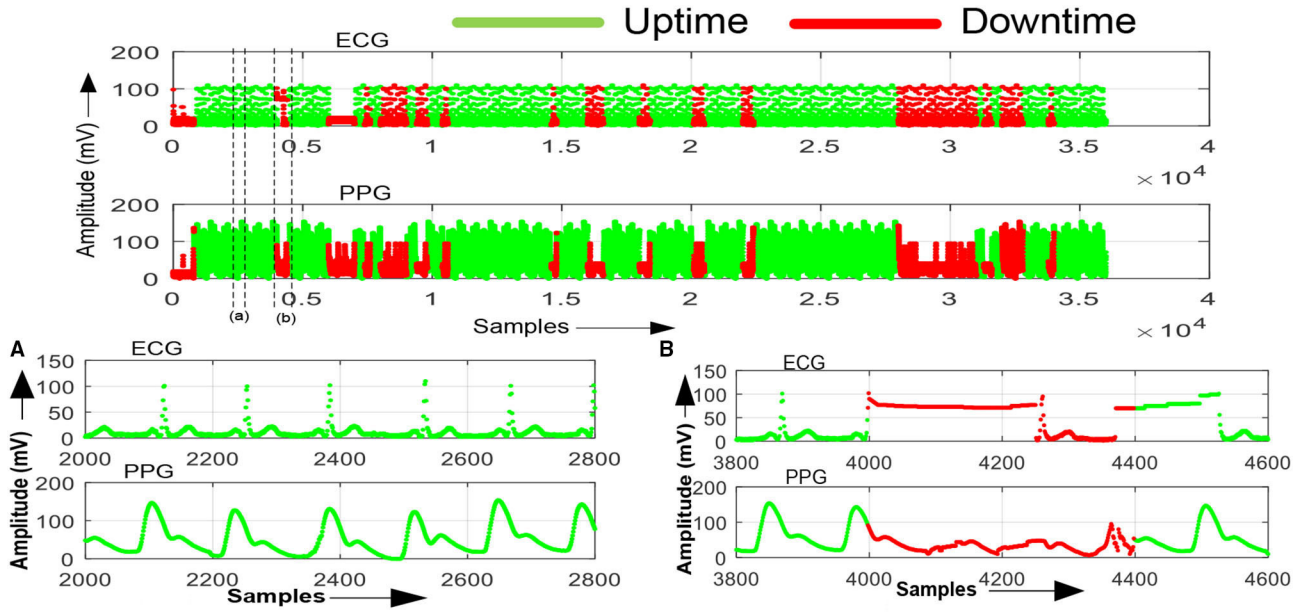


FIGURE 9 | ECG and PPG processed signals data captured using MATLAB tool w.r.t 2oo2 configuration. **(A)** System uptime—normal signal data with no-fault (green). **(B)** System downtime—abnormal signal data with a fault (red).

TABLE 4 | System availability results by configuring the fault-tolerant multimodal sensor system in 2oo2 system configuration mode.

Fault-tolerant multimodal sensor system configuration	Signals	Up-time (normal Signal) (Avg. min) (~"t"–60 min)	Down-timev (abnormal Signal) (Avg. min) (~"t"–60 min)	System availability configuration (%)	System availability average of ~50 h (%)	Improvement in system availability using 2oo2 design approach % increase (%)	Data analytics inferences on system health—common causes
Concept-1 (ECG–PPG) Evaluation on 50 subjects (1-h/person)	Channel-A (ECG)	53.4	6.6	89	90	57.8	Main hardware causes: • ECG/PCG Sensor probe contact fault. • Power-related faults
	Channel-B (PPG)	55.2	4.8	92%	89		
	Fault-tolerant-output (ECG–PPG)	58.2	1.8	97	95		
Concept-2 (ECG1–ECG2) Evaluation on 50 subjects (1-h/person)	Channel-A (ECG1)	54.6	5.4	91	89	62.5	Main software causes: • Data Sync delay Issues between channels due to algorithm computation times • Inaccuracy in detecting the pulses or due to Noise causes
	Channel-B (ECG2)	56.4	3.6	94	91		
	Fault-tolerant-output (ECG1–ECG2)	58.8	1.2	98	96		
Concept-3 (PPG1–PPG2) Evaluation on 50 subjects (1-h/person)	Channel-A (PPG1)	57.6	2.4	96	91	46.4	
	Channel-B (PPG2)	57.0	3.0	95	94		
	Fault-tolerant-output (PPG1–PPG2)	58.8	1.2	98	97		

improvement in system availability significantly improved from 45 to 55%.

However, to validate the improvement function at the system level resilience, some more tests needed to be performed with various sensors to evaluate function performance better. It was observed that, during experimentation, there were limitations such as a signal drift issue during capture and appropriate sync mechanisms needed to improve the signals and channels. The drift of ±2 pulses was noticed as a limitation of the system due to the mediums of the different sensors and the capturing

and measurement periods. Furthermore, the system response time of <5 s was noticed as a limitation. Detailed analysis of the uptime period of signals confirmed that the count of truthful pulses in both channels was almost the same during long-term signal capture with negligible or no incorrect pulses and undetected pulses recorded. In contrast, it can mitigate drift issues during short-term signal capture in the design by improving the synchronization mechanism for capturing the signals between the channels.

Further investigation on the uptime and down signals could help the understanding of the causes of various systemic faults within the sensor system. The related fault data captured were analyzed and used to provide inferences in **Table 4**. Additional analysis of this corresponding prognostic health data was out of the scope of this study as it required defining normal or abnormal vital parameter signal classifications and identifying support mechanisms within the system. In this study, as the focus was on system availability with reduced alarms by processing the signal data through the evaluation of the safety function, only minimal conditions to infer the health of the system were used.

The presented experimental data were captured from each channel with HR measured data. The inverted logic level of the alarm signal was logged for the 1-h duration per subject, with consideration for evaluating the system. The analysis results showed significant improvements in meeting the objectives and a similar systematic approach to further apply this method to other parameter evaluations for safety improvements.

CONCLUSION

In this paper, the concept of a fault-tolerant safety-related 2oo2 design approach implemented and evaluated in the configurable medical CHMS, which is a research platform, addressed the effective detection of functional faults, thus improving the uninterruptable function of the targeted medical system by reducing the false or spurious alarms. This framework found a significant reduction in the generation of insignificant alarms and increased uninterruptable system availability by 45–55%. These findings and the design approach were important contributions to issues related to present medical patient monitoring systems without significant impact on cost since it uses the existing system configuration of PPG and ECG signals, along with FPGA technology devices. While eliminating identified issues was specifically focused on, the conceptual design approach may suit medical monitoring systems, implying that the findings are likely to be of importance to the design of medical monitoring and control systems. In terms of future research, it is particularly suggested to use diverse algorithms and sensors or evaluation with a combination of these with effective predictive system maintenance, which could help eliminate spurious alarms with a reduced downtime of the system and produce more accurate data vital parameters.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because Non Disclosure Agreement with the patients. Requests to access the datasets should be directed to Prasadraju Lakkamraju, prasadraju.lvr@gmail.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Citizens Hospitals Hyderabad. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

The author would like to thank the Faculty of Center for VLSI and Embedded Systems Technologies (CVEST), the International Institute of Information Technology (IIIT)-Hyderabad, the Ministry of Human Resource Development (MHRD), the Government of India, and the School of Computing and Electrical Engineering (SCEE), Indian Institute of Technology (IIT)-Mandi for their valuable support during this research.

REFERENCES

1. Prasada Raju L, Anumukonda M, Roy Chowdhury S. Safety-related studies on non-invasive biomedical signals and its aptness usage in design of fault tolerant multimodal human health monitoring system. In: *Doctoral Consortium* (Prague). (2019) p. 3–14.
2. Alemzadeh H, Iyer RK, Kalbarczyk Z, Raman J. Analysis of safety-critical computer failures in medical devices. *IEEE Secur Privacy.* (2013) 11:14–26. doi: 10.1109/MSP.2013.49
3. Laible U, Bürger T, Pritschow G. A fail-safe dual channel robot control for surgery applications. *Safety Sci.* (2004) 42:423–36. doi: 10.1016/j.ssci.2003.09.009
4. Li Z, Yang C, Burdet E. Guest editorial an overview of biomedical robotics and bio-mechatronics systems and applications. *IEEE Trans Syst Man Cybern Syst.* (2016) 46:869–74. doi: 10.1109/TSMC.2016.2571786
5. Hsiao, C-C, Horng J, Lee R-G, Lin R. Design and implementation of auscultation blood pressure measurement using vascular transit time and physiological parameters. *IEEE Int Conf Syst Man Cybern.* (2017) 2017:2996–3001. doi: 10.1109/SMC.2017.8123084
6. US Food and Drug Administration. *Medical Device Databases*. Available online at: www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Databases/default.htm (accessed March 28, 2019).
7. Simone LK. Software-related recalls: an analysis of records. *Biomed Instrum Technol.* (2013) 47:514–22. doi: 10.2345/0899-8205-47.6.514
8. Hravnak M, Pellathy T, Chen L, Dubrawski A, Wertz A, Clermont G, et al. A call to alarms: current state and future directions in

the battle against alarm fatigue. *J Electrocardiol.* (2018) 51:S44–8. doi: 10.1016/j.jelectrocard.2018.07.024
9. Hegde V, Raheja D. Design for reliability in medical devices. In: *2010 Proceedings-Annual Reliability and Maintainability Symposium (RAMS)*. San Jose, CA: IEEE (2010). p. 1–6. doi: 10.1109/RAMS.2010.5448077
10. Kaushik R, Desreumaux P, Mouret JB. Adaptive prior selection for repertoire-based online adaptation in robotics. *Front Robot AI.* (2020) 6:151. doi: 10.3389/frobt.2019.00151
11. Jin L, Liao B, Liu M, Xiao L, Guo D, Yan X. Different-level simultaneous minimization scheme for fault tolerance of redundant manipulator aided with discrete-time recurrent neural network. *Front Neurorobot.* (2017) 11:50. doi: 10.3389/fnbot.2017.00050
12. Alemzadeh H, Di Martino C, Jin Z, Kalbarczyk ZT, Iyer, RK. Towards resiliency in embedded medical monitoring devices. *Presented at the IEEE/IFIP 42nd International Conference on Dependable Systems and Networks Workshops (DSN-W)*. Boston, MA (2012). doi: 10.1109/DSNW.2012.6264662
13. Saxena D, Raychoudhury V. Design and verification of an NDN-based safety-critical application: a case study with smart healthcare. *IEEE Trans Syst Man Cybern Syst.* (2019) 49:991–1005. doi: 10.1109/TSMC.2017.2723843
14. Sheng H, Schwarz M, Borcsok J. Modeling a safety-related system for continuous non-invasive blood pressure monitoring. *Biomedical Circuits and Systems Conference (BioCAS)*. Rotterdam: IEEE (2013). doi: 10.1109/BioCAS.2013.6679687
15. International Electrotechnical Commission. *IEC 61508 Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems*. Geneva: International Electrotechnical Commission (1998).
16. Börcsök J. *Electronic Safety Systems: Hardware Concepts, Models, and Calculations*. Heidelberg: Hüthig (2004).
17. EN CS. *50129-2003 Railway Applications: Communications, Signaling and Processing Systems-Safety Related Electronic Systems for Signaling*. London: British Standards Institution (BSI) (2003).
18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Int J Nurs Studies*. (2010) 47:931–6. doi: 10.1016/j.ijnurstu.2009.10.001
19. Myles PS, Cui J. I. Using the Bland-Altman method to measure agreement with repeated measures. Brit J Anaesth. (2007) 99:309–11. doi: 10.1093/bja/aem214
20. Sinex JE. Pulse oximetry: principles and limitations. *Amer J Emerg Med.* (1999) 17:59–66. doi: 10.1016/S0735-6757(99)90019-0
21. Selvaraj N, Jaryal A, Santhosh J, Deepak KK, Anand S. Assessment of heart rate variability derived from finger-tip photo-plethysmography as compared to electrocardiography. *J Med Eng Technol.* (2008) 32:479–84. doi: 10.1080/03091900701781317
22. Vescio B, Salsone M, Gambardella A, Quattrone A. Comparison between electrocardiographic and earlobe pulse photoplethysmographic detection for evaluating heart rate variability in healthy subjects in short-and long-term recordings. *Sensors.* (2018) 18:844. doi: 10.3390/s18030844
23. Pang Z, Yang G, Khedri R, Zhang YT. Introduction to the special section: convergence of automation technology, biomedical engineering, and health informatics toward the healthcare 4.0. *IEEE Rev Biomed Eng.* (2018) 11:249–59. doi: 10.1109/RBME.2018.2848518
24. Sidek KA, Khalil I, Jelinek HF. ECG biometric with abnormal cardiac conditions in remote monitoring system. *IEEE Trans Syst Man Cybern Syst.* (2014) 44:1498–509. doi: 10.1109/TSMC.2014.2336842
25. Miao F, Fu N, Zhang YT, Ding XR, Hong X, He Q, et al. A novel continuous blood pressure estimation approach based on data mining techniques. *IEEE J Biomed Health Inform.* (2017) 21:1730–40. doi: 10.1109/JBHI.2017.2691715
26. Cobus V, Heuten W. To beep or not to beep? Evaluating modalities for multimodal ICU alarms. *Multimod Technol Interact.* (2019) 3:15. doi: 10.3390/mti3010015
27. Gabriel A. Design and evaluation of safety instrumented systems: a simplified and enhanced approach. *IEEE Access.* (2017) 5:3813–23. doi: 10.1109/ACCESS.2017.2679023
28. Song X, Zhai Z, Lv R, Guo Y, Zhu P. Optimal design of the *k*-out-of-*n*: G (F) majority voter. *IEEE Access.* (2017) 5:22647–56. doi: 10.1109/ACCESS.2017.2761776
29. Lakkamraju P, Anumukonda M, Chowdhury SR. Improvements in accurate detection of cardiac abnormalities and prognostic health diagnosis using artificial intelligence in medical systems. *IEEE Access.* 8:32776–82 (2020). doi: 10.1109/ACCESS.2020.2965396

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.

frontiers in Public Health

ORIGINAL RESEARCH
published: 16 November 2021
doi: 10.3389/fpubh.2021.745524

OSA Patient Monitoring Based on the Beidou System

Cai Liangming[1†], Cai Xiaoqiong[2†], Du Min[3], Miao Binxin[4], Lin Minfen[4], Zeng Zhicheng[4], Li Shumin[4], Ruan Yuxin[4], Hu Qiaolin[4] and Yang Shuqin[4]*

[1] School of Physics and Information Engineering, Zhicheng College, Fuzhou University, Fuzhou, China, [2] Third Institute of Oceanography, State Oceanic Administration, Xiamen, China, [3] Laboratory of Eco-Industrial Green Technology Wuyi College, Wuyishan, China, [4] College of Zhicheng, Fuzhou University, Fuzhou, China

OPEN ACCESS

Edited by:
Nianyin Zeng,
Xiamen University, China

Reviewed by:
Paocheng Huang,
Fujian Agriculture and Forestry University, China
Xiaoyan Hong,
University of Alabama, United States

***Correspondence:**
Cai Liangming
71960127@qq.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 22 July 2021
Accepted: 12 October 2021
Published: 16 November 2021

Citation:
Liangming C, Xiaoqiong C, Min D, Binxin M, Minfen L, Zhicheng Z, Shumin L, Yuxin R, Qiaolin H and Shuqin Y (2021) OSA Patient Monitoring Based on the Beidou System. Front. Public Health 9:745524. doi: 10.3389/fpubh.2021.745524

This paper presents an OSA patient interactive monitoring system based on the Beidou system. This system allows OSA patients to get timely rescue when they become sleepy outside. Because the Beidou position marker has an interactive function, it can reduce the anxiety of the patient while waiting for the rescue. At the same time, if a friend helps the OSA patients to call the doctor, the friend can also report the patient's condition in time. This system uses the popular IoT framework. At the bottom is the data acquisition layer, which uses wearable sensors to collect vital signs from patients, with a focus on ECG and SpO2 signals. The middle layer is the network layer that transmits the collected physiological signals to the Beidou indicator using the Bluetooth Low Energy (BLE) protocol. The top layer is the application layer, and the application layer uses the mature rescue interactive platform of Beidou. The Beidou system was developed by China itself, the main coverage of the satellite is in Asia, and is equipped with a high-density ground-based augmentation system. Therefore, the Beidou model improves the positioning accuracy and is equipped with a special communication satellite, which increases the short message interaction function. Therefore, patients can report disease progression in time while waiting for a rescue. After our simulation test, the effectiveness of the OSA patient rescue monitoring system based on the Beidou system and the positioning accuracy of OSA patients have been greatly improved. Especially when OSA patients work outdoors, the cell phone base station signal coverage is relatively weak. The satellite signal is well-covered, plus the SMS function of the Beidou indicator. Therefore, the system can be used to provide timely patient progress and provide data support for the medical rescue team to provide a more accurate rescue plan. After a comparative trial, the rescue rate of OSA patients using the detection device of this system was increased by 15 percentage points compared with the rescue rate using only GPS satellite phones.

Keywords: IoT–internet of things, OSA patient rescue system, Beidou indicator, STM32 microcontroller implementation, android

INTRODUCTION

Sleep respiratory diseases include obstructive sleep apnea (OSA) and central nervous system sleep apnea (CSA), etc., the research shows that 80% of the patients with sleep apnea syndrome have snoring phenomena (1). Long term sleep apnea can cause dysfunction of the body's system and multiple diseases, such as increasing the risk of obesity and heart disease, and even sudden death (2). According to the World Health Group, the high-risk groups of sleep respiratory diseases mainly involve obese patients and the elderly (3). In middle-aged men and women in the United States, OSA incidence is high, the number of patients with sleep disordered breathing in the world is a large group, and nearly 80–90% of them have apnea syndrome but they are often undiagnosed (4). Therefore, the monitoring and auxiliary regulation of sleep disordered breathing are of great significance for the prevention and diagnosis of sleep and related diseases. One method is to help patients achieve a normal Body Mass Index (BMI), but the patients cannot always return to normal because their perseverance is insufficient, so the method cannot achieve the goal of reducing sleep disorders (SD) (5).

There are ∼5,000 standard sleep laboratories in the world, and most sleep laboratories use polysomnography (PSG) as the gold standard for detecting sleep disorders (6). PSG can only be used in specialized clinics or hospitals under constant medical supervision to collect physiological signals from patients with sleep disorders, and then use professional sleep experts to diagnose patients who have sleep disorders based on physiological signals detected by PSG (7). This means that patients must go to specialized medical institutions, which will inevitably increase the burden on patients. Moreover, PSG requires the integration of many sensors on the human body, which is considered invasive, so PSG screening can interfere with sleep (8). In addition, PSG is costly and time consuming, because the PSG test needs professional sleep experts. Therefore, it would be difficult for PSG to become a method for long-term monitoring of patients with sleep disorders (9). Limited data show that patients with severe sleep apnea who do not receive effective treatment will increase the risk of sudden cardiac death. If there is no heart disease, only sleep apnea, the possibility of death is very small. Long-term sleep apnea, if not well-treated, will increase the incidence of a series

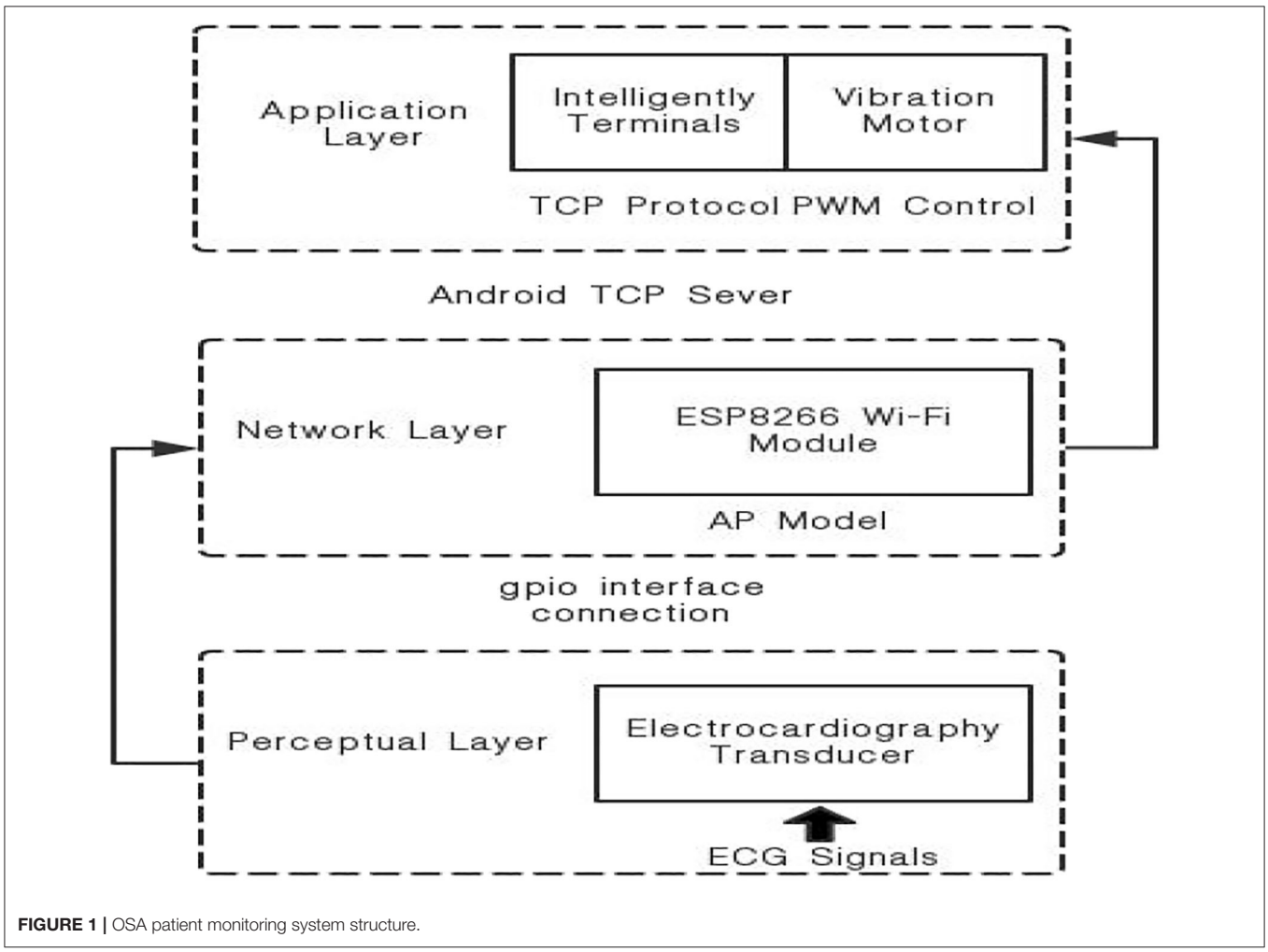


FIGURE 1 | OSA patient monitoring system structure.

of cardiovascular diseases including hypertension, pulmonary hypertension, coronary heart disease, heart failure, etc., and will also make the therapeutic effect of these diseases worse. For example, about 70% of refractory hypertension patients with poor blood pressure control have sleep apnea. Therefore, screening sleep apnea is very important for hypertension patients whose blood pressure is difficult to control. Sleep apnea may also increase the incidence of diabetes, fatty liver, stroke, and depression. It can make people sleepy and lack attention during the day, which also increases the possibility of accidents such as traffic accidents or the operation of other dangerous equipment.

Due to problems such as OSA not being effectively solved by surgery, and sleep correction using CPAP causing discomfort to patients with sleep disorders, the data suggests that 80% of the patients with sleep disorders have not been effectively diagnosed or treated in time. There is therefore a need for an Internet of

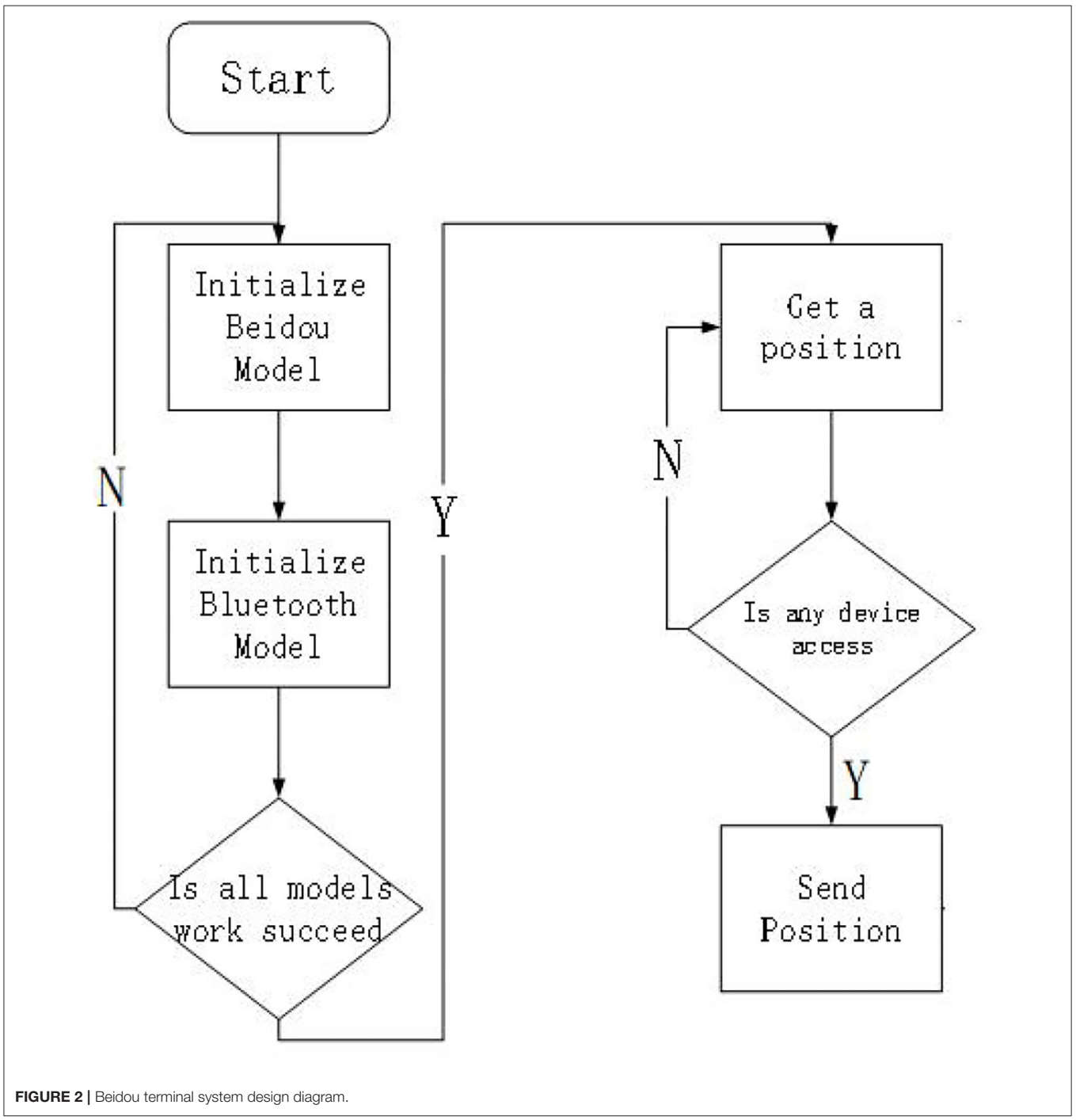


FIGURE 2 | Beidou terminal system design diagram.

Things (IoT) framework system to monitor OSA patients. This system enables OSA patients to be rescued in time when they are drowsy outdoors. Because the Beidou position marker has an interactive function, it can reduce the anxiety of the patient while waiting for rescue. At the same time, if a friend helps an OSA patient to call a doctor, the friend can also report the patient's condition in time.

The system uses a popular IoT framework. At the bottom is the data acquisition layer, which uses wearable sensors to collect vital signs from patients, with a focus on ECG and SpO2 signals. The middle layer transmits the collected physiological signals to the network layer of the Beidou indicator using the Bluetooth Low Energy (BLE) protocol. The top layer is the application layer, and the application layer uses the mature Beidou rescue interactive platform. Since the GPS indicator does not include a communication satellite, it does not have an SMS function. OSA patients can only passively wait for rescue. In addition, due to the lack of satellites in Asia and the insufficient density of ground-enhanced systems, OSA patients have large positioning errors. The Beidou system is developed by China itself, the main coverage of the satellite is in Asia, and is equipped with a high-density ground enhancement system. Therefore, the Beidou model improves the positioning accuracy and is equipped with a special communication satellite, which increases the short

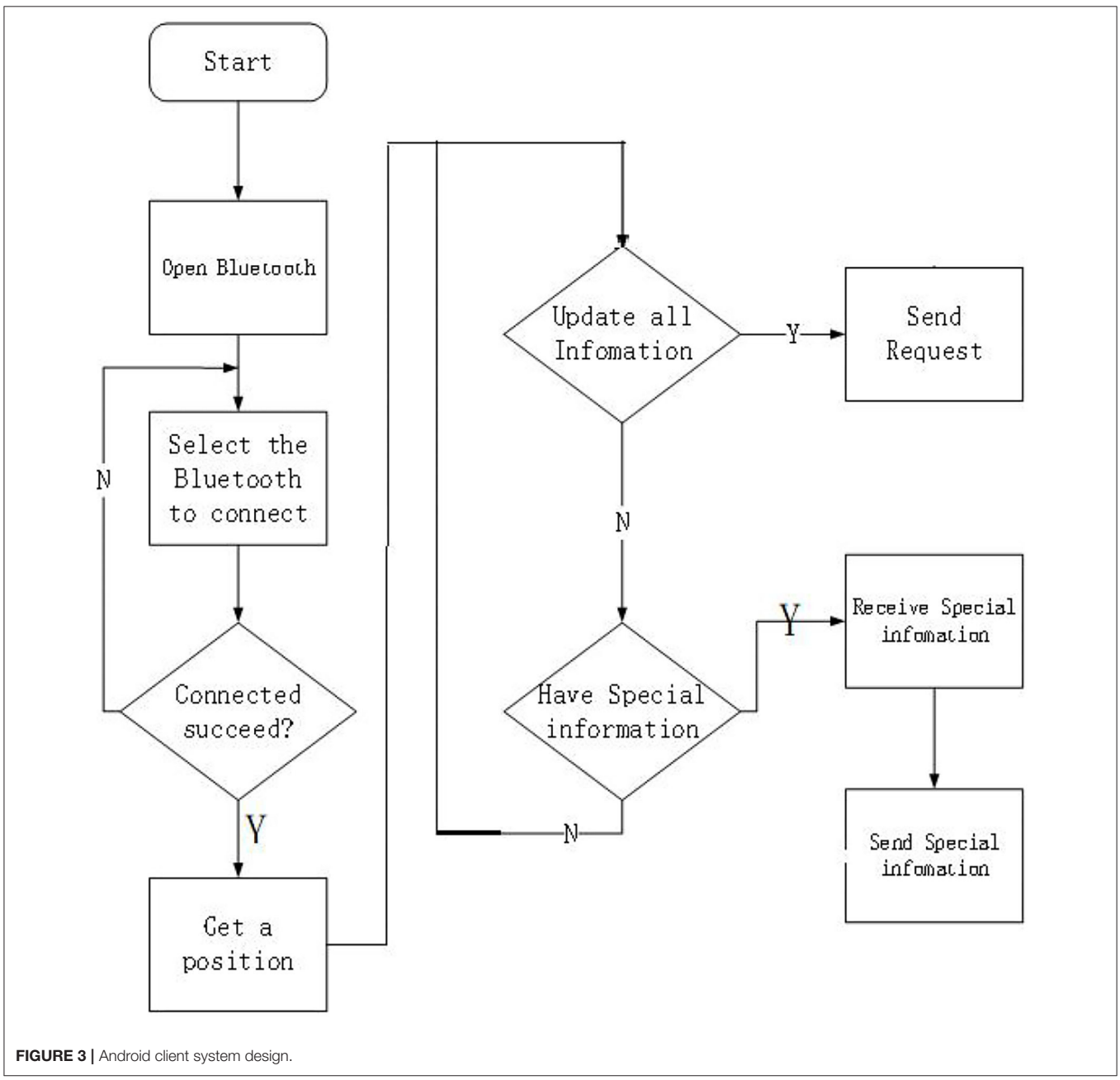


FIGURE 3 | Android client system design.

message interaction function. Therefore, patients can report disease progression in a timely manner while waiting for a rescue.

After our simulation test, the effectiveness of the OSA patient rescue monitoring system based on the Internet of Things framework and the positioning accuracy of OSA patients have been greatly improved. Especially when OSA patients work outdoors, the signal coverage of cell phone base stations is relatively weak. The satellite signal is well-covered, plus the SMS function of the Beidou indicator. As a result, the system can be used to provide timely patient progression and provide data support to the medical rescue team to provide a more accurate rescue plan. After a comparative test, the rescue rate of OSA patients using the system's detection equipment increased by 15 percentage points compared to the rescue rate using only GPS satellite phones.

The rest of this paper is structured as follows. Part 2 presents the latest literature on this area of research. Part 3 presents the methods and simulation results of a sleep disorder monitoring system based on the *Beidou system*. Section Experimental study summarizes and suggests the direction for future improvements.

RELATED WORK

At present, the golden standard for the diagnosis of sleep apnea is polysomnography (PSG) (10). However, PSG is expensive, so it is necessary for special sleep respiratory monitors to connect electroencephalogram (EEG), eye movement, EMG, and other wires to the body surface of patients, so their application is limited to a certain extent (11). In order to reduce the measured physiological signals, the detection method of sleep dyspnea based on SpO2, ECG (Electrocardiogram) signal has become a hot research topic (12).

In the research of detecting sleep dyspnea based on ECG signal, the methods of OSA detection of ECG signal were studied

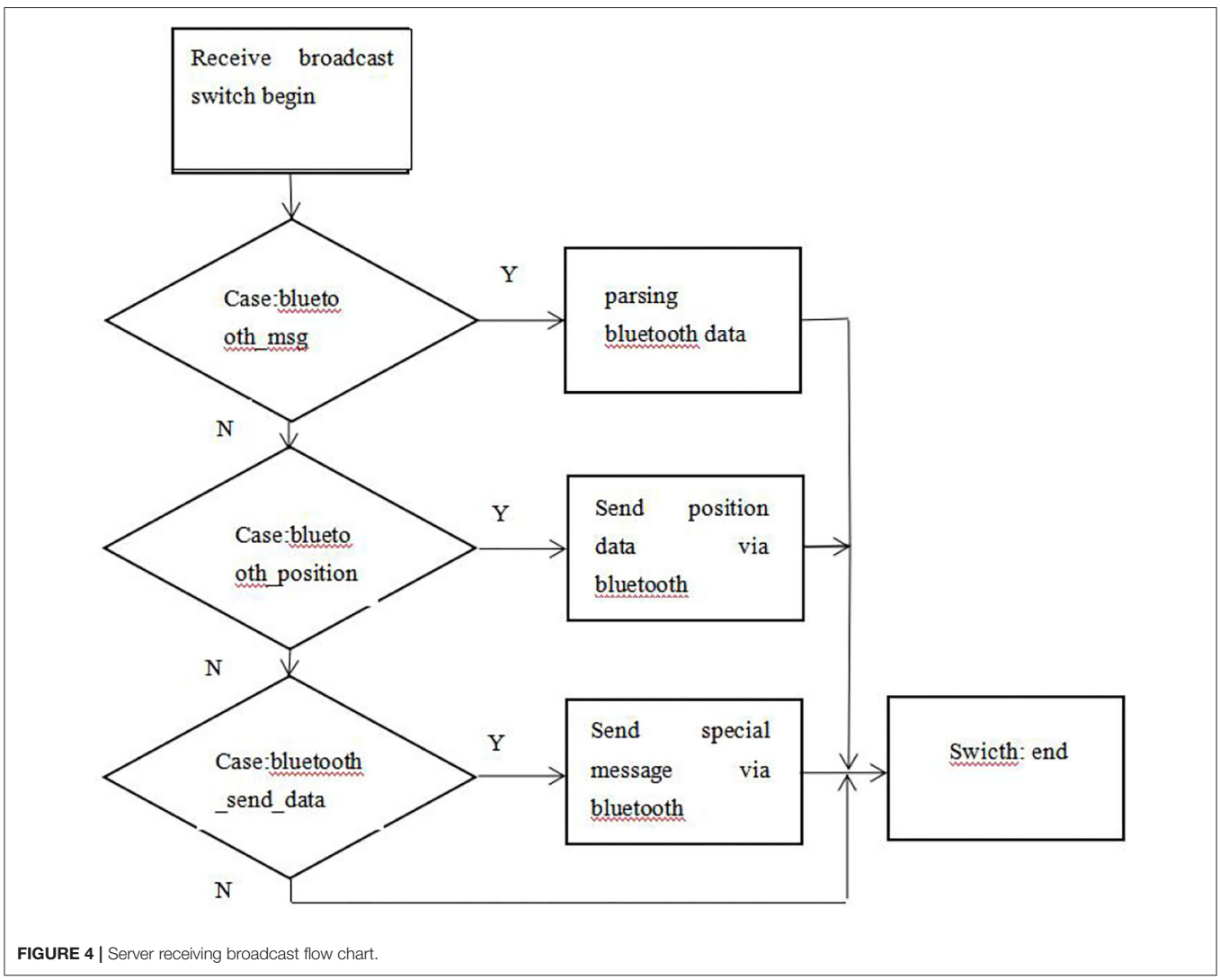


FIGURE 4 | Server receiving broadcast flow chart.

by the University of Wisconsin, Taiwan Jiaotong University and Peking University (13–18). At present, there are several CFDA certified devices in China to diagnose sleep dyspnea by sticking electrodes in the human body (19). The latest research shows that by monitoring the ECG signal, sending it to the neural network through frequency domain and time domain analysis,

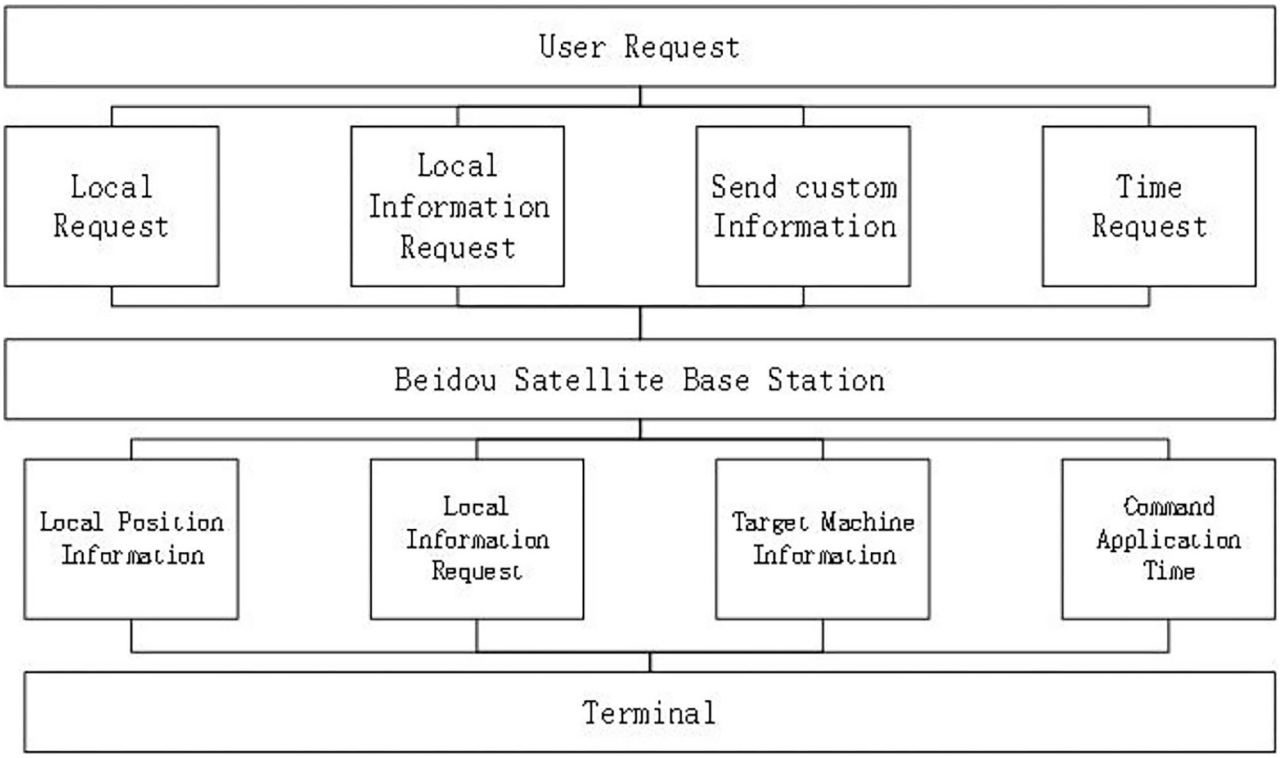


FIGURE 5 | Beidou terminal function service flow chart.

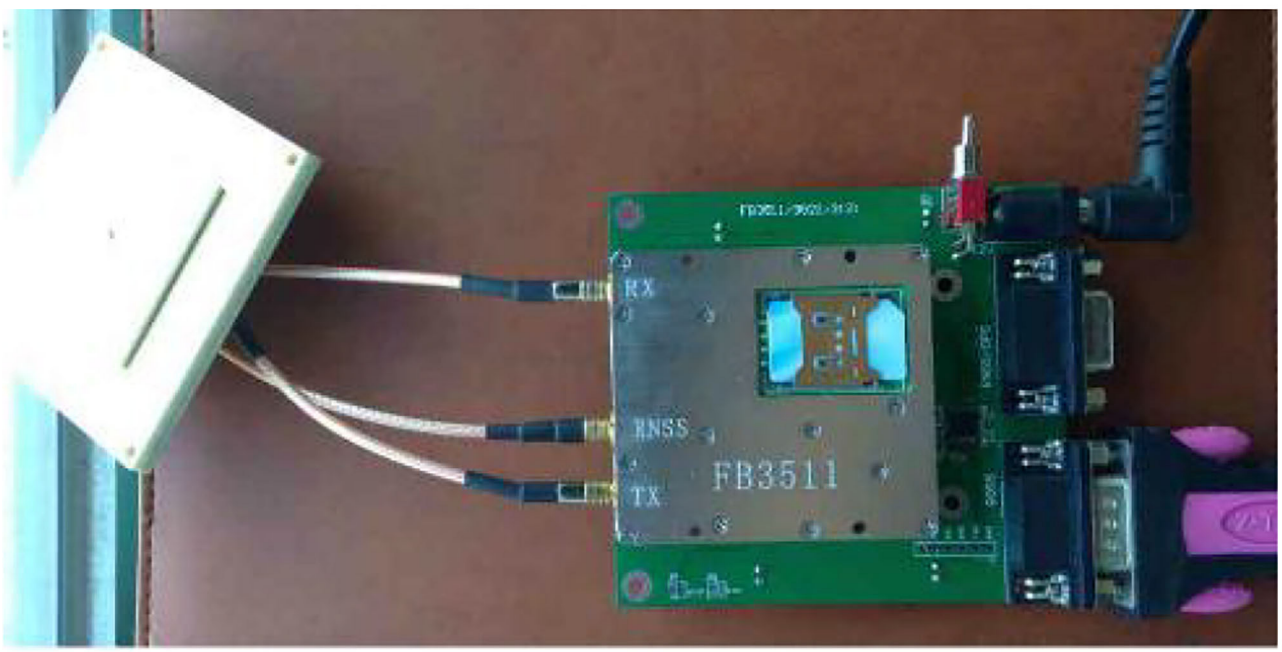


FIGURE 6 | Beidou interactive location-indicating machine.

and establishing a classification model through statistical analysis of the apnea-hypopnea index (AHI), the automatic detection of SDB can be realized to a certain extent (20). However, ECG signal acquisition requires patients to wear at least two ECG electrodes for dynamic monitoring for a long time, which brings serious physical and mental burden to patients. Therefore, finding a physiological signal that can replace ECG and is easier to obtain has become a research direction. Like the ECG signal, there is a slight difference in the interval between consecutive pulse waves. Pulse rate variability (PRV) is an analysis method to study the physiological conditions of the human body that these small changes can reflect. Studies have shown that using photoplethysmography (PPG) signals to analyze PRV in healthy people can replace HRV to reflect changes in the autonomic nervous system (ANS), and there is a good correlation between the two (21). Further studies have confirmed that PRV in SDB patients also has a good correlation with HRV, thus shifting SDB detection research from ECG signals to PPG signals that are easier to obtain (22–26). Amir et al. (27) conducted PSG synchronous PPG monitoring on 74 volunteers and found that the use of PPG signals to determine sleep breathing status indicated that PPG and PSG have a good consistency and can be used for clinical sleep breathing monitoring.

The Beidou satellite navigation system provides positioning, navigation, and timing services (28), which are divided into open service (29) and authorized service (30). Licensing services: in addition to free and open services to the world, there are services that require an authorization, which is divided into different levels and is distinguished between military and civilian use (31). The Beidou satellite navigation system is very important when calling for help and position when in distress at sea. In view of this situation, IMO has developed a set of global maritime distress and safety systems (GMDSS) (32). The systems consist of three major parts: INMARSAT (Inmarsat system) (33) and COS-PAS-SARSAT (Polar-orbiting Satellite Search and Rescue System) (34), ground-based radio communication system (i.e., Coast Station) (35), and Maritime Safety Information dissemination system (36).

The above research proves the feasibility of using machine learning to study the assistance positioning system for OSA patients. However, most of the above studies are based on small sample data sets and lack large-scale clinical applications. In addition, too many features may be selected in the feature engineering, resulting in too many calculations. Or there are too few functions selected, resulting in low classification accuracy. In addition, when selecting a deep learning classification method, manually setting parameters is too inefficient. At the same time, it may overtrain and lead to test overfitting and, because most research methods are only based on small data sets, in practical applications, the generalization ability of this model is low. Therefore, this topic aims to use machine learning algorithms on the cloud platform to standardize rescue positioning data from different devices.

TABLE 1 | Success effect of Beidou sending information.

	Time	Longitude	Latitude
T1	12:32:39	119°46'46.6"	46°46'37.6"
T2	12:34:37	119°46'45.9"	46°46'36.9"
T3	12:36:38	119°46'45.8"	46°46'37.2"
T4	12:38:38	119°46'46.3"	46°46'37.3"
T5	12:40:39	119°46'46.1"	46°46'37.1"
T6	12:42:38	119°46'46.6"	46°46'37.5"
T7	12:44:39	119°46'46.6"	46°46'37.3"
T8	12:46:39	119°46'46.5"	46°46'37.2"
T9	12:48:39	119°46'46.4"	46°46'37.3"
T10	12:50:39	119°46'46.5"	46°46'37.3"

METHOD AND SIMULATION RESULTS

The hardware component of the IOT-based sleep disorder monitoring system consists of three parts. The bottom layer is the hardware terminal of the indicator mark, the middle layer is the transmission medium, and the upper layer is the application platform. The beacon terminal includes an ARM main control board, a 4th generation mobile communication module, a Beidou communication module, a Bluetooth communication module, an antenna interface module, and a power management module. In addition, the sides of the housing are provided. There are data transmission interfaces, SOS buttons, and charging ports. There are several sealing rings and fixing screws at the bottom of the housing. However, when a sleep disorder patient is working outdoors, when the drowsiness state occurs, the SOS button can be used to send the help information to the Beidou satellite. Through the auxiliary ground-based augmentation system, Beidou Satellite sends the current status and position coordinates of the sleep-disabled patients to the Beidou rescue platform. The information released by the rescue platform enables the rescue team to quickly and accurately find the location to be rescued. As shown in **Figure 1**.

Design of the Perceptual Layer

Hardware Design

The whole system is based on ARM main control board, 4th generation mobile communication module, Beidou communication module, Bluetooth communication module, antenna interface module and power management module. In addition, the sides of the housing are provided. There are data transmission interfaces, SOS buttons and charging ports. There are several sealing rings and fixing screws at the bottom of the housing. However, when a sleep disorder patient is working outdoors, when the drowsiness state occurs, the SOS button can be used to send the help information to the Beidou satellite. The Beidou second-generation passive positioning and Beidou active short message communication are combined. The location information of OSA patients can be obtained through the Beidou module, while OSA patients can send disease information through Android mobile phone software operation. The emergency call function is provided in an emergency, and the OSA patient can send his own location information and simple condition information, thereby realizing the emergency

rescue function of the OSA patient monitoring system based on the IOT framework.

Function Design of Obtaining Position Information

The STM32 development board communicates with the Beidou module by serial port. It uses the command of RNSS format and uses BDGGA to get its own location information and stores it in the cache. When a mobile phone is connected with the terminal through Bluetooth, the data is forwarded to the mobile phone if the mobile phone makes a data request. The mobile phone is transformed into latitude and longitude coordinates through the received Beidou position information and displayed.

Design of the Network Layer

Functional Design of the Mobile Phone

OSA patients can use the Android phone to send AT commands to initialize the Beidou indicator, and obtain their own location information through BLE communication. The latest rescue information and the latest location information obtained through the refresh function. The OSA can also provide a special message input function through the Beidou indicator, input the condition information through the mobile phone and forward it to the terminal, and then the condition message and location information can be sent to the Beidou rescue platform by the Beidou indicator.

Design of Communication Protocol Between End and Terminal of Mobile Phone

It is very necessary for the mobile terminal to communicate with the terminal using Bluetooth to maintain complete communication. The mobile phone terminal obtains whether the communication is normal by sending instructions, and the terminal receives the correct return value by continuously sending test information, and providing an LED flicker function display.

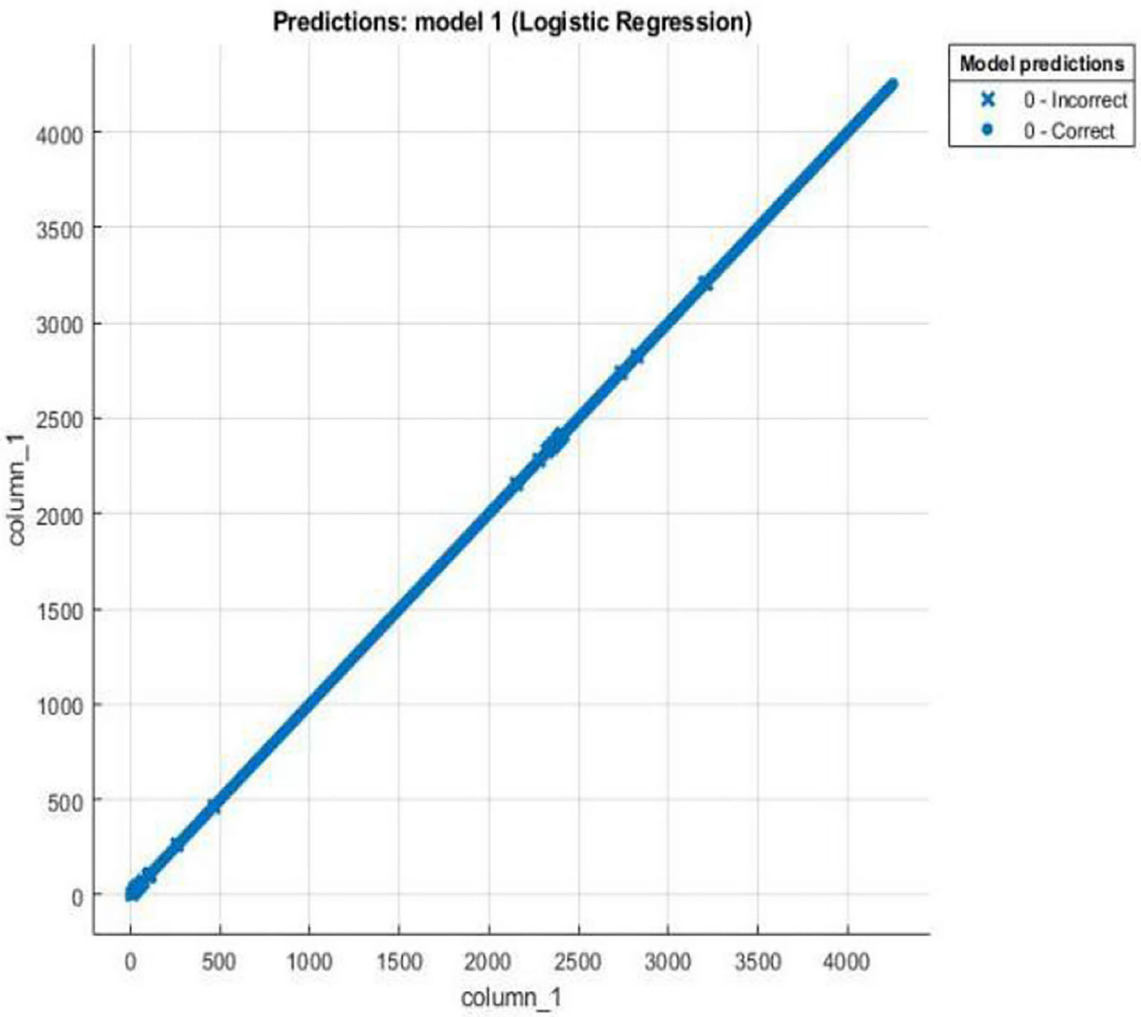


FIGURE 7 | Scatter plot from logistic regression.

Terminal Automatic Help Function

By setting the timer, the function of obtaining the position can be used automatically, and the acquired position information can be sent periodically. The B4 function can be started by triggering the button B3 automatic function to stop the automatic function.

The main consideration is to obtain location information, and how to send information for help and other aspects. In the event of a problem, the rescue party can get the message in time, and the suitor can also get the necessary information anytime and anywhere. The interactive function of the display object is mainly realized by Bluetooth module and STM32 main control board.

Software Design

Terminal Part System Design

Figure 2 illustrates the initialization process of the Beidou indicator, which provides a simple led indicator during normal operation. When you are able to communicate properly, you can view location information from your phone software.

Android Client Design

Figure 3 illustrates the initialization process for the Android phone side of OSA patients. The Beidou indicator mark sends a simple condition message to the mobile terminal, and presses the button SOS on the Beidou terminal to immediately send the condition information to the fixed number already stored in the Beidou indicator. The short message content includes Beidou location information and simple disease information. The Beidou indicator uses the Beidou module to send information. OSA patients can use 4G mobile phones to update data and display location and SMS status information for OSA patients.

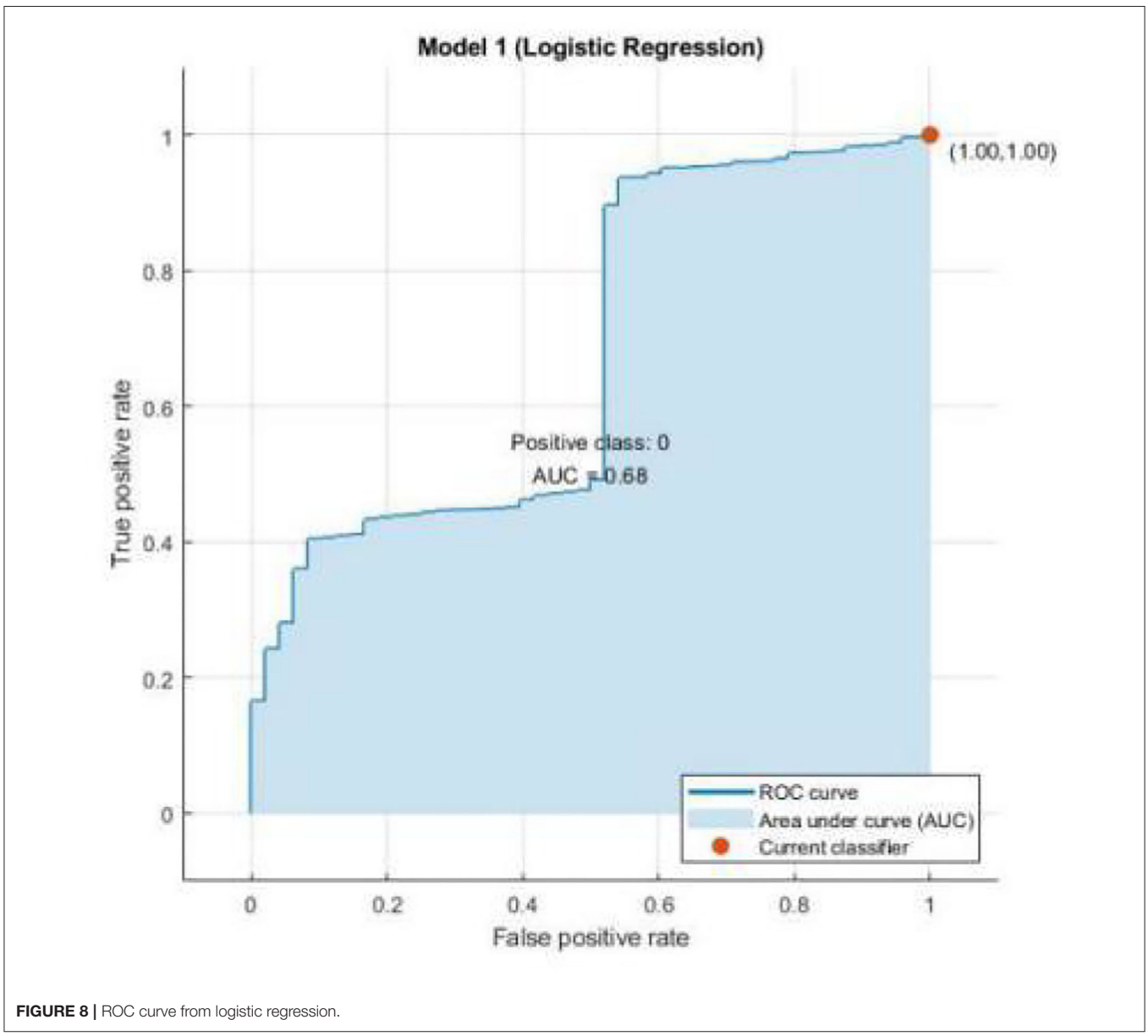


FIGURE 8 | ROC curve from logistic regression.

Data Transmission

The Server Reception Broadcast Process

The default encoding mode for Android phones is utf-8, which is used for serial communication to receive and send gdb32 encoded Chinese data. When you need to convert an object to a String object, the encoding system may produce inconsistencies, so the server also encodes and converts characters when performing transceiver control. The Server still does not stop when the Activity application is paused, so receiving information should not stop when the application is not in use at this time. Activity needs to bind to Server (**Figure 4**).

Terminal function service flow, as shown in **Figure 5**.

EXPERIMENTAL STUDY

Implementation Platform

The Beidou indicator uses the STM32F103RBT chip development board. The code running on the beacon is compiled on the mdk4.9 platform and downloaded using the CooCox development tool. Android phones use android stdio as the development platform for the Android sdk version. The entire mark is shown in **Figure 6**.

Android Client

The Beidou indicator uses a switch control to control the switch of the Bluetooth module. If Bluetooth is turned on, you can choose to connect to the Beidou.

Send success messages and location for success as shown in **Table 1**.

Logistic regression is a statistical tool based technique to deal with machine learning problems. We evaluated the results using logistic regression. Logistic regression works with sigmoid function. the accuracy of the LR model is very high (90%). In particular, LR models using all 10 features show better accuracy than other existing models using RQA functionality with an accuracy rate of 90%.

Most importantly, models using LR are still the best in terms of prediction accuracy, precision, recall, and F-measure.

The results of scatter plot from logistic regression is shown in **Figure 7**. It is very clear from the figure that logistic regression has performed efficiently to solve the issue of OSA patient monitoring. The ROC curve is shown in **Figure 8**. Parallel co-ordinate plot also shows that the prediction is accurate. From the confusion matrix in **Figure 9**, it can be seen that the accuracy is 90%. Moreover, the figure also shows the true positive rate and false negative rate. The **Figure 10** shows that positive predictive value is 90 % and false discovery rate is 10%. **Figure 11** consists of true class and predicted class.

A customized fitting model is shown in **Figure 12**. The coefficients of the model are a = 6.989, b = 0.0002727 and c = 247.1 (237.6, 256.7). The R-square value: 0.9222 and root mean squared error is 382.8.

Figure 13 shows that the result is also fitted with exponential fitting [$f(x) = a^*\exp(b * x)$]. The coefficients of the model are a = 430.5, and b = 0.0005877. The R-square value: 0.8975 and root mean squared error is 439.5. The Gauss fitting model is shown in **Figure 14**. The equation of gauss fitting model is $f(x) = a1^*\exp[-((x-b1)/c1)^2]$. The coefficients

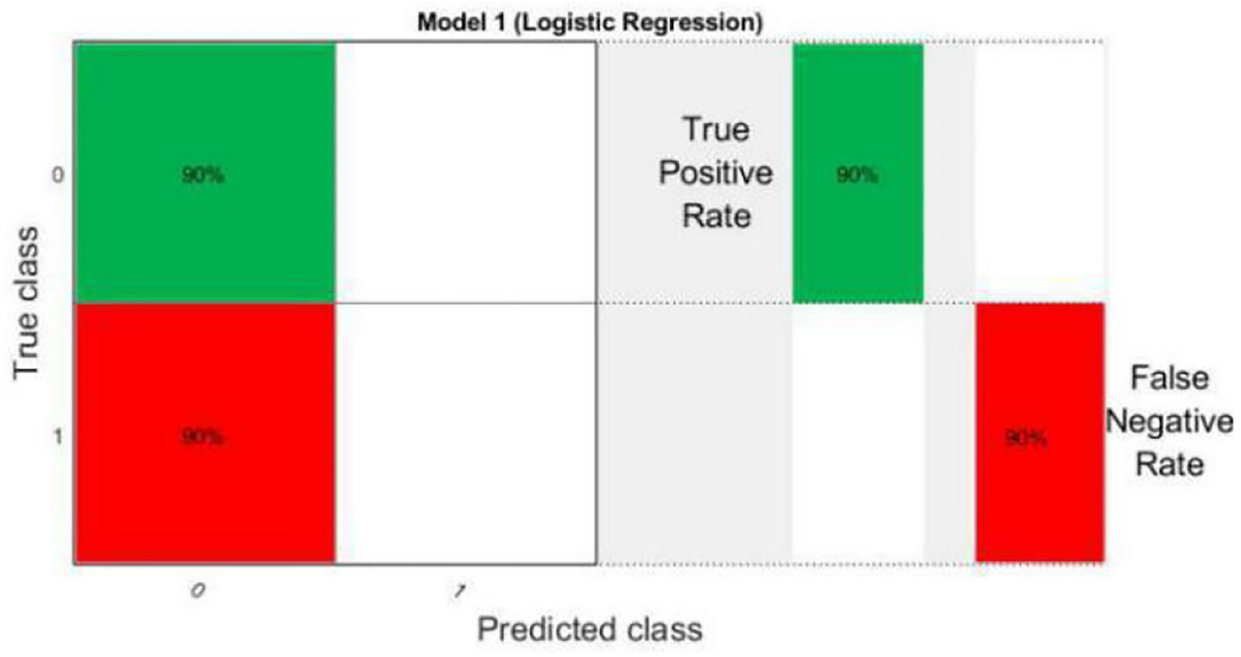


FIGURE 9 | Confusion matrix plot from logistic regression.

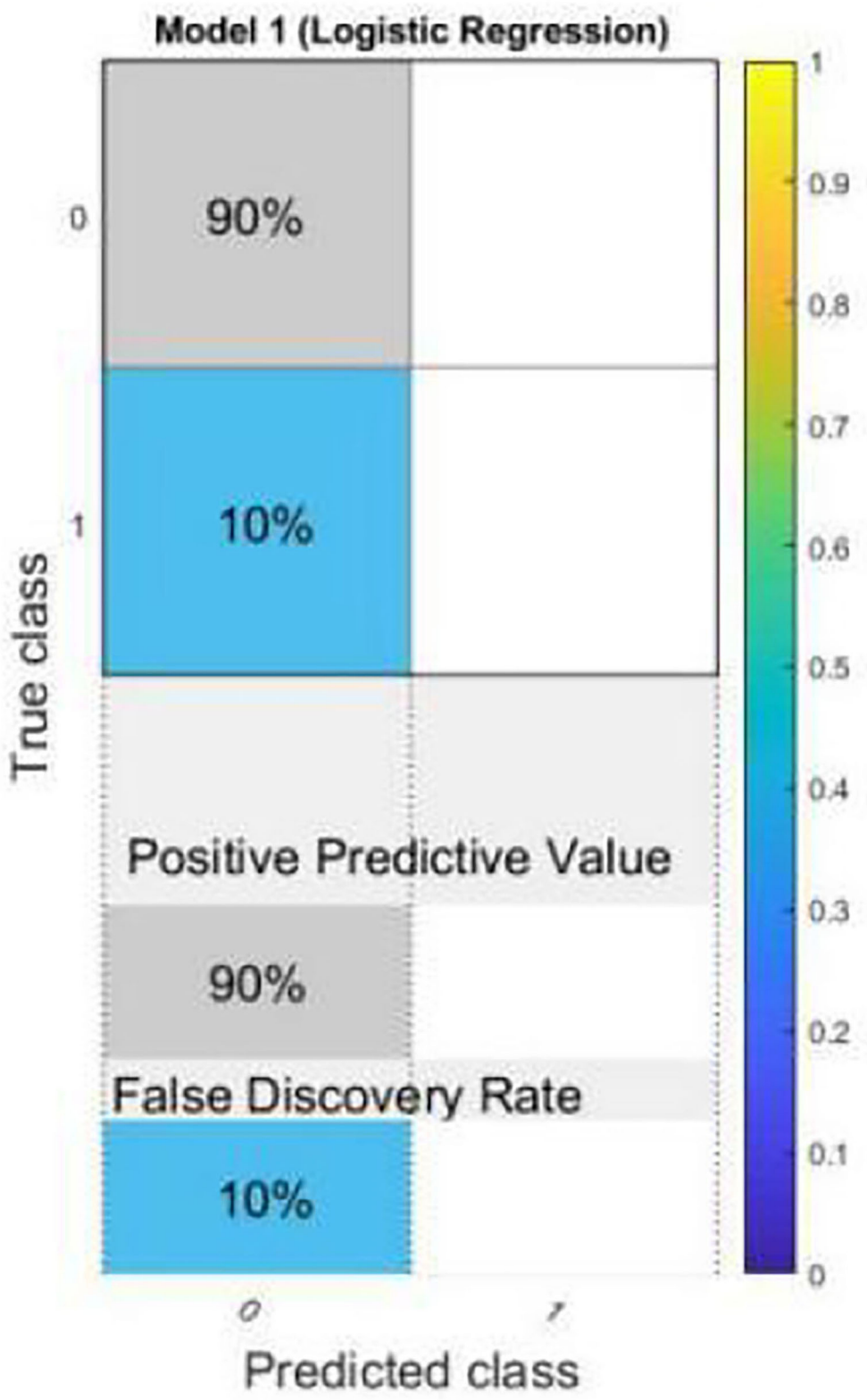


FIGURE 10 | Confusion matrix plot with positive predictive value and false discovery rate from logistic regression.

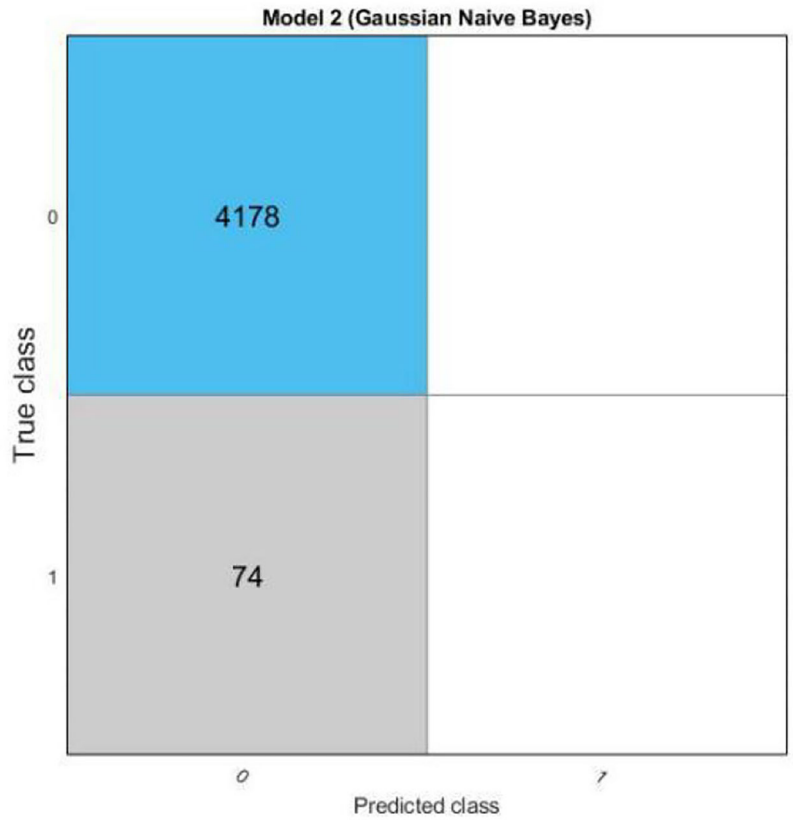


FIGURE 11 | Confusion matrix plot with true class and predicted class from logistic regression.

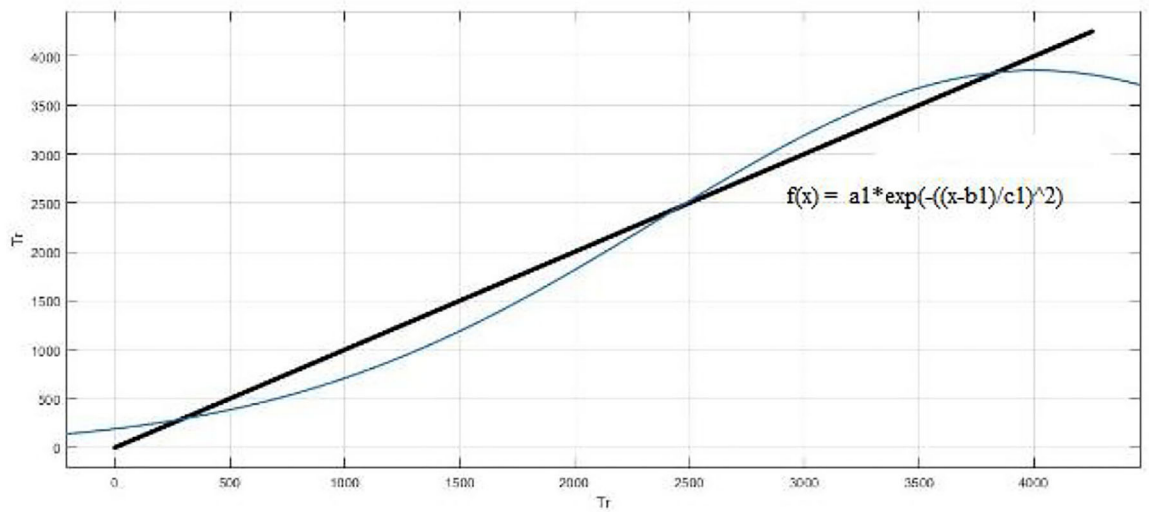


FIGURE 12 | Customized fitting model with $f(x) = a^*[\sin(x-pi)] + b^*[(x-10)^2] + c$.

of the model are a1 = 3855, b1 = 4005, and c1 = 2310. The R-square value: 0.9795 and root mean squared error is 196.6.

The motivation of the current work is to locate and rescue patients with OSA. Whether the positioning is accurate or not is actually a problem of binary classification of rescue targets. We

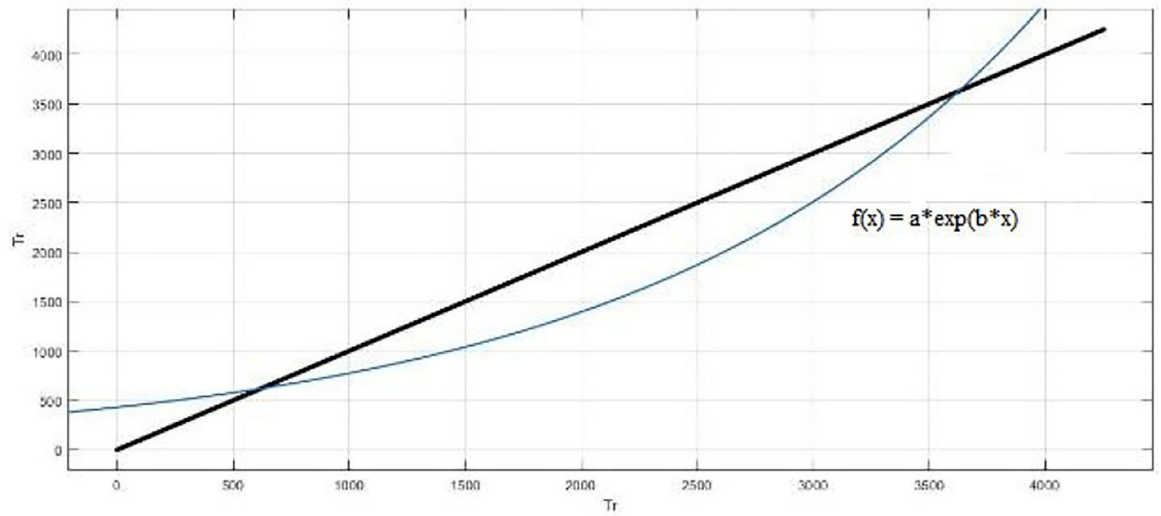


FIGURE 13 | Exponential fitting model.

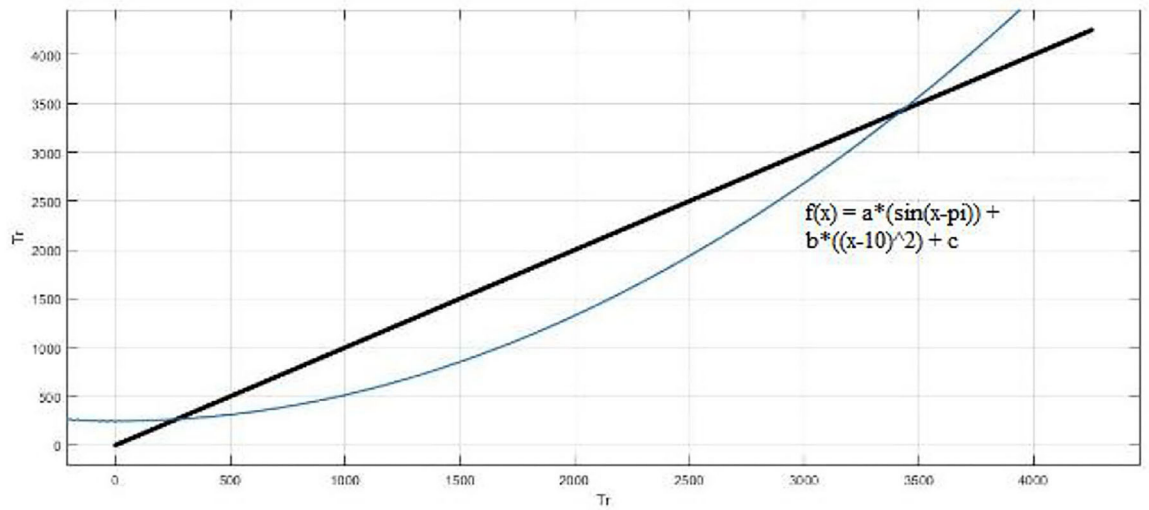


FIGURE 14 | Gauss fitting model.

use logistic regression methods to create classification models. The test result data set is divided into training and test data. Among them, 70% of the data is used for training the model, and 30% of the data is used for the test results. The simulation results confirmed that logistic regression can be used to classify the rescue positioning results of patients with sleep disorders. The positioning accuracy of this system is 90%.

CONCLUSIONS

The OSA sleep disorder patient monitoring system is based on the IoT framework structure. The entire system provides real-time disease reporting information, precise location information, and physician assistance information for OSA patients. The terminal hardware is realized by the Beidou indicator module, the 4G mobile communication module, and the BLE communication module. Through the secondary development of the Android mobile phone SDK development kit, it is convenient for users to operate. The special operation of shielding the Beidou indicator is made to make this OSA patient monitoring and rescue system easier to integrate into the existing hospital emergency system. The system currently only provides the Android mobile phone client as the input and receiving end of the message, which is convenient for the user, who is able to interact with the doctor in real time through the smart phone. The OSA monitoring system has self-healing capabilities and is extremely robust, providing system reliability. This system test shows that the OSA patient system we developed has lower power consumption and

a simpler hardware composition. In addition, our OSA patient monitoring and rescue system has a stable and simple operation while meeting design requirements. In short, our system is an innovative application of IoT technology in modern medicine, with a particular focus on the effective management of OSA patients. In the future, we will explore how to further improve system performance by integrating more sensors to monitor OSA patients in a more comprehensive manner.

The simulation results of this research show that the use of machine learning techniques (such as logistic regression) is very suitable for building a classifier for the accuracy of the rescue positioning of patients with sleep disorders. LR has the best performance in all aspects of prediction accuracy, accuracy, recall, and prediction.

These high-accuracy results will encourage researchers to more rigorously study the accuracy of sleep disorder rescue positioning data, and use logistic regression for binary output in more practical classification problems. In the next phase or research, we plan to use deep learning, convolutional neural networks, and transfer the learning to the classification of sleep disorder rescue positioning accuracy. In addition, it aims to collect data from hospitals and use machine learning algorithms to evaluate and predict the accuracy of sleep disorder rescue positioning. It is possible to manufacture devices with IOT to monitor sleep abnormalities in real time.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

CL, CX, and DM conceived the study, designed the experiments, analyzed the data, and wrote the whole manuscript. MB and LM provided the preprocessed data. ZZ, LS, and RY carried out experiments. HQ helped to analyze the data and experiments result. YS revised the manuscript.

FUNDING

This work was supported by the Fujian Provincial Scholarship Council, the National Science Foundation of China (No. 61773124), the Fujian Social Science Foundation (FJ2021BF029), the Fujian Natural Science Foundation (2018J01796), the Fujian Science and Technology Department Foundation (2019I1009), Project of Youth Foundation of Fujian Education Department (JAT191104), the Foundation of Fuzhou University Zhicheng College (KJ20190001), and the Foundation of National College Student Innovation and Entrepreneurship Project (ZJ1934). In part by The Open Fund of Provincial Key Laboratory of Eco-Industrial Green Technology-Wuyi College.

REFERENCES

1. Mencar C, Gallo C, Mantero M, Tarsia P, Carpagnano GE, Foschino Barbaro MP, et al. Application of machine learning to predict obstructive sleep apnea syndrome severity. *Health Informatics J*. (2019) 1:1–20. doi: 10.1177/1460458218824725
2. Nikolaidis K, Kristiansen S, Goebel V, Plagemann T. Learning from higher-layer feature visualizations. *Preprints arXiv*. (2019) arXiv:1903.02313v1.2019.
3. Ravichandran V, Murugesan B, Balakarthikeyan V, Shankaranarayana SM, Ram K, Preejith PS, et al. RespNet: a deep learning model for extraction of respiration from photoplethysmogram. *Annu Int Conf IEEE Eng Med Biol Soc*. (2019) 2019:5556–9. doi: 10.1109/EMBC.2019.8856301
4. Van Steenkiste T, Groenendaal W, Deschrijver D,Dhaene T. Automated sleep apnea detection in raw respiratory signals using long short-term memory neural networks. *IEEE J Biomed Health Inform*. (2019) 23:2354–64. doi: 10.1109/JBHI.2018.2886064
5. Cen L, Yu ZL, Kluge T, Ser W. System for obstructive sleep apnea events detection using convolutional neural network. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2018). p. 3975–8. doi: 10.1109/EMBC.2018.8513363
6. Islam SMS, Hassan M, Ali Al-Jumaily A, Claxton S. Deep learning of facial depth maps for obstructive sleep apnea prediction. In: *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*. IEEE (2018). p. 154–7. doi: 10.1109/iCMLDE.2018.00036
7. De Falco I, De Pietro G, Sannino G, Scafuri U, Trunfio GA. Deep neural network hyper-parameter setting for classification of obstructive sleep apnea episodes. In: *2018 IEEE Symposium on Computers and Communications (ISCC)*. IEEE (2018).
8. Lakhan P, Ditthapron A, Banluesombatkul N, Wilaiprasitporn T. Deep Neural networks with weighted averaged overnight airflow features for sleep apnea-hypopnea severity classification. *Preprint arXiv*. (2018) arXiv:1808.10845v1. doi: 10.1109/TENCON.2018.8650491
9. McCloskey S, Haidar R, Koprinska I, Jeffries AB. Detecting hypopnea and obstructive apnea events using convolutional neural networks on wavelet spectrograms of nasal airflow. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Cham: Springer (2018). p. 361–72. doi: 10.1007/978-3-319-93034-3_29
10. Jansen C, Hodel S, Penzel T, Spott M, Krefting D. Feature relevance in physiological networks for classification of obstructive sleep apnea. *Physiol Meas*. (2018) 39:124003. doi: 10.1088/1361-6579/aaf0c9
11. Urtnasan E, Park JU, Lee KJ. Multiclass classification of obstructive sleep apnea/hypopnea based on a convolutional neural network from a single-lead electrocardiogram. *Physiol Meas*. (2018) 39:065003. doi: 10.1088/1361-6579/aac7b7
12. Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag*. (2012) 29:82–97. doi: 10.1109/MSP.2012.2205597
13. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015) 521:436–44. doi: 10.1038/nature14539
14. Pathinarupothi RK, Vinaykumar R, Rangan ES, Gopalakrishnan EA, Soman KP. Instantaneous heart rate as a robust feature for sleep apnea severity detection using deep learning. In: *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE (2017). p. 293–6. doi: 10.1109/BHI.2017.7897263
15. Pathinarupothi RK, Prathap D, Rangan ES, Gopalakrishnan EA. Single sensor techniques for sleep apnea diagnosis using deep learning. In: *2017 IEEE International Conference on Healthcare Informatics*. IEEE (2017). p. 524–9. doi: 10.1109/ICHI.2017.37
16. Lan L, You L, Zhang Z, Fan Z, Zhao W, Zeng N, et al. Generative adversarial networks and its applications in biomedical informatics. *Front. Public Health*. (2020) 8:164. doi: 10.3389/fpubh.2020.00164

17. Zeng N, You Y, Xie L, Zhang H, Ye L, Hong W, et al. A new imaged-based quantitative reader for the gold immunochromatographic assay. *Optik*. (2018) 152:92–9. doi: 10.1016/j.ijleo.2017.09.109
18. Zeng N, Zhang H, Li Y, Liang J, Abdullah M. Dobaie. Denoising and deblurring gold immunochromatographic strip images via gradient projection algorithms, *Neurocomputing*. (2017) 247:165–172. doi: 10.1016/j.neucom.2017.03.056
19. Siqueira LC, Jacob AFL, de Santana ÁL. *Multimodal Low-Invasive System for Sleep Quality Monitoring and Improvement BT–Beyond the Internet of Things: Everything Interconnected*. Batalla JM, Mastorakis G, Mavromoustakis CX, Pallis E, editors. Cham: Springer International Publishing. (2017). p. 223–42. doi: 10.1007/978-3-319-50758-3_9
20. Nam Y, Kim Y, Lee J. Sleep monitoring based on a TriAxial accelerometer and a pressure sensor. *Sensors*. (2016) 16:750. doi: 10.3390/s16050750
21. Bulte CSE, Keet SWM, Boer C, Bouwman RA. Level of agreement between heart rate variability and pulse rate variability in healthy individuals. *Eur J Anaesthesiol*. (2011) 28:34–8. doi: 10.1097/EJA.0b013e32834088c4
22. Xianghua Q, Shoushui W, Chengyu L, Jing T, Shuangyan L. Comparison between heart rate variability and pulse rate variability during different sleep stages for sleep apnea patients. *Technol Health Care*. (2017) 25:435–45. doi: 10.3233/THC-161283
23. Khandoker AH, Karmakar CK, Palaniswami M. Comparison of pulse rate variability with heart rate variability during obstructive sleep apnea. *Med Eng Phys*. (2011) 33:204–9. doi: 10.1016/j.medengphy.2010.09.020
24. Dehkordi P, Garde A, Karlen W, Wensley D, Ansermino JM, Dumont GA. Pulse rate variability compared with heart rate variability in children with and without sleep disordered breathing. *Conf Proc IEEE Eng Med Biol Soc*. (2013) 2013:6563–6. doi: 10.1109/EMBC.2013.6611059
25. Zeng N, Zuo S, Zheng G, Ou Y, Tong T. Editorial: artificial intelligence for medical image analysis of neuroimaging data. *Front Neurosci*. (2020) 14:480. doi: 10.3389/fnins.2020.00480
26. Zeng N, Wang Z, Li Y, Du M, Liu X. Inference of nonlinear state-space models for sandwich-type lateral flow immunoassay using extended Kalman filtering. *IEEE Trans Biomed Eng*. (2011) 58:1959–66 doi: 10.1109/TBME.2011.2106502
27. Amir O, Barak-Shinar D, Henry A, Smart FW. Photoplethysmography as a single source for analysis of sleep-disordered breathing in patients with severe cardiovascular disease. *J Sleep Res*. (2012) 21:94–100. doi: 10.1111/j.1365-2869.2011.00927.x
28. Qingwu T. *Design and implementation of a monitoring system based on Beidou/GPS engineering transportation vehicles* (Master's thesis). Hunan University, Changsha, China (2017). Available online at: https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201802&filename=1018039753.nh
29. Light LX. *China's Satellite Industry Depth Studies*. Shanghai Shenyin Wanguo Securities Research Institute Co (2015). p. 33–9.
30. Zihui Y, Bin X. The development history and development trend of Beidou satellite navigation system. *J Navig Posit*. (2022) doi: 10.16547/j.cnki.10-1096.20220101
31. Min Y, Baochen S. Beidou satellite navigation system officially provides regional services. *Satellite Appl*. (2013) 5–8.
32. Yu M. *STM32 Self-Study Notes*. Beihang University Press (2015).
33. Richardson M, Wallace S. *Fall in Love With a Raspberry PI*. Lee where the Greek Translation of Science Books (2015). p. 120–50.
34. St. Stmicroelectronics. *STM32F10X Firmware Library Chinese explanation of V2.0[eb/ol].um0427*. User Manual 5:200–35.
35. *The AT Command Reference [EB/OL]*. Microsoft Technical Support (2014). p. 88–95.
36. *The Android Overview [EB/OL]*. Open Handset Alliance (2015). p. 180–205.

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