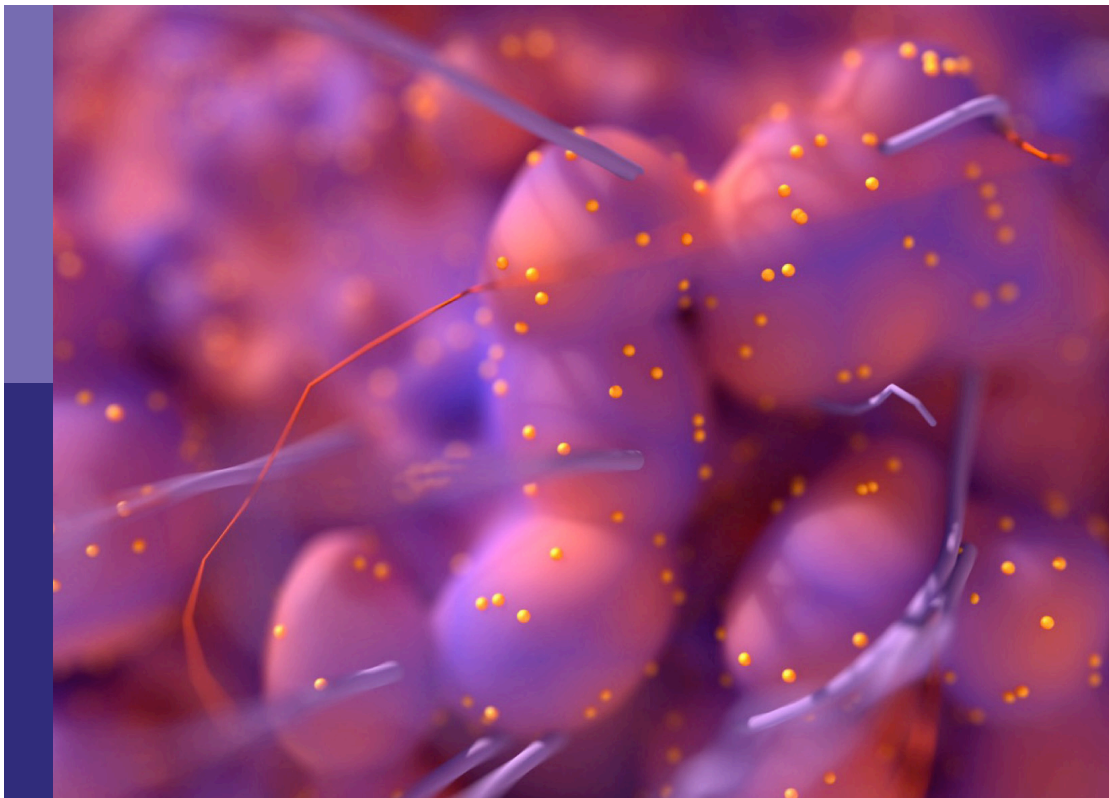
frontiers | Research Topics

Deep learning approaches in image-guided diagnosis for tumors

Edited by
Shahid Mumtaz, Victor Hugo C. Alburquerque and Wei Wei

Published in
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-83251-569-3
DOI 10.3389/978-2-83251-569-3

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

Deep learning approaches in image-guided diagnosis for tumors

Topic editors

Shahid Mumtaz — Instituto de Telecomunicações, Portugal

Victor Hugo C. Alburquerque — Federal University of Ceara, Brazil

Wei Wei — Xi'an University of Technology, China

Citation

Mumtaz, S., Alburquerque, V. H. C., Wei, W., eds. (2023). *Deep learning approaches in image-guided diagnosis for tumors*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-83251-569-3

Table of contents

frontiers in Oncology

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fonc.2022.860532

An Assisted Diagnosis Model for Cancer Patients Based on Federated Learning

Zezhong Ma[1,2,3,4], Meng Zhang[3,5], Jiajia Liu[4], Aimin Yang[1,2,3,4,5], Hao Li[3,5], Jian Wang[3,5], Dianbo Hua[6] and Mingduo Li[7,8]*

[1] Hebei Engineering Research Center for the Intelligentization of Iron Ore Optimization and Ironmaking Raw Materials Preparation Processes, North China University of Science and Technology, Tangshan, China, [2] Hebei Key Laboratory of Data Science and Application, North China University of Science and Technology, Tangshan, China, [3] The Key Laboratory of Engineering Computing in Tangshan City, North China University of Science and Technology, Tangshan, China, [4] College of Science, North China University of Science and Technology, Tangshan, China, [5] Tangshan Intelligent Industry and Image Processing Technology Innovation Center, North China University of Science and Technology, Tangshan, China, [6] Beijing Sitairui Cancer Data Analysis Joint Laboratory, Beijing, China, [7] State Key Laboratory of Process Automation in Mining and Metallurgy, Beijing, China, [8] Beijing Key Laboratory of Process Automation in Mining and Metallurgy, Beijing, China

OPEN ACCESS

Edited by:
Shahid Mumtaz, Instituto de Telecomunicações, Portugal

Reviewed by:
Marcin Wozniak, Silesian University of Technology, Poland
Omar Abu Arqub, Al-Balqa Applied University, Jordan

***Correspondence:**
Aimin Yang
aimin@ncst.edu.cn

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

Received: *23 January 2022*
Accepted: *08 February 2022*
Published: *03 March 2022*

Citation:
Ma Z, Zhang M, Liu J, Yang A, Li H, Wang J, Hua D and Li M (2022) An Assisted Diagnosis Model for Cancer Patients Based on Federated Learning. Front. Oncol. 12:860532. doi: 10.3389/fonc.2022.860532

Since the 20th century, cancer has been a growing threat to human health. Cancer is a malignant tumor with high clinical morbidity and mortality, and there is a high risk of recurrence after surgery. At the same time, the diagnosis of whether the cancer is *in situ* recurrence is crucial for further treatment of cancer patients. According to statistics, about 90% of cancer-related deaths are due to metastasis of primary tumor cells. Therefore, the study of the location of cancer recurrence and its influencing factors is of great significance for the clinical diagnosis and treatment of cancer. In this paper, we propose an assisted diagnosis model for cancer patients based on federated learning. In terms of data, the influencing factors of cancer recurrence and the special needs of data samples required by federated learning were comprehensively considered. Six first-level impact indicators were determined, and the historical case data of cancer patients were further collected. Based on the federated learning framework combined with convolutional neural network, various physical examination indicators of patients were taken as input. The recurrence time and recurrence location of patients were used as output to construct an auxiliary diagnostic model, and linear regression, support vector regression, Bayesling regression, gradient ascending tree and multilayer perceptrons neural network algorithm were used as comparison algorithms. CNN's federated prediction model based on improved under the condition of the joint modeling and simulation on the five types of cancer data accuracy reached more than 90%, the accuracy is better than single modeling machine learning tree model and linear model and neural network, the results show that auxiliary diagnosis model based on the study of cancer patients in assisted the doctor in the diagnosis of patients, As well as effectively provide nutritional programs for patients and have application value in prolonging the life of patients, it has certain guiding significance in the field of medical cancer rehabilitation.

Keywords: cancer, machine learning, federated learning, cancer recurrence, diagnostic model

INTRODUCTION

Since the 20th century, the improvement of information storage capacity and the continuous improvement of information processing speed have promoted the rapid development of the data storage industry and big data information technology, and at the same time produced a huge thrust for the birth and development of emerging industries. At present, the amount of data output in the medical field is increasing exponentially. Through effective data resource storage and transmission management technology, combined with big data mining technology, the utilization efficiency and intelligence of data in the medical field have been improved (1, 2), giving medicine rapid development of the field has injected new impetus. This paper studies the influencing factors of postoperative recurrence of cancer patients, and proposes a federated learning model suitable for predicting the auxiliary diagnosis and prediction of cancer patients. The clinical data of patients is collected and combined with the prediction model to predict the location of cancer recurrence in recovered patients. Further Assisting doctors in diagnosis and improving the survival rate of cancer patient's has certain significance in the fields of cancer care, rehabilitation and clinical diagnosis.

With the continuous improvement of the material living standard of human society, the living environment and lifestyle of human beings have also changed correspondingly. The cancer problem that comes with it has become one of the most serious problems threatening human health. According to the International Agency for Research on Cancer According to the estimated data of "Global Cancer Incidence and Mortality in 2018" (GLOBOCAN2018) (3), there were approximately 18.1 million new cancer cases and 9.6 million cancer deaths worldwide in 2018. In 1971, the United States first proposed the concept of "tumor rehabilitation" (4), the main purpose of which is to help cancer patient's recover their mental, physical and physical functions under cancer conditions and limited treatment. The way of cancer rehabilitation mainly depends on the nature of the tumor and the stage of development of the tumor (5). Early detection and rehabilitation of cancer have greatly improved the survival rate and quality of life of patients. However, the factors affecting cancer patient's' recurrence after surgery are complex, so it is very challenging to predict the condition and trend of cancer patient's after surgery. In this regard, many scholars have done some work. Based on the evaluation data of cancer patients, some scholars have classified and predicted benign or malignant tumors, predicted postoperative recurrence time, and predicted the type of tumor. And trend research (6–8), the continuous development of machine learning and deep learning fields has also played a huge role in assisting cancer diagnosis and treatment (9, 10). However, there are two problems in the development of machine learning technology. On the one hand, data security is difficult to guarantee, and privacy protection issues are becoming more and more serious. On the other hand, because data sharing has become a new trend, and in order to prevent leakage of data among enterprises, data protection has been strengthened, and data in the era of big data has been reduced. Sharing, machine learning has encountered obstacles in data sharing training, resulting in the phenomenon of "data islands" (11). In the medical environment, the phenomenon of data islands also exists among hospitals. In order to break the phenomenon of "data islands", Google proposed the concept of federated learning (12) in 2016, which was originally used to solve Android mobile terminals. The problem of users updating the model locally, the design goal is to carry out between multiple parties or multiple computing nodes under the premise of ensuring information security during big data exchange, protecting terminal data and personal data privacy, and ensuring legal compliance and efficient machine learning. Among them, the machine learning algorithms that can be used in federated learning are not limited to neural networks, but also include important algorithms such as random forests. This model effectively solves the problem of privacy protection during data sharing between various enterprises. This model not only improves the security of data sharing between enterprises, on the other hand, because of the data sharing between enterprises, the accuracy of the training model also increases. At the same time, diagnosing whether the cancer is recurring *in situ* and predicting the time of recurrence are crucial for the patient's next rehabilitation treatment. According to statistics, about 90% of cancer-related deaths are caused by failure to prepare for cancer recurrence and cancer cell metastasis. Therefore, research on accurately predicting the location of cancer recurrence and resetting time under the premise of ensuring data security is for cancer Clinical diagnosis and treatment are of great significance.

AUXILIARY DIAGNOSIS MODEL CONSTRUCTION RELATED WORK

At present, cancer has always been a worldwide medical problem. With the gradual increase in the incidence of cancer, traditional cancer rehabilitation forecasts and cancer rehabilitation programs given by doctors through their own experience can no longer meet the needs of patients, and cancer is generally difficult to achieve a complete cure. The effect of cancer treatment is often limited to improving symptoms, with the goal of improving the quality of life of patients during the survival period and prolonging life span. Although cancer patients cannot be completely cured, more and more advanced technologies are applied in the medical field. The current 5-year survival rate of patients with advanced cancer has increased from 2%~5% decades ago to 16%~23% today. In the future, the accumulation of cancer patient data and the vigorous development and application of artificial intelligence will have more advantages than traditional medical models. In the long run, the application of artificial intelligence will surely drive the field of cancer rehabilitation diagnosis to high-end, personalized, precise, and intelligent. This research will adopt the convolutional neural network algorithm based on the federated learning framework to predict the recurrence time and location of cancer patients. On the one hand, federated learning has sufficient guarantee for the safety of cancer patient data. On the

other hand, cancer patient's data is predicted under the federated learning framework, which provides a guarantee for the safety of patient data among hospitals, greatly increases the amount of training data, and makes the training model more accurate in the end, which benefits multiple parties. Finally, the doctor outputs the results and patient information through the model. Analysis of the correlation degree, timely intervention in the rehabilitation process of patients, in order to improve the survival time of patients.

Federated Learning

In the context of the gradual maturity of machine learning and the realization of automatic identification and intelligent decision-making, in order to solve the problem of data privacy protection, federated learning (13–16) emerged as a potential solution.

Since the training data is still stored locally in the participants during the federated learning process, this mechanism can not only realize the sharing of the training data of each participant, but also ensure the protection of the privacy of each participant (17). The basic workflow of federated learning mainly includes:

1. Participants download the initialized global model from the cloud server, use the local data set to train the model, and generate the latest local model update (model parameters).
2. The cloud server collects various local update parameters and updates the global model through the model averaging algorithm. Because of the unique advantage of federated learning-a unified machine learning model can be trained from the local data of multiple participants under the premise of protecting data privacy.

Its main innovation is to provide a distributed machine learning framework with privacy protection features. Its working principle is as shown in **Figure 1**, and it can cooperate with thousands of participants in a distributed manner for a specific machine learning model. Iterative training, an iterative process of federated learning is as follows:

1. The client downloads the global model $\lambda_{t,\,k+1}$ from the server.
2. Client k trains local data to obtain local model $\lambda_{t,\,1}$
3. Clients of all parties upload local model updates to the central server.
4. The server performs a weighted aggregation operation after receiving the data from all parties to obtain the global model λ_t.

Among them, represents the local model update of the *t-th* round of communication of the *k-th* client, and represents the global model update of the *t-th* round of communication.

It can be seen from the introduction and flow chart that the federated learning technology has the following characteristics.

1. The original data participating in the federated learning is kept on the local client, and only the model update information is interacted with the central server, and there is no data transmission in plain text.

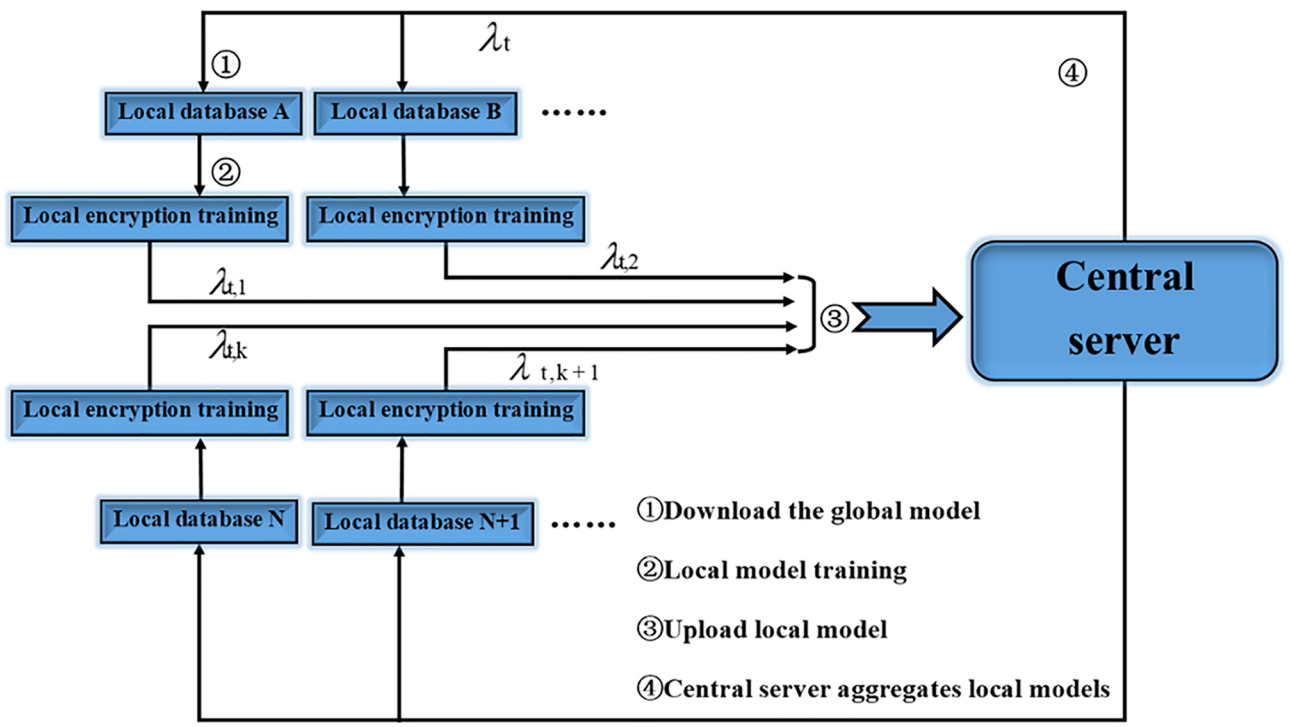


FIGURE 1 | Federated learning workflow.

2. The model jointly trained by the participants of federated learning will be shared by all parties who contribute training data.
3. The final model accuracy of federated learning is similar to that of centralized machine learning, and the accuracy is stronger.
4. The higher the quality of the training data of the federated learning participants, the higher the global model accuracy.

Federated learning can be divided into three categories: Horizontal Federated Learning, Vertical Federated Learning, Federated Transfer Learning. Horizontal federated learning is essentially the union of samples. The scope of application is where there is a large overlap of participant data features and a small overlap of user data. The data that can be used for joint modelling training is that part of the data where both parties have the same data characteristics but the users are not identical. For the part of the data, the horizontal federated learning application scenarios are more extensive. For example, between banks A and B in the same region, their businesses are similar (features similar), but users are different (different samples). Another example is the patient data of Hospital A and Hospital B for a particular case, which is also perfectly suitable for horizontal federal learning. There is data A in data B. Under the framework of the federated horizontal learning model, the server only conducts joint training for the common features of data A and data B and the parameters are returned to the participants. For this study, we have a total of three hospitals participating together. We select the experimental data strictly in combination with the characteristics of horizontal federated learning, and finally establish the model under the condition of ensuring that the data of each hospital is protected.

Localized Differential Privacy Protection Method

Differential privacy is a privacy definition first proposed by Cynthia Dwork in 2006 (18), which was developed in a specific scenario of statistical disclosure control. Differential privacy provides a kind of information theory security guarantee, so that the output result of the function is insensitive to any specific record in the data set.

Differential privacy can be divided into centralized differential privacy and localized differential privacy according to the different ways of data mobile phones. The two are different from the stages of differential data. Centralized differential privacy requires a trusted third party to collect data and perform data differential work in a unified manner. However, the current problem is that it is difficult to find a trusted third party in our lives. Therefore, in the context of federated learning, localized differential privacy can fit well with the encryption process required by the federated learning framework. The data is preprocessed using the idea of localized differential privacy, and then the federated learning framework is used for subsequent operations to fully improve the data. The safety of the user and the safety of the user.

Localized differential privacy can transfer the data privacy processing process to each participant in federated learning, and the participants will process and protect the data themselves, which will further improve the security of the data, which is defined as (19, 20): for any one Localized differential privacy function $f(x)$, its domain (domain) is $Dom(f)$, range (range) is $Ran(f)$, for any input x, $x' \in Dom(f)$, output $y \in Ran(f)$, we call function f provides (ξ) -localized differential privacy protection, y is the final output, currently only When it meets:

$$\Pr[f(x) = y] \leq e^{\xi} \Pr[f(x') = y] \quad (1)$$

In the above formula, ξ represents the privacy budget.

The concept of localized differential privacy is similar to the concept of federated learning. In fact, we have combined the idea of localized differential privacy in the realization of this research.

Convolutional Neural Network

CNN was proposed in 1998 by Yann LeCun of New York University (21). CNN is essentially a multilayer perceptron. The key to its success lies in the way it uses local connections and shared weights. On the one hand, it reduces the number of weights and makes the network easy to optimize, and on the other hand, it reduces overfitting risk (22). CNN is a kind of neural network. Its weight sharing network structure makes it more similar to biological neural network, which reduces the

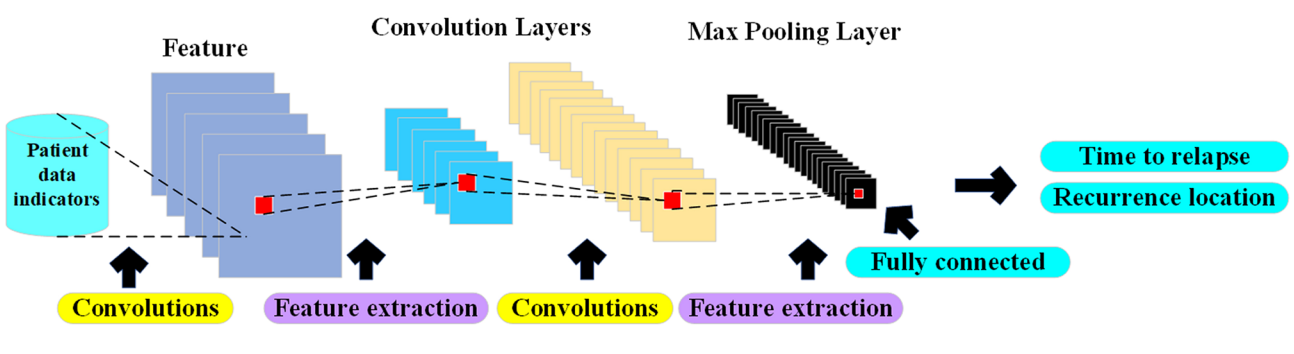


FIGURE 2 | Auxiliary diagnosis CNN structure diagram.

complexity of the network model and reduces the number of weights. The model structure of CNN is shown in **Figure 2**.

With the continuous increase in the incidence of cancer, the difficulty in detecting early cancer symptoms, and the various uncertainties in rehabilitation after cancer surgery, cancer has become the number one killer of human health today. In response to this difficulty and uncertainty, the CNN algorithm is widely used in cancer CT image detection and cancer postoperative recurrence prediction with its strong recognition ability and high prediction accuracy. The American artificial intelligence AI uses deep learning (CNN) to diagnose and treat cancer (23), and trains a deep convolutional neural network model to detect cancerous transformation of normal cells by letting artificial intelligence algorithms learn cancer CT images that far exceed the number of consultations of human doctors in a lifetime. The purpose is to achieve the purpose of early detection and early treatment.

In addition to showing high performance in recognizing cancer CT images, CNN also shows its powerful side in cancer prediction. As we all know, prostate cancer is the most common and the second most deadly cancer among American men. The classification of prostate cancer based on histological image Gleason classification is of great significance in patient risk assessment and treatment planning. In response to this problem, the regional convolutional neural network model was used to detect epithelial cells to predict the risk of cancer. After model training and experimental testing, the accuracy rate reached 99.8% (24).

In summary, CNN, as one of the deep learning algorithms, has mature theoretical foundations and experimental cases for cancer CT image detection and recognition or cancer incidence prediction, which provides theoretical guidance for the cancer rehabilitation medical recommendation system designed in this paper.

CNN is an artificial neural network with high recognition ability. In CNN, there are multiple neuron connections between each layer of the network. The convolution kernel is actually a user-defined size and weight matrix, which acts on the local perception domains in different regions of the same image, and extracts each local perception domain. And generate input values for the next layer of neurons. The convolutional layer convolves the input features, and the pooling layer reduces the size of the feature map through spatial invariance averaging or maximum operation. The activation function we use ReLU (22). The main advantage of CNNs is that they are easier to train and have fewer parameters than fully connected networks with the same number of hidden units. The feature map is shown in formula (2). The pooling layer performs secondary extraction of input features through specific pooling rules, and its feature map is shown in formula (3).

$$H_i = f(H_{I-1} \otimes \omega_i + b_i) \tag{2}$$

$$H_j = f(pooling(H_{I-1}) + b_j) \tag{3}$$

Among them, H_i is the feature map, $f(x)$ is a nonlinear activation function, "⊗" is the convolution operation of the convolution kernel and the feature map, ω is the weight vector b is the bias, pooling (x) is a pooling rule, for example Average pooling layer, maximum pooling layer and random pooling layer.

The structure of the convolutional neural network designed in this paper takes into account the sample data as 6 indicators. The size of the convolutional layer is 3×3×128, 3×3×256, 3×3×512, and the pooling layer is uniformly designed to have a size of 2×2.

(1) Convolutional layer: the *j-th* feature image of the first layer is expressed as:

$$X_j = g(\sum_{x_i \in M_j} x_i * k_{ij} + b_j) \tag{4}$$

Among them, the nonlinear activation function *g*. The set of feature maps connected between the l-1th layer and the *j-th* feature map of the lth layer is denoted as M_j, which means the set of input feature images. The offset is denoted as b_j. The convolution kernel connecting the *i-th* feature map in the l-1 layer and the *j-th* map in the l-*th* layer is denoted as k_{ij}

(2) Pooling layer: The pooling layer is denoted as the l-*th* layer, and the *j-th* feature map x_j of the lth layer is expressed as:

$$x_j = w_j pool(X_j) + b_j \tag{5}$$

Among them, the weight coefficient $x_j = w_j\ pool(X_j) + b_j$ is denoted as w_j, and the real number is taken in the general experiment. The bias is denoted as b_j, and the pooling function is denoted as *pool*(). There are maximum pooling, average pooling, random pooling, and LP pooling.

(3) Fully connected layer: the output vector x^l of the fully connected layer:

$$x^l = (\beta^l)^T v^{l-1} + b^l \tag{6}$$

Among them, the vector generated by the feature map of the pooling layer of the l-1 layer or the output vector of the feature map of the convolutional layer is denoted as v^{l-1}, the bias is denoted as b^l, and the weight coefficient matrix is denoted as β^l.

(As early as 2004, the Stanford University Medical and Mathematics Interdisciplinary Research of Cancer Patient Rehabilitation Intelligence Evaluation Model has been established, and it has shown its huge application prospects in clinical applications for hundreds of thousands of American cancer patients. In China, we In conjunction with the Stellite Cancer Data Analysis Laboratory, Hebei and Shanxi and other hospitals, it has long-term evaluated and tracked the rehabilitation process of thousands of cancer patients in China, analyzed their clinical data and rehabilitation data, and established Model of the system suitable for Chinese patients)

Federated Learning Model Based on Convolutional Neural Network

In order to optimize the recurrence time and the accuracy of the recurrence location of cancer patients in the rehabilitation stage, and to solve the insufficient amount of data in a single hospital (25, 26), this question paper proposes a cancer patient-assisted diagnosis model that combines federated learning and convolutional neural networks. Use federated learning to protect user data privacy and expand the amount of data, and

at the same time allow participants to collaborate to train a global model without sharing each other's private data. For each participant, the local data needs to be pre-processed, including digitization, and standardized to convert the original data into a standard data format, and then the local data is the first step to protect the local data with local differential privacy.

The iterative process completes the training of parameters locally for the convolutional neural network model deployed by the third party, and the parameters include the convolution kernels and offset terms of each layer. Post-encryption training on patient data using homomorphic encryption, followed by uploading of parameters. After receiving the model parameters uploaded by the client, the server will iterate the model according to the configuration of the central server, update the parameters of the current model, and persist it for the next round of training parameter upload and aggregation before returning it to the participants. The iterative process of our overall model follows the basic federated learning iterative process, in which we fuse a convolutional neural network adapted to cancer patient data samples and configure the model to form continuous iterations. According to the data supply characteristics of each participant, we adopt the rules of horizontal federated learning and unify the data standards.

In actual training, we consider that each participating hospital only exchanges encrypted correlation coefficients with the server. This experiment is based on the case where the data scale of each hospital is equal or the difference is not large. The training model algorithm is as follows: **Algorithm 1** Shown.

Algorithm 1 CNN-FL model based on local differential privacy

1 For Iteration t do
/* Service-Terminal: * /
2 $\omega_t = \frac{1}{Q}\sum_{q=1}^{Q}\omega_t^k$;
3 Send ω_t to each participant;
/* participants: * /
4 for Participant q do
5 Do Localized differential privacy
6 $\omega_t^q = \omega_t$;
7 For Local epoch e do
8 $\omega_t^q = \omega_t^q - \eta\frac{\partial}{\partial X^q}Z_{\wp}$
9 End
10 End
11 End

Among them, Q represents the participant of Q; ω_t represents the global model parameter during iteration t; ω_t^k represents the model parameter of the q th participant at iteration t; η represents the learning rate; X^k represents the training of the q-*th* participant data set. It should be noted that differential privacy protection is added to the user side of the algorithm, and

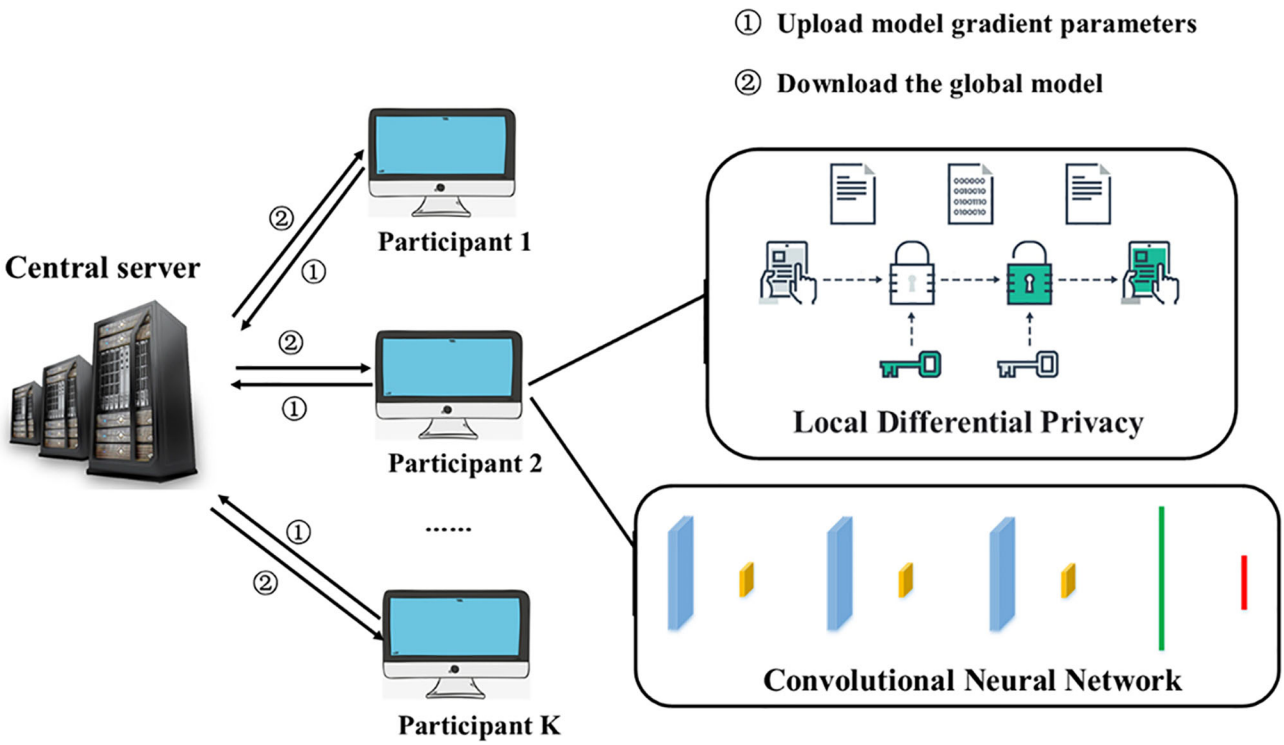


FIGURE 3 | CNN-FL model structure.

after a partial differential privacy model is initially formed, it constitutes a global model for user homogenization upload parameter training.

As shown in Algorithm 1 above, each participant needs to use the local data set to train the CNN model. The model is shown in **Figure 3** below. In each iteration, each model participant first uploads the current model correlation coefficient to the server. The global model is updated by averaging the latest correlation coefficients of each participant. In the next iteration, each participant downloads the latest global model parameters and uses local data to train the CNN model. Iterate continuously until the overall model is optimal.

The current CNN-FL model has 100 participants. After 50 iterations, the global model parameter is ω_{50}. At the 51st iteration, each participant's initial local model parameter $\omega_{51}^{k} = \omega_{50}$ (the value interval of k is [1, 100]), After 51 iterations, the global model is updated to $\omega_{51} = \frac{1}{100}\sum_{k=1}^{100}\omega_{51}^{k}$.

CONSTRUCTION OF REHABILITATION DATA SAMPLE SET BASED ON FEDERATED LEARNING

According to the characteristics of federated learning and long-term medical consensus (27, 28), ASC carcinogenic factor research report and TIES.IO cancer assessment data, 12 factors that affect cancer recurrence are comprehensively selected: gender, age, basic score, tumor score, immune score, basic Nutrition score, nutritional comparison score, safe intake score, total nutrition score, microenvironment score, psychological score, aerobic activity score, collect data from cancer patients and score, use statistical correlation coefficient Pearson correlation coefficient, Spearman correlation coefficient to compare the sample Enter the indicators for correlation research, and finally determine 6 influencing factors: tumor score, immune score, basic nutrition score, psychological score, microenvironment score, aerobic exercise and advanced homework. The Pearson correlation coefficient is shown in **Table 1**. The related items of each index score and their weights are shown in **Table 2**. The weights are given based on the experience of doctors and experts (29). The data set in this article was collected from Shanxi Provincial People's Hospital and Hebei Tumor Hospital, etc., a collaborative experiment in Beijing, China The office is responsible for cancer data evaluation and data processing, as well as liaison with various hospitals.

It can be seen from **Table 2** that each index has a certain correlation, and we can see that the tumor index and the immune index are positively correlated, indicating that the stronger the immune index, the weaker the tumor index. Moreover, there is a negative correlation between psychological indicators and tumor indicators. The more ideal these indicators, the longer the patient will have to relapse.

EXPERIMENTAL SIMULATION AND RESULT ANALYSIS

Cancer Aided Diagnosis Model

We build a cancer-assisted diagnosis model. Based on the evaluation criteria of each indication described above, we trained and tested the machine learning-assisted diagnosis and treatment model on 500 sets of data (5 groups of 500 different cancer patient's), and first established the input and output vectors, Among which sample input and output: the six major indicators of immune indication, tumor indication, microenvironment indication, psychological indication, nutrition indication, aerobic exercise and advanced homework as input, the predicted recurrence time and recurrence position As an output, the experimental results are shown in **Figure 4** below. We combined medical knowledge and intelligent diagnosis and treatment models to set the prediction error range of cancer recurrence time to ±6 months. During the experiment, we used the linear model of machine learning, the tree model, and the neural network of the multi-layer perceptron (MLP) in deep learning is used to test the cancer-assisted diagnosis model.

Based on multiple iterative experiments and multiple experiments, combined with the absolute error of the cancer recurrence time, the accuracy of the prediction of the recurrence

TABLE 1 | Cancer Patient Index Evaluation Criteria.

Finger syndrome	Related terms and Weights
Immune score	CD3+CD4+CD8+/CD45+ (4); CD3+CD4+/CD45+ (8); CD4+/CD8+ (10); CD3+CD16+CD56+/CD45+ (6); CD3-CD56+ (5); CD4+CD25+ (1); Exercise ECG (X ± SD) (2); Sports Leather (X ± SD) (2).
Tumor score	Size (10); Placeholder (10); Violate the relationship (10); Angiogenesis (10); Pathological typing (3); CTC value (9); Differentiation (10); Mutation target (1).
Basic nutrition score	Total nutrition (6); Balanced nutrition (3); Nutrition safety assessment (5); Cancer cell proliferation (10); Immune cell proliferation (10); Angiogenesis (8); Amino acid evaluation (5); Proteomics evaluation (10).
Psychological score	Life event scale (1); Cornell Medical Index (2); Self-rating anxiety scale (5); Self-rating depression scale (5); Baker Anxiety Scale (5); Baker Depression Questionnaire (5); Pittsburgh sleep Quality index (4); Texas Social Behavior Questionnaire (3); Family function assessment (1); Exercise ECG (X ± SD) (2); Sports Leather (X ± SD) (2).
Microenvironment score	O2 (3); PH value (4); Interstitial pressure (2); Inflammatory response (7); Vascular permeability (6); CTC value (9); Proteomic analysis (8).
Exercise and advanced work	Aerobic exercise (4); Advanced social work (3); Texas Social Behavior Questionnaire (3).

TABLE 2 | Correlation analysis table of each index.

Pearson correlation	Pearson-cor/Sig.	Imm	Tum	Mic	Heart	Nut	Aer
Imm	Pearson-cor	1	0.30*	0.04	0.17	0.21	-0.07
	Sig.		0.03	0.79	0.25	0.14	0.64
Tum	Pearson-cor	0.30*	1	0.09	-0.38**	0.10	0.03
	Sig.	0.03		0.54	0.01	0.48	0.86
Mic	Pearson-cor	0.04	0.09	1	0.21	-0.20	-0.003
	Sig.	0.79	0.54		0.14	0.18	0.99
Heart	Pearson-cor	0.17	-.038**	0.21	1	-0.15	-0.02
	Sig.	0.25	0.01	0.14		0.32	0.92
Nut	Pearson-cor	0.21	0.10	-0.20	-0.15	1	-0.19
	Sig.	0.14	0.48	0.18	0.32		0.19
Aer	Pearson-cor	-0.07	0.03	-0.003	-0.02	-0.19	1
	Sig.	0.64	0.86	0.99	0.92	0.19	

*At the 0.05 level (Sig.), the correlation is significant, **At the 0.01 level (Sig.), the correlation is significant.

time of the five types of cancer patient's is between 65%-85%. This result proves that the model has practical application value. The doctor can complete the diagnosis and treatment of the patient based on the results and refer to the various indicators of the patient, and this result is only completed in a unilateral modeling situation with a limited amount of data, and then we can determine whether the recurrence of the cancer patient has metastasized to another location, where 1.0 is The original position, 2.0 is that the cancer cells have metastasized. The test was performed using the constructed auxiliary diagnosis model. The test results of the neural network using the multi-layer perceptron (MLP) are shown in **Figure 5**.

It can be clearly seen from the figure above that the MLP network has been trained and tested on 100 sets of samples. The final performance of the MLP network is 90%, and the algorithm that predicts the location of cancer recurrence (that is, whether the cancer cell has metastasized to other parts of the body) can reach 90%. It was unstable. We subsequently simulated the patients, and showed excellent application prospects under the condition of unilateral machine learning modeling and insufficient data.

From the **Table 3**, we know whether the cancer cells of the patient in the sample have metastasized and whether they have recurred in situ. By predicting simulation and simulation, only one-way machine learning modeling can achieve very impressive results. Then, we proposed a method A convolutional neural network-assisted diagnosis model based on federated learning. On the one hand, this model protects the data privacy of patients. On the other hand, through the joint modeling of multiple hospitals, they can share each other's data but protect each other's privacy.

Convolutional Neural Network Aided Diagnosis Model Based on Federated Learning

We analyze the disadvantages of data processing of cancer patient's based on machine learning. Among them, machine learning adopts unilateral modeling, and the data is unprotected. The amount of data in a single hospital is not sufficient, but it still achieves good results, but diagnosis and treatment based on machine learning The model is difficult to truly enter the application level and faces many data security issues. Therefore, we established a convolutional neural network-assisted diagnosis model based on federated learning, built the FL-CNN model based on the federated learning framework, and used privacy protection methods. The model parameter transmission updates the model, and after several rounds of parameter updates, we analyze the cancer recurrence time and the accuracy rate of the recurrence location in five types of cancer patients. The data volume used under the federated learning framework (after differential privacy) is shown in **Table 4** below, based on the federation The experimental results of the learned convolutional neural network cancer recurrence time simulation model are shown in **Figure 6** below.

From **Figure 6** above and **Figure 4**, it can be clearly seen that the parameters updated after multiple iterations of each participant through the federated learning framework based on the convolutional neural network-based auxiliary diagnosis model have obvious accuracy under 500 simulated simulation samples. Under the condition of an absolute error of ±6 months, although a certain proportion of noise data has been added to the participants to make each participant achieve homogeneity, the final experimental results indicate that the model simulates recurrence in various cancer patient's The time accuracy can reach more than 90%. Through the comparison of the two methods, the federated learning enables the model to be trained locally on the basis of the patient data privacy and security, and the updated parameters are returned to update the overall model, which expands The patient data sample further improves the accuracy of the model. The intelligent diagnosis model can assist doctors in diagnosing cancer patients to a certain extent. It has application prospects and is of great significance for prolonging the lives of cancer patients. It also provides a way for doctors to diagnose patients. Effective reference and **Table 5** is a comparison of the above figure with the recurrence time and the pros and cons of the model. After that, we conducted corresponding experiments on whether the cancer cells metastasized when the patient's cancer recurred, and used the diagnosis and treatment model to assist doctors in the

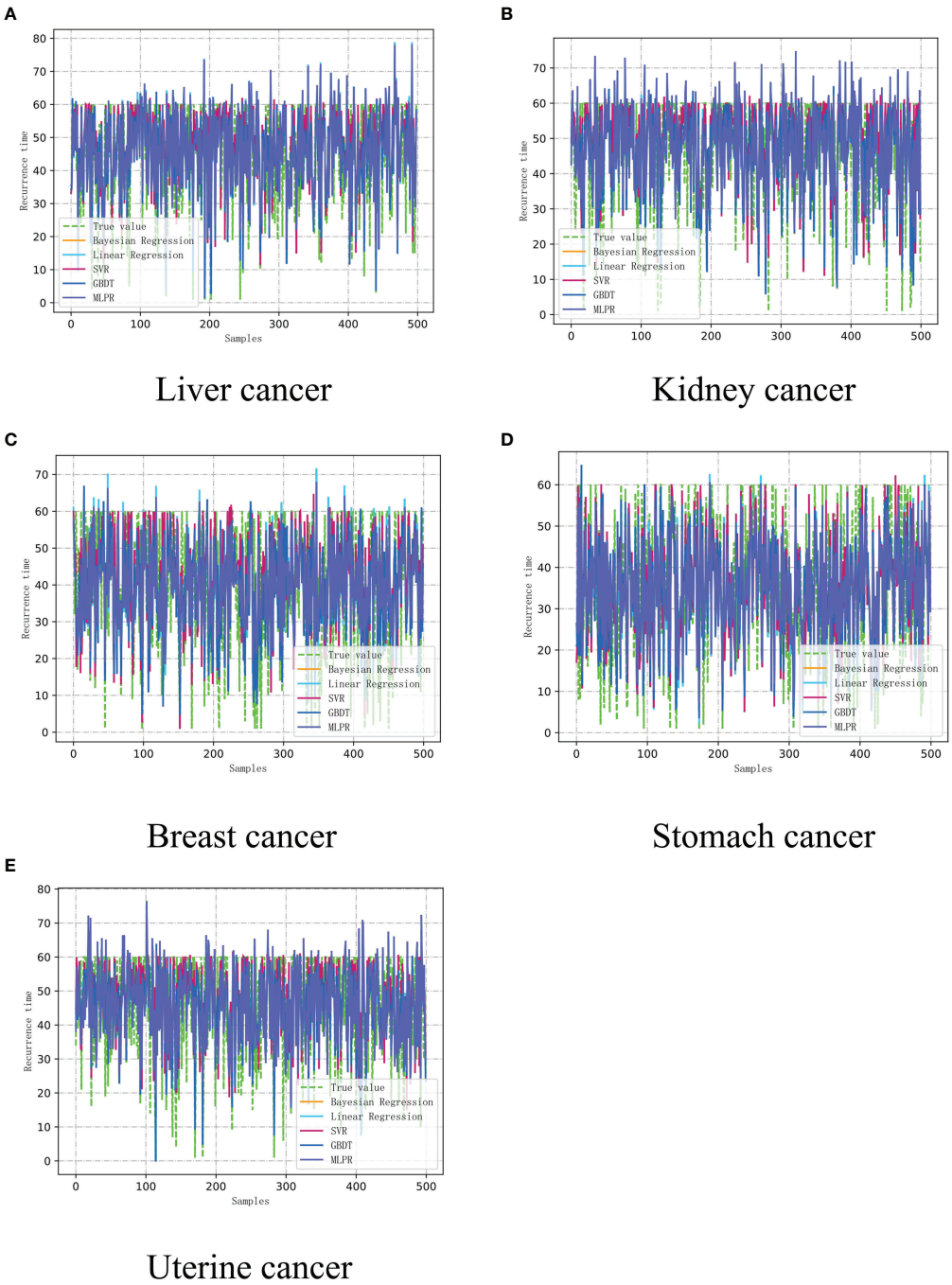


FIGURE 4 | The recurrence time model of cancer assisted diagnosis based on machine learning. **(A)** Liver cancer. **(B)** Kidney Cancer. **(C)** Breast cancer. **(D)** Stomach cancer. **(E)** Uterine cancer.

diagnosis and rehabilitation of cancer patients. With the iteration and parameter return of the model under the federated learning framework, the recurrence position gradually stabilized at Around 90%, we finally learned through the federated learning model that the recurrence time is also in a different state with the changes of various indicators. This shows that regulating different indicators can help cancer patients to a certain extent. The interactive interface is designed to facilitate the doctor's understanding, see **Figure 7** below.

The auxiliary diagnosis system based on the federated learning framework has greatly increased the data sample size, increased the number of data iteration rounds, and guaranteed

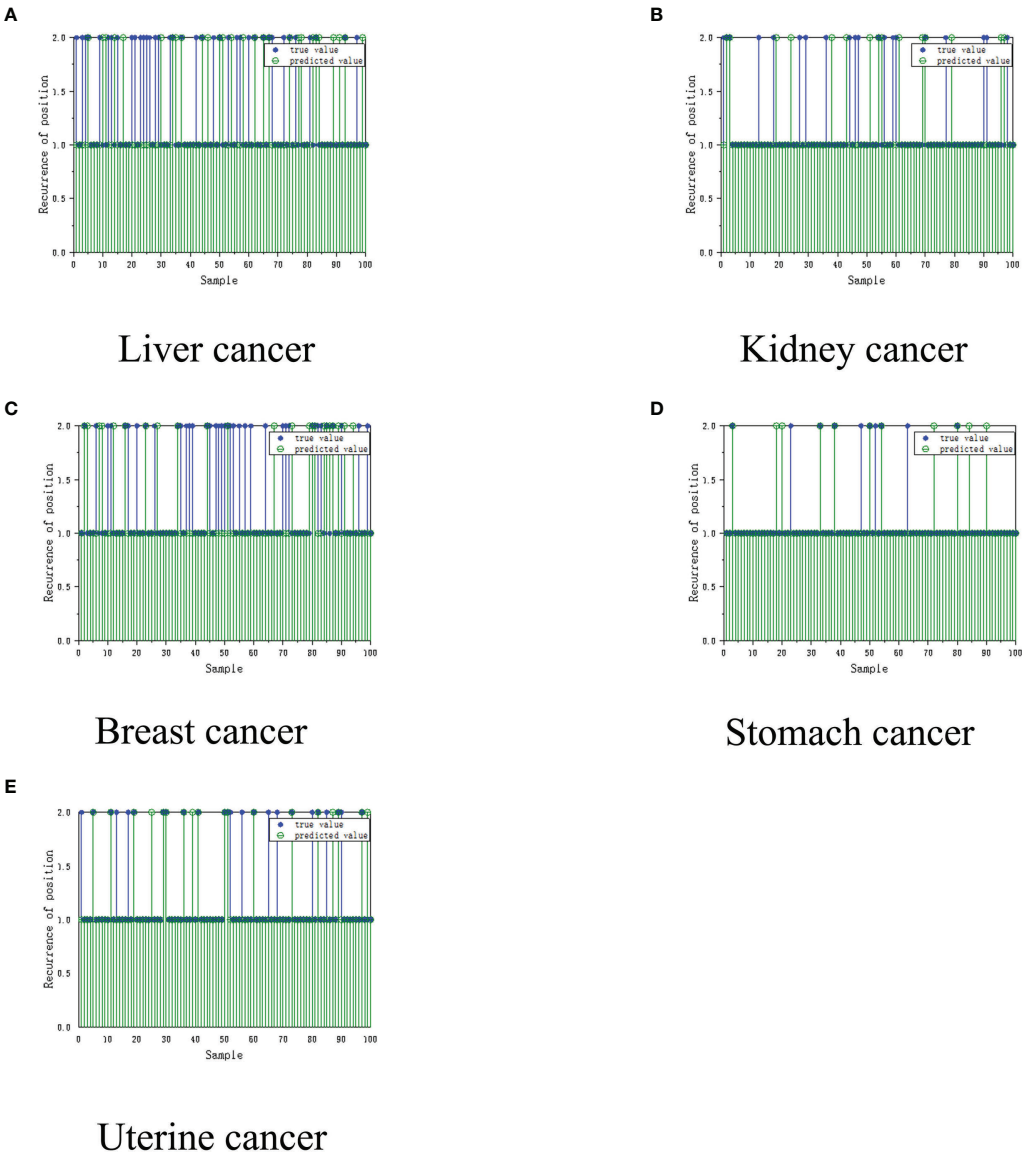


FIGURE 5 | Recurrence location model for cancer assisted diagnosis based on machine learning. **(A)** Liver cancer. **(B)** Kidney Cancer. **(C)** Breast cancer. **(D)** Stomach cancer. **(E)** Uterine cancer.

data security, greatly improving the accuracy of the model, and we have given an interactive design diagram, In order to facilitate the application of this model, through the accurate model and analysis of the correlation between various indications and cancer recurrence time, doctors can finally use the model and medical knowledge to intervene in the patient's rehabilitation process, affect the patient's rehabilitation indications, and improve patient survival rate.

TABLE 3 | Unilateral modeling and simulation of recurrence location.

MLP neural network	Liver cancer	Kidney Cancer	Breast cancer	Stomach cancer	Uterine cancer
In-situ (simulation)	60%	77%	61%	91%	86%
Transfer (simulation)	40%	23%	39%	9%	14%
In situ (actual)	62%	80%	60%	89%	76%
Transfer (actual)	38%	20%	40%	11%	24%

TABLE 4 | Introduction to the data set of participants.

Overall model data information (after localized differential privacy)	Hebei A Hospital, China	Shanxi B Hospital, China	Beijing C Hospital, China
Liver cancer	800-900 (1000)	800-900 (1000)	800-900 (1000)
Kidney Cancer	300-350 (400)	300-350 (400)	300-350 (400)
Breast cancer	300-350 (400)	300-350 (400)	300-350 (400)
Stomach cancer	175-215 (250)	175-215 (250)	175-215 (250)
Uterine cancer	200-250 (300)	200-250 (300)	200-250 (300)

CONCLUSION

In the context of the continuous improvement of international privacy protection laws and regulations, data security gradually being valued by the public, and the prevalence of "data islands", this article combines multiple hospitals with cancer patient data, and uses federated learning, neural networks, and localized differential privacy. Based on the homogenization of the center, a set of federated learning auxiliary diagnosis models for cancer patients was constructed, and the unilateral modeling machine learning assisted diagnosis models were compared. The accuracy and safety of the models have been greatly improved. The

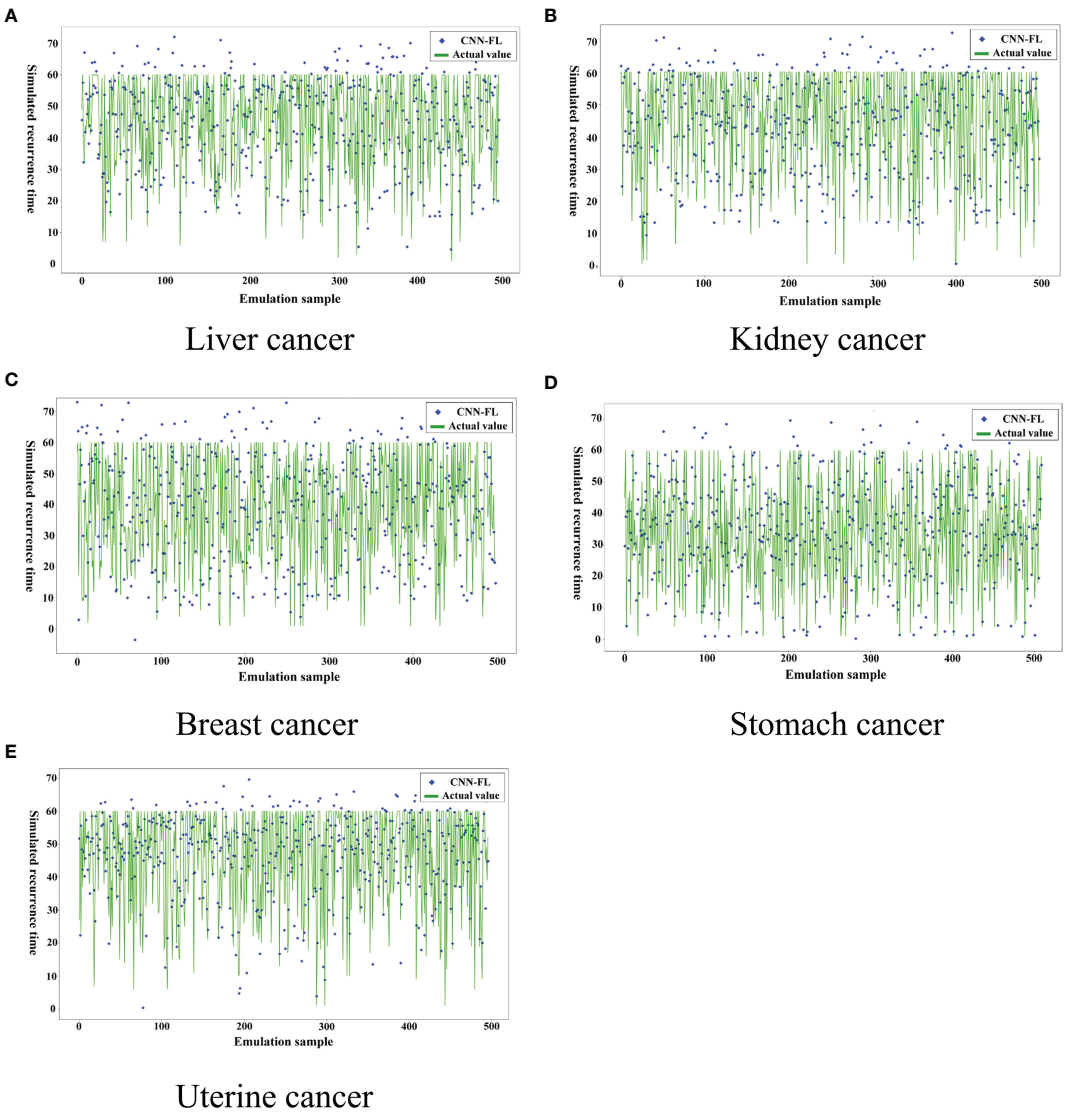


FIGURE 6 | Recurrence time model of cancer assisted diagnosis based on federated learning.

TABLE 5 | Model comparison.

Comparison of advantages and disadvantages	The amount of data	Safety	Accuracy (this experiment)	Participants
Unilateral modeling intelligent diagnosis model	Restricted	Low	65%-85%	Unilateral
A model for assisted diagnosis of cancer patients based on federated learning	Unrestricted	high	90%>	Multi-party joint

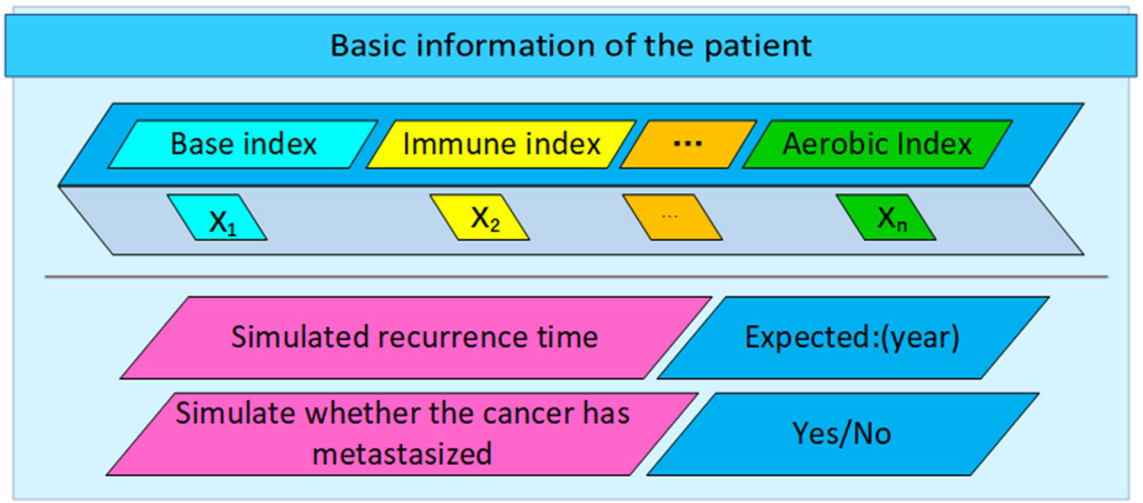


FIGURE 7 | Interactive design of auxiliary diagnosis and treatment system.

federated learning model effectively expand the training data of cancer patients, and protect the privacy and security of cancer patient's' data. This study only cooperated with three hospitals and one biological laboratory. Although the amount of data has increased significantly, there are still shortcomings for fatal cancers. It is expected that in the future, the number of participants will gradually increase. As the model is constantly updated, the cancer intelligent diagnosis and treatment system will eventually play its value.

Federated learning is one of the methods that can solve the current data security sharing problem. Compared with artificial intelligence methods that are widely used in all walks of life, such as unilateral machine learning and fuzzy systems, federated learning shows great advantages such as improving data privacy protection and expanding data volume. We have achieved gratifying results by applying it and applying it to the rehabilitation of cancer patients. It is expected to come, with continuous exploration and innovation. The phenomenon of "data islands" between various industries and enterprises will be broken, and data from all parties will be shared more reasonably and safely, allowing artificial intelligence to be applied to all corners of us, and we will continue to work on assisted diagnosis solutions for cancer patient's The research, combined with more advanced mathematical models and machine learning models, and safer federated learning privacy protection methods, is to find the best solution for cancer patients on the premise of protecting patient data security and expanding cancer patient data.

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

ZM contributed to the conception of the study, experimental part and writing the main body of the paper. MZ participated in the important experimental part of the paper. AY contributed to the construction of the overall thesis framework and the revision of some thesis content, as well as the follow-up communication of the thesis. DH is in charge of collation of experimental data and liaison with major hospitals. The rest of the authors are responsible for the alignment of data samples, construction of experimental models and other issues, and can be counted as authors with equal contributions. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by: 1. Key Science and Technology Project of Hebei Provincial Department of Education (North China University of Science and Technology, Project Number: JYG2020001); 2. Key Basic Research Project Fund of Hebei

Provincial Department of Science and Technology (North China University of Science and Technology, Project Number: 20270902D); 3. Project funded by the Natural Science Foundation of Hebei Province. (North China University of Science and Technology, Project Number: E2021209024); 4. State Key Laboratory of Process Automation in Mining & Metallurgy and Beijing Key Laboratory of Process Automation in Mining & Metallurgy, the fund project number is: BGRIMM-KZSKL-2018-10.

ACKNOWLEDGMENTS

The authors thank all the participants in this study. Thanks to all authors who contributed to this paper.

REFERENCES

1. Wu JH, Wei W, Zhang L, Wang J, Robertas D, Li J, et al. Risk Assessment of Hypertension in Steel Workers Based on LVQ and Fisher-SVM Deep Excavation. *IEEE Access* (2019) 7:23109–19. doi: 10.1109/ACCESS.2019.2899625
2. Orujov F, Maskeliūnas R, Damaševičius R, Wei W. Fuzzy Based Image Edge Detection Algorithm for Blood Vessel Detection in Retinal Images. *Appl Soft Comput* (2020) 94:106452. doi: 10.1016/j.asoc.2020.106452
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492
4. Fu JB, Molinares DM, Morishita S, Silver JK, Bruera E. Retrospective Analysis of Acute Rehabilitation Outcomes of Cancer in Patients With Leptomeningeal Disease. *Pm&r* (2020) 12(3):263–70. doi: 10.1002/pmrj.12207
5. Cenik F, Mhr B, Palma S, Keilana M, Crevenna R. Role of Physical Medicine for Cancer Rehabilitation and Return to Work Under the Premise of the "Wiedereingliederungsteilzeitgesetz". *Wiener klinische Wochenschrift* (2019) 131.19:455–61. doi: 10.1007/s00508-019-1504-7
6. Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine Learning Models in Breast Cancer Survival Prediction. *Technol Health Care* (2016) 24 (1):31–42. doi: 10.3233/THC-151071
7. Asri H, Mousannif H, Al Moatassime H, Noel T. Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. *Proc Comput Sci* (2016) 83:1064–9. doi: 10.1016/j.procs.2016.04.224
8. Agarap AFM. On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Datasetc]. In: *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing* Vietnam: ACM (2018). p. 5–9. doi: 10.1145/3184066.3184080
9. Hornbrook MC, Goshen R, Choman E, O'Keeffe-Rosetti M, Rust KC. Early Colorectal Cancer Detected by Machine Learning Model Using Gender, Age, and Complete Blood Count Data. *Digestive Dis Sci* (2017) 62(10):2719–27. doi: 10.1007/s10620-017-4722-8
10. Yang A, Han Y, Liu CS, Wu JH, Hua DB. D-TSVR Recurrence Prediction Driven by Medical Big Data in Cancer. *IEEE Trans Ind Inf* (2020) 17(5):3508–17. doi: 10.1109/TII.2020.3011675
11. Ye M, Wang Y. Research on the Legal System of Breaking the Data Island in the Era of Artificial Intelligence. *J Dalian Univ Technol (SOCIAL Sci EDITION)* (2019) 40(05):69–77. doi: 10.19525/j.issn1008-407x.2019.05.009
12. Mcmahan HB, Moore E, Ramage D, Hampson H, Arcas B. Communication-Efficient Learning of Deep Networks From Decentralized Data C]//Aarti Singh, Jerry Zhu. In: *Artificial Intelligence and Statistics*. Fort Lauderdale: Cornell University (pre-published) (2017). p. 1273–82.
13. Shokri R, Shmatikov V. *Privacy-Preserving Deep Learning C]//Proceedings of the 22nd ACMSIGSAC Conference on Computer and Communications Security*. Denver, CO, USA: ACM (2015) p. 1310–21.
14. Yoo JH, Jeong H, Lee J, Chung TM. *Federated Learning: Issues in Medical Application*. Cornell University (pre-published) (2021). doi: 10.1007/978-3-030-91387-8_1
15. Konen J, Mcmahan HB, Yu FX, Richtárik P, Bacon D. Federated Learning: Strategies for Improving Communication Efficiency. *arXiv* (2016) 16:1–10.
16. Nishio T, Yonetani R. *Client Selection for Federated Learning With Heterogeneous Resources in Mobile Edgec]//Proceedings of ICC*. Shanghai, China.
17. Li T, Sahu AK, Talwalker A, Smith V. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Process Mag* (2020) 37(3):50–60. doi: 10.1109/MSP.2020.2975749
18. Cynthia D, Frank M, Kobbi N, Adam S. Calibrating Noise to Sensitivity in Private Data Analysis. In: *Theory of Crytptography Conference*. Springer (2006). p. 265–84. doi: 10.1007/11681878_14
19. Li N, Ye Q. Mobile Data Collection and Analysis With Local Differential PrivacyC]. In: *2019 20th IEEE International Conference on Mobile Data Management (MDM)*. IEEE (2019). doi: 10.1109/MDM.2019.00-80
20. Ye Q, Meng X, Zhu M, Zheng H. Overview of Localized Differential Privacy Research. *J Software* (2018) 29(7):25. doi: 10.13328/j.cnki.jos.005364
21. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based Learning Applied to Document Recognition. *Proc IEEE* (1998) 86(11):2278–324. doi: 10.1109/5.726791
22. Aubert B, Vazquez C, Cresson T, Parent S, De Guise JA. Toward Automated 3d Spine Reconstruction From Biplanar Radiographs Using CNN for Statistical Spine Model Fitting. *IEEE Trans Med Imaging* (2019) 38 (12):2796–806. doi: 10.1109/TMI.2019.2914400
23. Yao M, Sohul M, Marojevic V, Reed JH. Artificial Intelligence Defined 5g Radio Access Networks. *IEEE Commun Mag* (2019) 57(3):14–20. doi: 10.1109/MCOM.2019.1800629
24. Li W, Li J, Sarma KV, Ho KC, Shen S, Knudsen BS, et al. Path R-CNN for Prostate Cancer Diagnosis and Gleason Grading of Histological Images. *IEEE Trans Med Imaging* (2018) 38(4):945–54. doi: 10.1109/TMI.2018.2875868
25. Liu X, Chen S, Song L, Woniak M, Liu S. Self-Attention Negative Feedback Network for Real-Time Image Super-Resolution. *J King Saud Univ - Comput Inf Sci* (2021) 4). doi: 10.1016/j.jksuci.2021.07.014
26. Woniak M, Sika J, Wieczorek M. Deep Neural Network Correlation Learning Mechanism for CT Brain Tumor Detection. *Neural Comput Appl* (2021) 6). doi: 10.1007/s00521-021-05841-x
27. Anno. New Biostratigraphic Data on the S.Cassiano Formation Around Sella Platform (Dolomites, Italy). *Allergie Et Immunol* (1994) 26(10):388–9. doi: 10.12785/amis/080617
28. Arqub OA, Abo-Hammour Z, Momani S, Shawagfeh N. Solving Singular Two-Point Boundary Value Problems Using Continuous Genetic Algorithm. *Abstract Appl Anal* (2014) 2012(5):1–25. doi: 10.1155/2012/205391
29. Rana S. Bio-Medical Data Processing With Machine Intelligence. (2021). doi: 10.13140/RG.2.2.22395.95524

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 | Frontiers in Oncology

METHODS
published: 27 May 2022
doi: 10.3389/fonc.2022.884523

Respiratory Prediction Based on Multi-Scale Temporal Convolutional Network for Tracking Thoracic Tumor Movement

Lijuan Shi[1,2], *Shuai Han*[1,2], *Jian Zhao*[2,3]*, *Zhejun Kuang*[2,3], *Weipeng Jing*[4], *Yuqing Cui*[1,2] *and Zhanpeng Zhu*[5]

[1] College of Electronic Information Engineering, Changchun University, Changchun, China, [2] Jilin Provincial Key Laboratory of Human Health Status Identification and Function Enhancement, Changchun, China, [3] College of Computer Science and Technology, Changchun University, Changchun, China, [4] College of Information and Computer Engineering, Northeast Forestry University, Harbin, China, [5] Department of Neurosurgery, The First Hospital of Jilin University, Changchun, China

Radiotherapy is one of the important treatments for malignant tumors. The precision of radiotherapy is affected by the respiratory motion of human body, so real-time motion tracking for thoracoabdominal tumors is of great significance to improve the efficacy of radiotherapy. This paper aims to establish a highly precise and efficient prediction model, thus proposing to apply a depth prediction model composed of multi-scale enhanced convolution neural network and temporal convolutional network based on empirical mode decomposition (EMD) in respiratory prediction with different delay times. First, to enhance the precision, the unstable original sequence is decomposed into several intrinsic mode functions (IMFs) by EMD, and then, a depth prediction model of parallel enhanced convolution structure and temporal convolutional network with the characteristics specific to IMFs is built, and finally training on the respiratory motion dataset of 103 patients with malignant tumors is conducted. The prediction precision and time efficiency of the model are compared at different levels with those of the other three depth prediction models so as to evaluate the performance of the model. The result shows that the respiratory motion prediction model determined in this paper has superior prediction performance under different lengths of input data and delay time, and, furthermore, the network update time is shortened by about 60%. The method proposed in this paper will greatly improve the precision of radiotherapy and shorten the radiotherapy time, which is of great application value.

Keywords: radiotherapy, respiratory motion prediction, deep learning network, empirical mode decomposition, temporal convolutional network

OPEN ACCESS

Edited by:
Shahid Mumtaz,
Instituto de Telecomunicações, Portugal

Reviewed by:
M. M. Manjurul Islam,
American International University-Bangladesh, Bangladesh
Thippa Reddy Gadekallu,
VIT University, India

***Correspondence:**
Jian Zhao
zhaojian@ccu.edu.cn

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

Received: *26 February 2022*
Accepted: *19 April 2022*
Published: *27 May 2022*

Citation:
Shi L, Han S, Zhao J, Kuang Z, Jing W, Cui Y and Zhu Z (2022) Respiratory Prediction Based on Multi-Scale Temporal Convolutional Network for Tracking Thoracic Tumor Movement. Front. Oncol. 12:884523. doi: 10.3389/fonc.2022.884523

1 INTRODUCTION

When a patient with cancer undergoes radiation therapy, the fluctuating movement of chest and abdomen caused by human respiratory motion makes the tumor unable to rest statically in the planning target volume (PTV), which causes it impossible to ensure the coverage of tumor by simply increasing the PTV area. Meanwhile, it is very likely for the organs at risk (OARs) around the tumor

to be destroyed during radiotherapy, thus causing secondary injury to the patients (1). Some studies have shown that, during breathing, some muscles (such as the diaphragm) move 20–130 mm, the lungs move an average of 8–10 mm, and the liver moves an average of 1–19 mm (2). Therefore, it is of great significance to reduce the adverse effects of human respiratory movement in the process of cancer treatment.

To address the problem of respiration-induced tumor displacement, many clinical initiatives have been proposed, including breath-holding techniques (3), passive compression techniques (4), respiratory gating techniques (5), and real-time tracking techniques (6). Breath-holding technique and passive compression technique both reduce the impact by actively controlling human respiration by itself or external equipment, which is very convenient, but the mandatory control makes the patient's tolerance poor and is not suitable for patients with pulmonary insufficiency. Respiratory gating technology tracks the location of the tumor by monitoring the patient's breathing and adjusting the radiation instrument to match a specific breathing cycle. Real-time tracking technology is currently one of the best methods to track tumors and improve treatment effects. It continuously adjusts the irradiation target area to track tumors in real time through *in vitro* marker signals (respiration laws).

Vedam et al. (7), Ozhasoglu and Murphy (8), and Fayad et al. (9) verified the correlation between respiration and tumor movement to varying degrees. CyberKnife, Exactrac, and Vero system are respiratory motion tracking systems applied in clinical practice. In the actual treatment, the machine system establishes the motion relationship between marker signals and tumor through the prediction model, so as to adjust the radiotherapy target position. A certain time delay is required during the adjustment process, which demands the establishment of prediction delay system through the external respiratory signal. The accuracy of delay prediction directly determines the target position in radiotherapy. The CyberKnife system has a system delay time of about 115 ms from data acquisition, calculation of tumor location, to adjustment of the radiation beam. The delay of Vero system is about 50 ms and that of Varian MLC system is about 420 ms (10, 11). To compensate for these delays, some prediction algorithm is used to calculate the future position of the target.

Conventional time series prediction models have been applied in the field of respiratory prediction, such as extended Kalman filter algorithm based on Kalman filter (12) combining with support vector machine (13), wavelet-based multi-scale regression (14), recursive least squares algorithm (15), and an autoregressive integrated moving average (ARIMA) model (16). With the development of deep learning, it has brought new possibilities to respiratory motion prediction. Deep learning can effectively mine time series information and semantic information, independently extract a large number of data features, and improve the prediction accuracy. To compensate for the system delay and improve the accuracy of respiratory motion prediction, this paper proposes a multi-scale enhanced time series convolution respiratory motion prediction model based on deep learning network. The main contributions are as follows:

(1) A multi-scale enhanced convolution and temporal convolution network (TCN) based on squeeze-and-excitation is proposed to establish a deep convolution neural network model for respiratory motion prediction.
(2) Aiming at the simplification of respiratory signal features, EMD algorithm is used to decompose the original complex sequence into several intrinsic mode functions (IMFs) with different time scales so as to increase the network fitting ability and improve the prediction precision.
(3) The underlying features of different receptive fields are extracted by using a multi-scale convolution kernel, and attention mechanisms are added to the feature space.
(4) The recurrent neural network (RNN) model is replaced by the TCN, which has higher precision and time efficiency than birectional long short-term memory (BiLSTM).

2 RELATED WORK

Deep learning is based on artificial neural network (ANN), which has stronger adaptability in the case of irregular breathing model and model. Some studies have shown (17, 18) that the ANN structure has certain advantages in the prediction of respiratory motion, especially when the respiratory signal is unstable and non-linear.

Convolutional neural network (CNN) can deal with data similar to grid structure through convolution operation and perform exceedingly well in many fields such as time series and image data; RNN has some advantages when learning the non-linear characteristics of sequences. LSTM is one of the classical algorithms of RNN series because of its introduction of the gate mechanism to make the network have a certain memory, so that the network can capture the long-distance dependence of the sequence and better overcome the disadvantage of gradient disappearance in RNN. This deep learning mechanism allows the automatic construction of a model from a problem or set of rules. When dealing with large amounts of data, the model can adapt to input new data or import new knowledge through other models, allowing it to solve almost any real-world task (19). Wang et al. (20) established BiLSTM network by composing forward and backward LSTM and applied it in the experiment respiratory data of 103 patients with malignant tumors. Through the experiment, they found that the best prediction effect was obtained when seven-slice BiLSTM was used, with an average absolute error of 0.074 mm and a root mean square error (RMSE) of 0.097 mm at a delay standard of 400 ms, which was three to five times higher than the prediction precision of ARIMA and multi-layer perceptron neural network (ADMLP-NN). Compared with traditional prediction models, the deep learning network with higher robustness can greatly improve the prediction precision, which can be applied to data of different patients and reduce the

interference of delay time. However, deeper network will lead to longer update time of prediction, which is not conducive to the update of prediction model. The Bidirectional Gated Recurrent Unit (Bi-GRU) rapid breathing prediction model was constructed by Yu et al. (21) by using a variant of LSTM–gating cycle unit (GRU), consequently reducing the time efficiency by about 30% compared with the LSTM model, which greatly improved the update time of the prediction network. Therefore, deep learning will be an emerging force driving progress in the field of respiratory motion prediction.

In general, the prediction accuracy of the model can be greatly improved by training the model on the clinical data of a limited number of patients (18, 22). However, when the model is applied to new patient data, the prediction effect is greatly discounted, and the generalization ability of the prediction model needs to be improved. Each patient has different physical conditions and respiratory states, and it is of great significance to design a general model to predict the respiratory signals of different patients (23). The establishment of a general model requires a large amount of patient data as support, so deep learning has good applicability, because deep learning has better learning and analysis capabilities under a large amount of data.

3 MATERIALS AND METHODS

3.1 Respiratory Movement Data

The data used in this paper are a publicly available dataset derived from the Institute of Robotics and Cognitive Systems, University of Lubeck, Germany (24). This dataset contains the respiratory data of 103 patients with horacoabdominal tumors, with a total of 306 respiratory motion trajectories. Three markers are installed on the chest and abdomen of each patient, and the trajectory data of the markers moving along with the respiratory movement were recorded. An optical tracking sampling instrument with a sampling frequency of 26 Hz is used for sampling work.

3.2 Research Methods

In this study, we built a respiratory motion prediction model and used *in vitro* marker signals to predict tumor motion trajectories. **Figure 1** shows the process of tumor motion and machine positioning during radiotherapy. First, a tumor motion area in the lungs that follows the patient's breathing is determined, and then, the tumor motion trajectory is further captured in this area. Considering the problem of mechanical and computer delays, the respiratory motion prediction model needs to determine the trajectory of the tumor after a period of delay, and finally perform radiotherapy to kill tumor cells.

The overall framework of the breathing motion prediction method based on the deep CNN is shown in **Figure 2**, which is mainly divided into two steps (1): data preprocessing and feature extraction: abnormal detection and correction of respiratory signals and extraction of features using EMD decomposition signals (2); respiratory motion prediction model: a deep respiratory motion prediction model composed of multi-scale convolution neural network including SEnet attention mechanism and TCN for the prediction of respiratory position at different delay times from 200 to 500 ms.

3.2.1 Data Preprocessing

In order to extract more information features and reduce the influence of interference information on prediction. First, the integrated model Bagging is used to detect and correct the abnormal interval, and then, the original series is decomposed into several IMFs containing different time scales by EMD algorithm, and finally, the dataset is divided as the input of depth prediction model.

3.2.1.1 Remove Outliers

Because of the long time of data acquisition, tumor patients sometimes have actions such as coughing, sneezing, or speaking

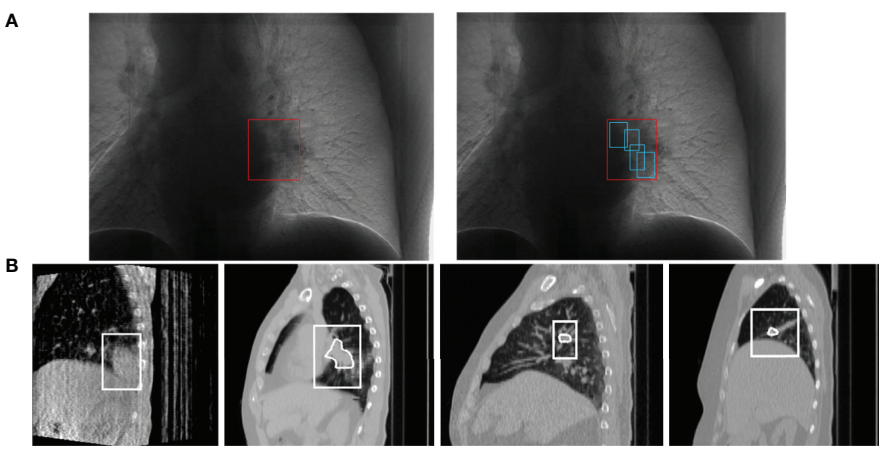


FIGURE 1 | Schematic diagram of lung tumor motion tracking, **(A)** is the process of tumor localization (25). Each of **(B)** is a 4DCBCT (four-dimensional cone beam CT) sequence image of tumor tracking at different stages in a respiratory process, obtained by the EELKTA Synergy XVI system in the University of Tokyo Hospital (26).

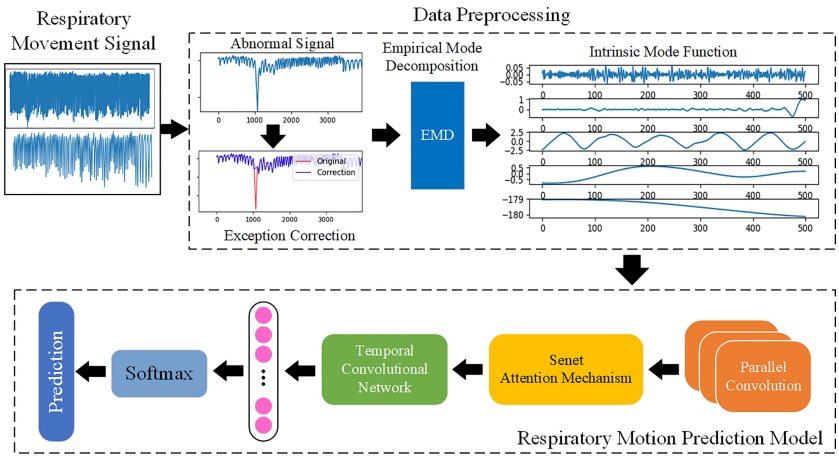


FIGURE 2 | The overall framework of respiratory motion. Use EMD to fully extract the features of the respiratory motion signal and learn the features by building a prediction model based on deep learning, so as to achieve accurate prediction.

during the acquisition process, which will greatly interfere the stability of respiratory trajectory, resulting in relatively intensive abnormal signals of respiratory data at a certain time segment. Therefore, this paper uses Bagging to deal with abnormal signals. Bagging mainly samples T sampling sets containing m training samples, then trains a base learner on the basis of each sampling set, and finally combines these base learners together (27). **Figure 3** shows a comparison diagram before and after processing an abnormal signal.

3.2.1.2 Empirical Mode Decomposition

Complex time series data will reduce the prediction precision of the prediction model, which will be alleviated to some extent by and the introduction of some decomposition algorithms in the pre-phase of data procession. Because the respiratory motion signal is a complex time series with non-linear, non-stationary, and univariate characteristics, when fitting this type of sequence with deep learning network, there are often problems such as gradient disappearance or explosion, and it is impossible to accurately identify the slight change characteristics of a certain time scale (28). Considering the multi-scale characteristics of time series, Fourier spectrum analysis and wavelet analysis are usually used to decompose the data to predict the better learning characteristics of the model. However, the limitations of these methods limit the operation of the prediction model to a certain extent, and empirical mode decomposition (EMD) can adaptively decompose complex signals. Compared with the above methods, EMD can more accurately reflect the original physical characteristics and local performance.

EMD decomposition is based on the following assumptions (29): the data have at least two extreme values (maximum and minimum); the local time–domain characteristics of the data are uniquely determined by the time scale between extreme points; if the signal is not extreme but contains an inflection point, then it

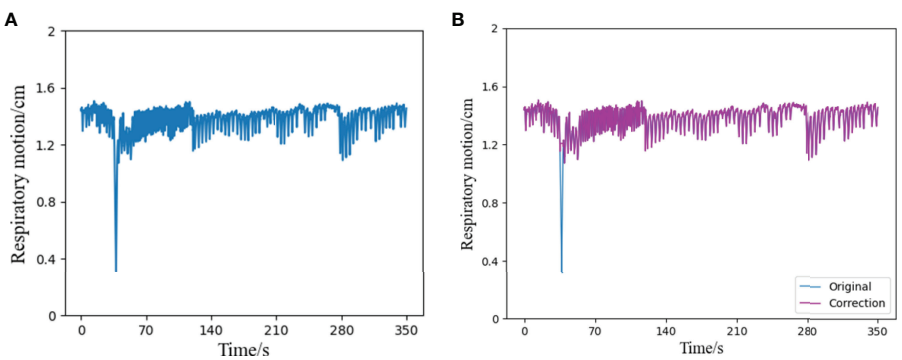


FIGURE 3 | Schematic diagram of outlier correction and comparison, **(A)** is a segment of the original respiratory signal, which contains an abnormal state in a certain time interval, **(B)** shows the result of the respiratory signal after the outlier correction algorithm. Compared with the original signal, it can be seen that the part containing outliers has been successfully corrected, and the rest remain unchanged.

can be differentiated once or more to obtain the extreme value. As for the given raw signal, $x(t)$ (t = 1,2…n),the EMD algorithm decomposition is described as follows:

- Extraction of the maximum and minimum values of $x(t)$: the upper and lower envelopes $X_{\max}(t)$,$X_{\max}(t)$ are formed by using the cubic spline difference to calculate their mean values m_1:
- Extraction details:

$$h_t = m(t) - m_1 \quad (1)$$

- Judgment of whether h_t IMF formation conditions: If it meets, then an IMF will be derived and the remaining volume $r(t) = x(t)$-$h(t)$ will be in lieu of (t) ; if not, then h_t will be in lieu of $x(t)$.
- Repetition of the above steps: When the standard deviation (0.2-0.3) is met the iteration will be ended.
- After the decomposition process, can be replaced by the following formula:

$$x(t) = \sum_{j=1}^{n} h_j(t) + r_n(t) \quad (2)$$

In this formula, n is the number of IMF; h_j (t) (j = 1.2,…n) are IMFs; and $r_n(t)$ is the final residual error, which indicates the central trend of $x(t)$.

For the generalization ability of the model, this paper uses the clinical respiratory data of 103 patients in the database and randomly selects a continuous signal (the total length of each signal is 10,000, about 7 min) from 306 respiratory trajectories as the model sample set. As shown in **Figure 4**, a series of length 10,000 is decomposed into nine IMF components and one residual (Res), and the physical meaning of each component of the IMF, whose order is divided according to the frequency from high to low, represents each frequency component of the raw signal. Because of the large amount of noise at high frequencies, the first two high-frequency IMF componets (IMF_0,IMF_1) are removed, and the remaining components will input the physical characteristics of the raw signal and into the prediction model. Because EMD is an adaptive decomposition, the respiratory data series of different patients will be decomposed into different amounts of IMF. Before being input into the network, it is necessary to supplement the number of IMFs of each original series in the whole database. The supplemented IMF components are filled with 0, so as to achieve a unified number of IMFs of each patient's respiratory series.

3.2.1.3 Division of Preprocessed Data

The training set, validation set, and test set are partitioned among the filtered IMF components. As shown in the division diagram **Figure 5**, the original sequence P = (p_1,p_2,…,p_i,…,p_{i+n}) is divided in a ratio of 6:2:2. In addition, the training set is indicated as P_{train} , P_{train} = $[p_1,p_2,\ldots,p_n]^T$; the valiadation set is denoted as P_{valid} , P_{valid} = $[p_j,\ldots, p_{j+n}]^T$; and the test set is denoted as P_{test} , P_{test} = $[p_j,\ldots, p_{j+n}]^T$. Take P_{train} as an example, form p-1 top$_n$ all sequences are isometric sequences, and each sequence contains the original sequence (X_1,X-2,…,X_n) and the delay time of the predicted value (t_1,…,t_n).

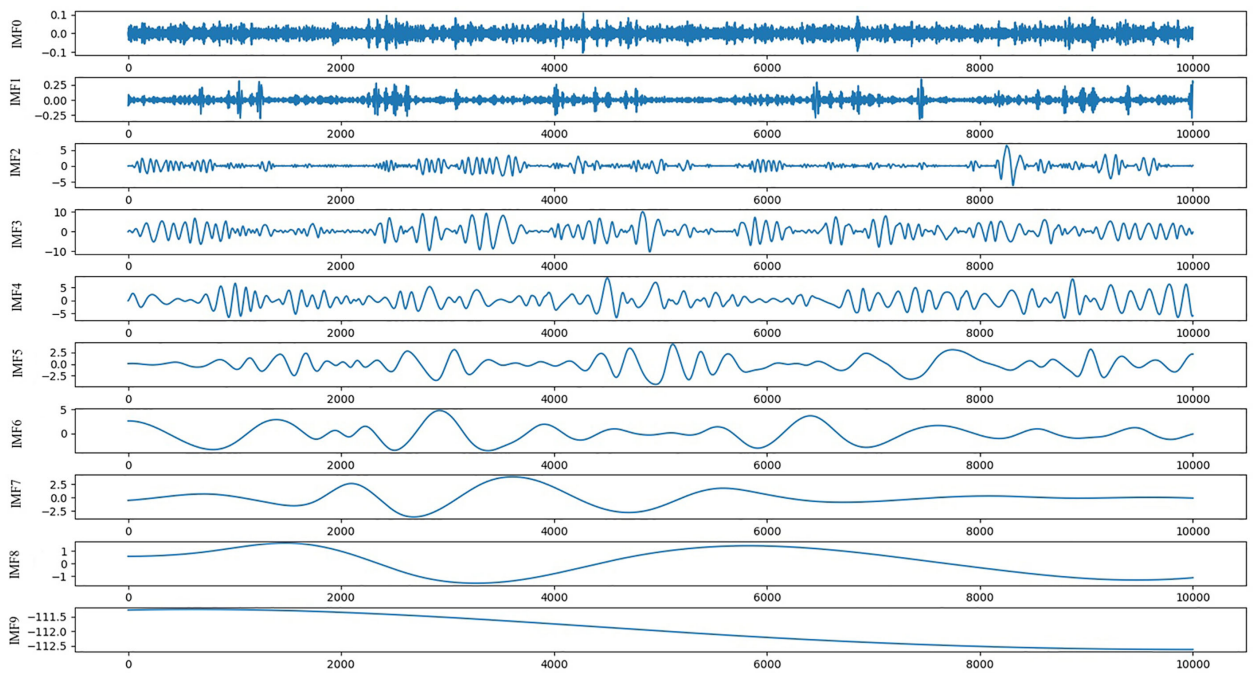


FIGURE 4 | Schematic diagram of IMFs and Res decomposed by EMD.

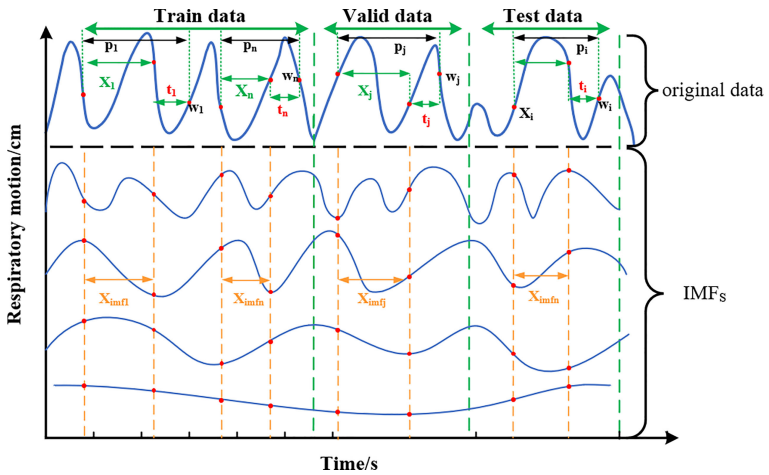


FIGURE 5 | Schematic diagram of training set, verification set, and test set classification. The original data at each end are decomposed by EMD to form multiple IMFs, and the data of equal length are intercepted as the feature input.

Network model input: After decomposition of the original sequence X, IMFs correspond to part of $Ximfs$, $X_{imfs} = [X_{imf1}, X_{imf2},\ldots, X_{imf1}, X_{imfi+n}]^T$, which is a stationary sequence containing multidimensional features. Target prediction value (label): observation point (w_1, w_2,…,w_n) after delay time t is the target prediction value, which is sampled from the original sequence and does not contain IMFs information. According to the equipment sampling frequency of 26 Hz, the corresponding delay time at t_i = 3,5,10, and 13 is about 100, 200, 400, and 500 ms, respectively.

3.2.2 Respiratory Motion Prediction Model

The deep convolution neural network model proposed in this paper for respiratory motion prediction includes three major parts. First, multi-scale convolution layers are used to extract features in parallel to find the optimal local sparse structure of the convolution network and obtain timing information fully. Second, the addition of a SEnet-based attention mechanism to the convolved feature channel increases the sensitivity of the model to the channel feature and automatically learns the importance of the different channel features. Last, TCNs are used to grasp long-time dependent information and assign each convolutional feature to a causal relationship, thereby predicting respiratory motion signals for a future period of time.

3.2.2.1 Squeeze and Exception Module

CNN has the ability of characterization learning, translates invariant classification of input information according to the hierarchical structure, and fuses spatial and channel information in the local receiving domain of each layer of network to construct local features. A squeeze-and-excitation module is proposed on the basis of CNN by Hu et al. (30), which improves the CNN characterization ability by improving the spatial coding quality at the feature level and clearly establishing the interdependence between convolutional feature channels.

3.2.2.2 Temporal Convolutional Network

The main characteristics of TCN include adopting a one-dimensional fully convolutional networks (FCNs) (31) to receive input sequences of any length as inputs and map them into output sequences of equal length at the same time; each time is calculated simultaneously, not serially on the time sequence, to improve the network operation efficiency; causal convolution is used, so that each convolution layer is causally related, which means that information "leakage" will not occur from future to the past. Briefly: TCN = 1D FCN + Causal convolutions (32).

3.2.2.2.1 Causal Convolutions. If the input sequence is shown as $X = (x_1.x_2,\ldots,x_r)$, then the prediction yt of the moment t can only be obtained through x_1 to x_{t-1}, which is input before moment t as what has been shown in the left half of **Figure 6A**. If the filter is defined as $F = (f_1,f_2,\ldots,f_k)$ and K is the number of filters, then the causal convolution at time x_t is as follows:

$$(F * X)(x_t) = \sum_{k=1}^{K} f_k x_{t-K+k} \tag{3}$$

There is a big defect in causal convolution. If a more distant x_{t-n} is needed as input to enlarge the receptive field, then a large number of convolution layers are needed, which increases the network depth and easily causes problems such as gradient disappearance and poor fitting effect.

3.2.2.2.2 Dilated Convolutions. Dilated convolution can be used to solve the above problems; meanwhile, it is also the convolution used by the TCN network. To obtain larger receptive field, the dilated convolution (d) introduces the concept of dilation factor, which allows the input interval adoption during the convolution. Adding to the dilation factor gives sequence X dilated convolution at x_t at which the expansion factor is d:

$$(F_d * X)(x_t) = \sum_{k=1}^{K} f_k \; x_{t-(K-k)d} \quad (4)$$

The right half of **Figure 6A** shows that $d = 1$ at input is a common convolution, with $d = 2$ for the first hidden layer and $d = 4$ for the second hidden layer, and the expansion factor increases exponentially by 2 as the network layer increases.

3.2.2.2.3 Residual Connections. The residual connection is added to the TCN network, which allows the network to transmit information across layers and solves the problems of gradient disappearance or explosion of deep network, and learning the overall transformation of input X changes into learning the partial modification of input X. In the TCN, residual blocks are used to replace convolution layers, which include dilated convolution with two layers and non-linear mapping. In addition, a WeightNorm and Droput regularization network is used in each layer, with a linear rectification function (Relu) as the activation function as shown in **Figure 6B**.

3.2.2.3 Network Layer of Respiratory Motion Prediction Model

The main body of respiratory motion prediction model is composed of multi-scale enhanced CNNs layer (CNN_SEnet) and a TCN layer. As shown in **Figure 7**, first, a multi-scale convolution channel is composed of a convolution layer containing different convolution kernels, and the sizes of each convolution kernel in each channel are 3 × 1, 5 × 3, and 7 × 5, respectively, with a step size of 1 and a convolution filter of 16. Setting convolution kernels at different scales allows the model to learn different local features in the sequence. For example, smaller convolution kernels can extract local subtle features and are more sensitive to instantaneous changes in the sequence; larger convolution kernels mainly extract local trend features and can control the overall features at a certain time scale. The input of the prediction model is $X_{imfs} = [X_{imf1}, X_{imf2}, \ldots, X_{imfi+n}]^{\mathrm{T}}$, in which the length of X_{imfn} is the IMFs containing a certain time length, about 100 to 400, and the width is the IMFs with different frequency components formed by the original sequence after EMD decomposition, about 10 to 15. Its length–width ratio gap is so large that the convolution kernel size is no longer set as the conventional 3 × 3 or 5 × 5 but set the convolution kernel of the above size, which can highlight the time–domain characteristics when the frequency–domain characteristics are ensured. Each scale channel contains a convolution layer of three above parameters for adequately extracting feature information in the sequence.

Second, to enhance the information representation ability of CNNs layer, SEnet attention mechanism is added after each CNNs channel, and the weight coefficient of each channel after convolution is learned, so that the model has more discrimination ability for the characteristics of each channel. Its network parameters are detailed in the literature. The activation function Relu and the maximum pooling layer with a 2 × 2 window are then performed for extracting important features and discard irrelevant features.

Then, the output of the three scale channels is combined through the connecting layer to form a richer information feature. Afterward, the causal relationship of each feature can be found out through the TCN layer, and the future information is predicted through the historical information feature. The number of filters in this module is set as 32; the convolution kernel size is 3; the dilation factor grows by 2^n; the number of stacks of residual blocks is 1, and the activation function is Relu. Last, the predicted target values were obtained through Flatten layer and full junctional layer.

3.2.3 Evaluation Criteria

In this paper, the mean absolute error (MAE), RMSE, and R2 determination coefficient (R2_score) are used as evaluation indexes of respiratory prediction algorithm. MAE is the mean

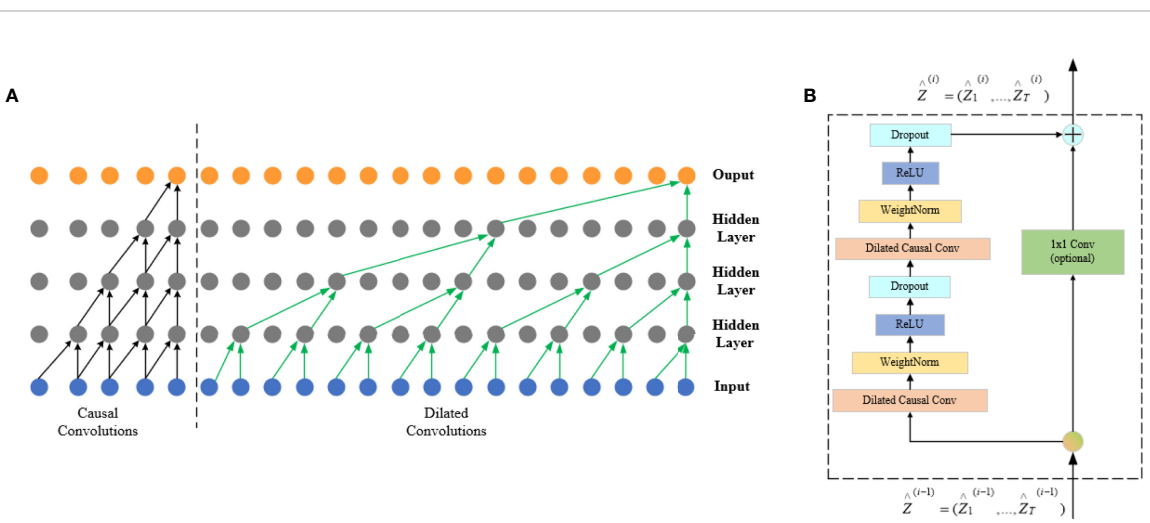


FIGURE 6 | FCN architecture. The left half of **(A)** is a causal convolution schematic and the right half is a dilated convolution schematic, and **(B)** is the TCN residual block.

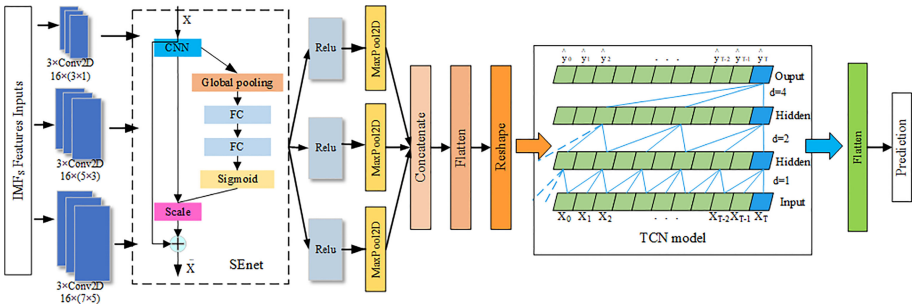


FIGURE 7 | Network layer of respiratory prediction model.

of the absolute value of the deviation between all individual observed value and the arithmetic mean. It is defined as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - y_i^*| \quad (5)$$

The RMSE is the square root of the ratio of the square of the deviation of the predicted value from the true value to the number of observations n, and it is defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y_i^*)^2} \quad (6)$$

R2_score is the overall fit of the regression equation, and the closer the value of R2 is to 1, the better the fit of the regression equation to the observed value is, which can be defined as follows:

$$RMSE = 1 - \frac{\sum_i (y_i - y_i^*)^2}{\sum_i (\bar{y}_i - y_i^*)^2} \quad (7)$$

In this equation, N is the number of data points; y is the actual respiratory motion trace; $y^\star$ is the trajectories of respiratory motion prediction; and $\sum_i(\bar{y}_i - y_i^*)^2$ is a benchmark model in the field of machine learning.

4 RESULTS AND DISCUSSION

4.1 Results

Table 1 and **Figure 8**, respectively, show the experimental results of the proposed EMD-SEnet-TCN multi-channel depth prediction model in this paper; in addition, the prediction results in this paper are all calculated according to the following parameters: epochs = 100, batch size = 128, optimizer = Adam, and learning rate = 0.001. Judging from the results, although the prediction precision decreases with the increase of delay time (t_i), the prediction accuracy is still ensured to some extent; when the length of model input data is increased, the network does not present gradient explosion or disappearance problems, which indicates that the proposed algorithm in this paper has the ability to overcome long-distance dependence and can make full use of historical information to predict the future information.

To verify the higher prediction precision of the model in this paper, a comparison is made with the Deep BiLSTM model proposed by Wang et al. (20) with the same dataset. **Figure 9** shows the comparison of these two algorithmic models under the same parameters (X_i = 50, t_i = 1.5, and 10). It can be seen from the figure that the prediction precision of the algorithm proposed in this paper is better at different delay times under the MAE and RMSE evaluation indexes.

Because of the limitations of different input samples, preprocessing operations, and experimental platforms under different models, to illustrate the superiority of this model more clearly, a comparison among three depth prediction models is conducted, including multi-convolution combined with BiLSTM network (CNN-BiLSTM), multi-channel convolution combined with TCN model (CNN-TCN), and multi-channel convolution combined with BiLSTM based on EMD (EMD-CNN-BiLSTM). **Table 2** shows the performance comparison results of the proposed algorithm (EMD-SEnetTCN) with the above three models at X_i = 100 and delay times at 80, 150, 240, 300, 400, 450, and 520 ms (t_i = 2, 4, 6, 8, 10, 12, and 14).

As shown in **Figure 10**, the prediction precision of each model is high, and there is no significant difference when the delay time is shorter than 240 ms. The MAE and RMSE are about 0.72% ~ 0.18% and 0.21% ~ 0.28%, respectively. When the delay time exceeds 240 ms, the better performance of EMD-SEnet-TCN becomes more and more obvious. To meet the clinical requirements, 400 ms is used as the standard delay time. Compared with CNN-TCN, the precision decrease of MAE and RMSE are by 13.7% and 9.2%, respectively, whereas for R2_score, the precision increases by 2%. The difference between

TABLE 1 | Results of respiratory prediction algorithm.

Input Length (Xi)	Latency (ms)	MAE (mm)	RMSE (mm)	R2 (None)
50	120 (t_i = 3)	0.009022	0.022503	0.989431
100	200 (t_i = 5)	0.016584	0.031588	0.979483
200	400 (t_i = 10)	0.035926	0.053782	0.941398
400	500 (t_i = 13)	0.048367	0.068925	0.908258

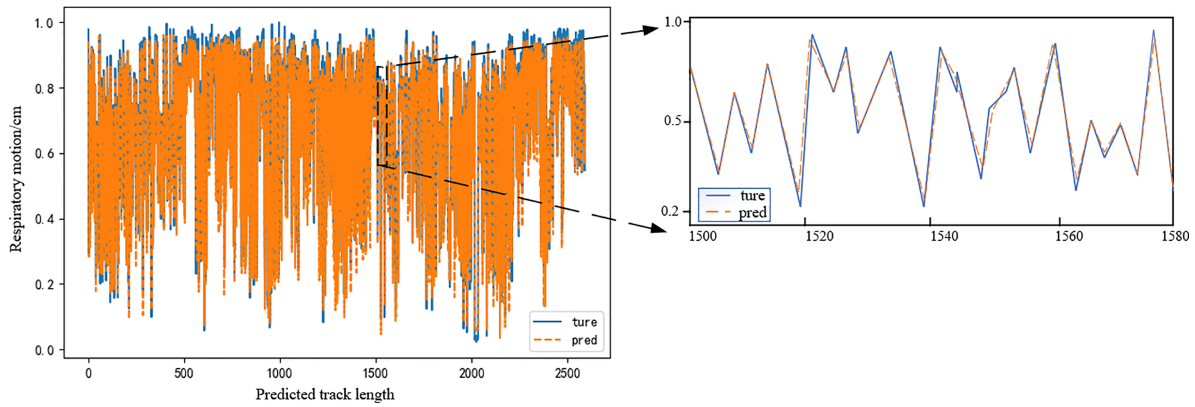


FIGURE 8 | Actual breathing trajectory and predicted trajectory. Model delay time is 400 ms.

these two models is whether EMD is used or not. Judging from the results, EMD is very effective for improving the precision of the model. Compared with CNN-BiLSTM, the precision values of MAE and RMSE decreased by 15% and 18.3%, respectively, whereas the precision value of R2_score increased by 1.4%; compared with EMD-CNN-BiLSTM, the precision values of MAE and RMSE decreased by 8.4% and 9.1%, respectively, whereas the precision value of R2_score increased by 1%, of MAE decreased by 8.4%, of RMSE decreased by 9.1%, and of R2_score increased by 1%. The prediction precision of this model is very close to that of this paper due to the similar structure of the two depth models and the difference lies in TCN and BiLSTM. EMD-SEnet_TCN not only has higher precision but also improves of prediction update time. The results show that, compared with other prediction models, the model in this paper has excellent performance at different delay times, and the prediction model performance will be further improved with the increase of delay time.

Figure 11 shows the prediction update time of different depth models in seconds per epoch. Although EMD-CNN-BiLSTM is slightly inferior to the model proposed in this paper in terms of prediction precision, the update time has reached 10 s per epoch, which is much longer than the update time of EMD-SEnet_TCN (2 s per epoch), failing to meet the clinical requirements; whereas the update time of CNN-TCNs is the shortest, only 1 s per epoch, without meeting the standard of prediction pricision; as for other prediction models, all perform poorly in terms of precision or update time. In general, the prediction model proposed in this paper greatly reduces the average update time with the guarantee of high prediction precision, so that the network can predict the target value quickly and accurately.

The input data length of the model affects the prediction precision to a certain extent. Generally, to lower the prediction error, the input data segment should be located near the target prediction value because the farther the distance is, the weaker the correlation is. In addition, if the data is too long, then there will be problems such as increased training time of the prediction model and gradient disappearance or explosion. To study the effect of different lengths of input data on the prediction results, the prediction errors of different data with lengths of 50, 100, 200, 400, and 600 at a delay time of 400 ms (t_i = 10) are compared. The results are shown in **Figure 12**.

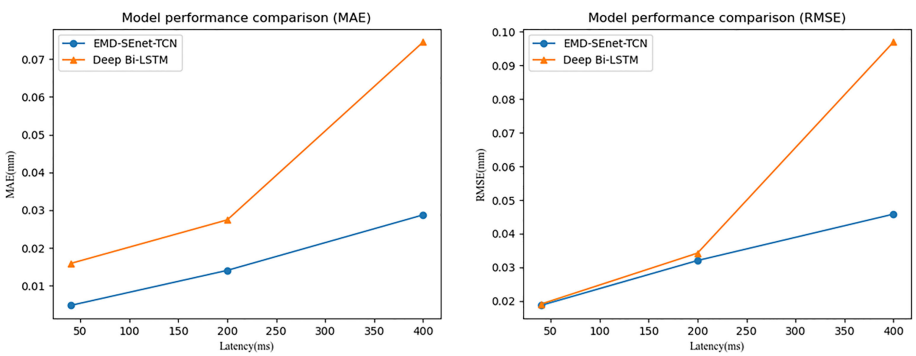


FIGURE 9 | Model performance comparison.

TABLE 2 | Results comparison of different respiratory prediction models.

Prediction model	Latency (ms)	MAE (mm)	RMSE (mm)	R2 (None)
EMD-SEnet-TCN	80	0.008797	0.01814	0.993157
	150	0.016442	0.026901	0.985422
	240	0.021495	0.033960	0.975841
	300	0.028391	0.042698	0.960636
	400	0.031789	0.0491499	0.951819
	450	0.038560	0.058295	0.928711
	520	0.045638	0.064746	0.910043
CNN-BiLSTM	80	0.013164	0.020244	0.994183
	150	0.015645	0.021739	0.989963
	240	0.026275	0.032612	0.966891
	300	0.040191	0.051979	0.939334
	400	0.044840	0.060412	0.917064
	450	0.051416	0.071154	0.890039
	520	0.059593	0.080446	0.857721
CNN-TCN	80	0.007193	0.009572	0.997983
	150	0.020939	0.028718	0.985953
	240	0.022003	0.031190	0.978862
	300	0.029881	0.045487	0.961351
	400	0.040356	0.054568	0.932332
	450	0.048732	0.070259	0.893678
	520	0.054356	0.077895	0.869860
EMD-CNN-BiLSTM	80	0.010331	0.022604	0.989376
	150	0.017215	0.025713	0.986681
	240	0.0257432	0.038429	0.969060
	300	0.029316	0.045936	0.954440
	400	0.037918	0.054366	0.937525
	450	0.040603	0.060994	0.911850
	520	0.048301	0.065581	0.907710

With the increase of input data (X_i) the prediction errors of different models increase, among which the gradient of CNN-BiLSTM disappears at X_i = 600 and both MAE and RMSE increase abnormally; EMD-CNN-BiLSTM and CNN-BiLSTM have better prediction precision when X_i is small, but the prediction error increases rapidly when X_i is big; CNN-TCN has a more stable prediction error fluctuation at different X_i whereas that of MAE and RMSE are big; comparing the above three models, EMD-SEnet_TCN displays excellent prediction performance in that it can cope with sequence information of various lengths and ensure certain prediction precision.

4.2 Discussion

Choosing different optimizer (Op) and learning rate (Lr) will affect the prediction results of deep prediction model. The optimizer is used to update and calculate the network parameters affecting the training and output of model, so that they approximate or reach the optimal value to minimize (or maximize) the loss function. The learning rate determines whether the objective function can converge to the local minimum value and when it can converges to the minimum value. The appropriate learning rate can make the target function converge to the local minimum value at appropriate time. SGD is a relatively commonly used optimizer, in which noise will be added when the gradient is randomly selected, and the update weight value does not reach the global optimum, which makes the accuracy rate decrease; Adagrad adopts an adaptive learning rate optimization algorithm to update the low-frequency parameters greatly while update the high-frequency parameters less; Adadelta is an improvement of Adagrad because it has an exponential decay average; RMSprop changes the gradient accumulation of Adagrad into an exponentially weighted moving average, improving the effect under non-convex settings; Adam combines the momentum advantages of RMSprop with SGD to form an optimizer with better performance.

Different optimizers display differently in various tasks, and it is not necessarily that the more advanced the version is, the better its results are. To select a better optimizer, the comparison of different optimizers is performed in **Table 3**. The learning rate controls the update speed of model parameters–Lr is too small, it will greatly reduce the network convergence rate and increase the training time; if it is too large, then it will lead to parameters oscillating on both sides of the optimal solution. **Table 3** below shows the prediction model performance results of different sizes of learning rates (0.1, 0.01, 0.001, and 0.0001).

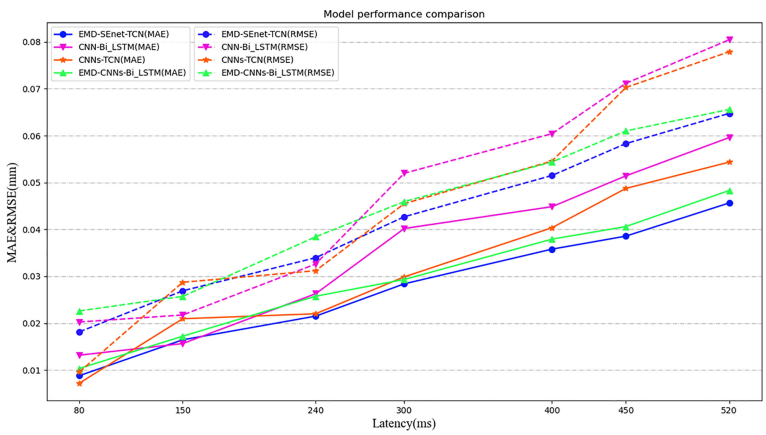


FIGURE 10 | Model performance comparison under different depths and different evaluation criteria. Under different predictive evaluation indicators, the performance of this model is compared with the other three models. The blue represents the model of this paper, the solid line represents the MAE indicator, and the dashed line represents the RMSE indicator.

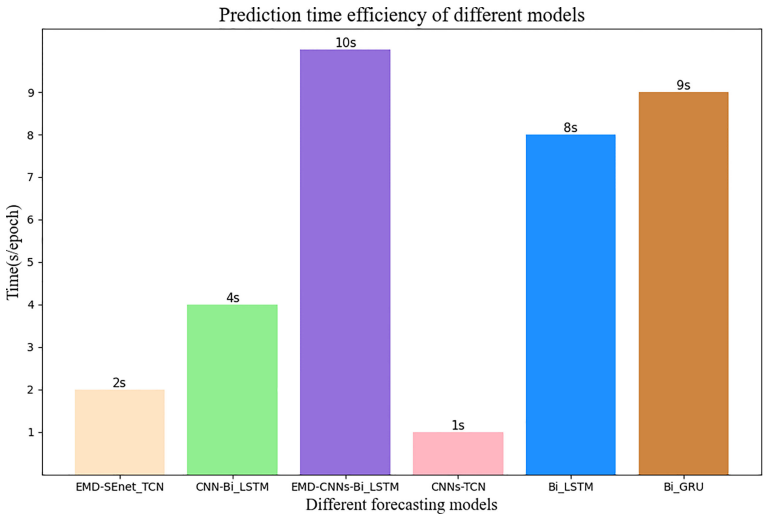


FIGURE 11 | Average update time of different prediction models. The number at the top of the bar graph represents time, and the color of wheat represents the model of this article.

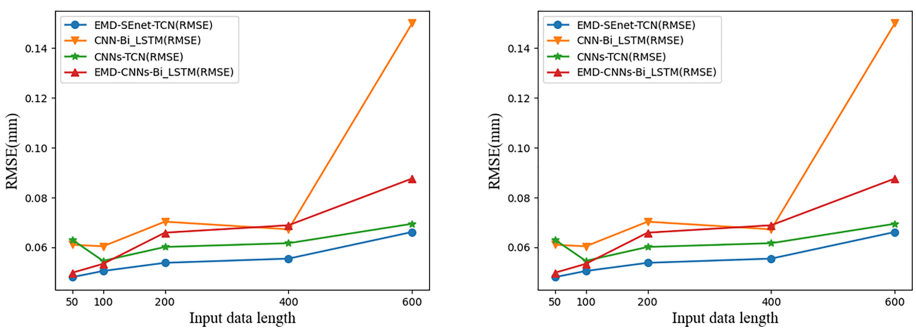


FIGURE 12 | Comparison of prediction results of input data with different lengths. Blue represents the model of this article.

All the results in **Table 3** are based on EMD-Senet-TCN prediction model with epochs = 100, batch size = 128, X_i = 100, t_i = 10 (400 ms). From **Table 3**, it can be seen that Op uses Adam. MAE and RMSE are the smallest and their prediction is the most accurate. Although Adadelta is an advanced version of Adagrad, it is not very effective when applied under the prediction model in this paper. The different learning rate settings were all obtained under Op = Adam, and the best result was obtained when Lr = 0.001, where when Lr = 0.01, the learning rate is too large to result in a model that could not converge and the regression coefficient was negative. It can be seen that the model in this paper uses Op = Adam and Lr = 0.001 to the best prediction results.

TABLE 3 | Effect of different parameters (Op, Lr) on EMD-SEnet_TCN.

Parameters	MAE (mm)	RMSE (mm)	R2 (None)
Op = SGD	0.087130	0.116361	0.702972
Op = Adam	0.035789	0.051499	0.941819
Op = Adagrad	0.049455	0.069139	0.895137
Op = Adadelta	0.166532	0.200430	0.118740
Op = RMSprop	0.039013	0.060148	0.910473
Lr = 0.1	0.308492	0.375175	−2.08768
Lr = 0.01	0.037244	0.053731	0.936662
Lr = 0.001	0.034789	0.049149	0.951819
Lr = 0.0001	0.041593	0.057317	0.927934

5 CONCLUSION

Respiratory motion brings great difficulties to the treatment of thoracoabdominal tumors, and respiratory motion prediction models are extremely important for precision radiotherapy. In this paper, a depth prediction model (EMD-SEnet-TCN) is proposed for the application of respiratory motion signals in radiation therapy for patients with cancer. The method was validated by using respiratory motion signals from multiple patients with malignant tumors in the database of the Institute of

Robotics and Cognitive Systems, University of Lübeck, Germany. The results of this paper show that (1) the depth prediction model method proposed in this paper is superior to other benchmark models in terms of delay prediction precision and time update efficiency (2); it verifies that the decomposition of complex respiratory motion signals by using EMD can further improve the prediction precision of the prediction model (3); the multi-scale CNN containing attention mechanisms has a better feature extraction ability for finite IMFs of respiratory motion signals. This work solves one of the major challenges for precise prediction of the state of patient respiratory motion signals, and in medical practice, the proposed method has important practical significance for precision radiation therapy.

The present study has some limitations. The first one is the correlation between the external respiratory signal and the internal tumor motion. In order for our technique to be applied clinically, another model needs to be designed to realize the correlation analysis in the future. The second is that whether the prediction technology in this paper achieves clinical application is the key to future research. On the basis of complying with legal and ethical requirements and respecting patient privacy, it is very important to determine a medical analysis platform that applies the deep learning framework in the future.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at http://signals.rob.uni-luebeck.de/.

AUTHOR CONTRIBUTIONS

Conceptualization: LS and JZ; Formal analysis: LS and SH; Supervision: ZK and WJ; Investigation: YC and ZZ; Writing—original draft: JZ and SH; Writing—review and editing: LS and JZ. All authors contributed to the article and approved the submitted version.

FUNDING

Funding sources: the Jilin Provincial Department of Science and Technology (Grant/Award Number: No.20190201195JC. 20200601004JC.20200301054RQ.20200404207YY), Science and Technology Development Plan of Jilin Province (Grant/Award Number: 20200403120SF), Natural Science Foundation of Jilin Province (Grant/Award Number: 20210101477JC), and the National Natural Science Foundation of China for supporting the research in this article (Grant/Award Number: No.61502052).

ACKNOWLEDGMENTS

The authors thanks the dataset provided by the Institute of Robotics and Cognitive Systems, University of Lübeck, Germany. We would like to thank the Jilin Provincial Department of Science and Technology and the National Natural Science Foundation of China for their support of this research, as well as the authors for their joint efforts.

REFERENCES

1. Zhao B, Yang Y, Li T, Li X, Heron DE, Huq MS. Dosimetric Effect of Intrafraction Tumor Motion in Phase Gated Lung Stereotactic Body Radiotherapy. *Med Phys* (2012) 39:6629–37. doi: 10.1118/1.4757916
2. Shirato H, Suzuki K, Sharp GC, Fujita K, Onimaru R, Fujino M, et al. Speed and Amplitude of Lung Tumor Motion Precisely Detected in Four-Dimensional Setup and in Real-Time Tumor-Tracking Radiotherapy. *Int J Radiat Oncol Biol Phys* (2006) 64:1229–36. doi: 10.1016/j.ijrobp.2005.11.016
3. Rydhög JS, de Blanck SR, Josipovic M, Jølck RI, Larsen KR, Clementsen P, et al. Target Position Uncertainty During Visually Guided Deep-Inspiration Breath-Hold Radiotherapy in Locally Advanced Lung Cancer. *Radiother Oncol* (2017) 123:78–84. doi: 10.1016/j.radonc.2017.02.003
4. Herfarth K, Debus J, Lohr F, Bahner M, Fritz P, Höss A, et al. Extracranial Stereotactic Radiation Therapy: Set-Up Accuracy of Patients Treated for Liver Metastases. *Int J Radiat Oncol Biol Phys* (2000) 46:329–35. doi: 10.1016/S0360-3016(99)00413-7
5. Oh SA, Yea JW, Kim SK. Statistical Determination of the Gating Windows for Respiratory-Gated Radiotherapy Using a Visible Guiding System. *PloS One* (2016) 11:e0156357. doi: 10.1371/journal.pone.0156357
6. Murphy MJ. Tracking Moving Organs in Real Time. *Semin Radiat Oncol (Elsevier)* (2004) 14:91–100. doi: 10.1053/j.semradonc.2003.10.005
7. Vedam S, Kini V, Keall P, Ramakrishnan V, Mostafavi H, Mohan R. Quantifying the Predictability of Diaphragm Motion During Respiration With a Noninvasive External Marker. *Med Phys* (2003) 30:505–13. doi: 10.1118/1.1558675
8. Ozhasoglu C, Murphy MJ. Issues in Respiratory Motion Compensation During External-Beam Radiotherapy. *Int J Radiat Oncol Biol Phys* (2002) 52:1389–99. doi: 10.1016/S0360-3016(01)02789-4
9. Fayad H, Pan T, François Clement J, Visvikis D. Correlation of Respiratory Motion Between External Patient Surface and Internal Anatomical Landmarks. *Med Phys* (2011) 38:3157–64. doi: 10.1118/1.3589131
10. Richter L, Ernst F, Martens V, Matthäus L, Schweikard A. Client/server Framework for Robot Control in Medical Assistance Systems. *Int J Comput Assist Radiol Surg* (2010) 5:306–7.
11. Depuydt T, Verellen D, Haas O, Gevaert T, Linthout N, Duchateau M, et al. Geometric Accuracy of a Novel Gimbals Based Radiation Therapy Tumor Tracking System. *Radiother Oncol* (2011) 98:365–72. doi: 10.1016/j.radonc.2011.01.015
12. Smith RL, Abd Rahni AA, Jones J. A Kalman-Based Approach With Em Optimization for Respiratory Motion Modeling in Medical Imaging. *IEEE Trans Radiat Plasma Med Sci* (2018) 3(4):410–20. doi: 10.1109/TRPMS.2018.2879441
13. Hong S, Bukhari W. Real-Time Prediction of Respiratory Motion Using a Cascade Structure of an Extended Kalman Filter and Support Vector Regression. *Phys Med Biol* (2014) 59:3555. doi: 10.1088/0031-9155/59/13/3555
14. Ernst F, Schlaefer A, Schweikard A. Prediction of Respiratory Motion With Wavelet-Based Multiscale Autoregression. *Int Conf Med Imag Comput Computer-Assisted Intervent* (2007) 4792:668–75. doi: 10.1007/978-3-540-75759-7_81
15. Ernst F, Schweikard A. Prediction of Respiratory Motion Using a Modified Recursive Least Squares Algorithm. *CURAC* (2008) 8:157–60. doi: 10.1.1.149.4134
16. Homma N, Sakai M, Takai Y. Time Series Prediction of Respiratory Motion for Lung Tumor Tracking Radiation Therapy. In: *Proceedings of the 10th WSEAS International Conference on Neural Networks.* Prague, Czech Republic: World Scientific and Engineering Academy and Society (WSEAS) (2009), 126–31. doi: 10.5555/1561799.1561822

17. Tsai TI, Li DC. Approximate Modeling for High Order Non-Linear Functions Using Small Sample Sets. *Expert Syst Appl* (2008) 34:564–9. doi: 10.1016/j.eswa.2006.09.023
18. Sun W, Jiang M, Ren L, Dang J, You T, Yin F. Respiratory Signal Prediction Based on Adaptive Boosting and Multi-Layer Perceptron Neural Network. *Phys Med Biol* (2017) 62:6822. doi: 10.1088/1361-6560/aa7cd4
19. Wei W, Ke Q, Nowak J, Korytkowski M, Scherer R, Woźniak M. Accurate and Fast Url Phishing Detector: A Convolutional Neural Network Approach. *Comput Networks* (2020) 178:107275. doi: 10.1016/j.comnet.2020.107275
20. Wang R, Liang X, Zhu X, Xie Y. A Feasibility of Respiration Prediction Based on Deep Bi-Lstm for Real-Time Tumor Tracking. *IEEE Access* (2018) 6:51262–8. doi: 10.1109/ACCESS.2018.2869780
21. Yu S, Wang J, Liu J, Sun R, Kuang S, Sun L. Rapid Prediction of Respiratory Motion Based on Bidirectional Gated Recurrent Unit Network. *IEEE Access* (2020) 8:49424–35. doi: 10.1109/ACCESS.2020.2980002
22. Tang S, Andres B, Andriluka M, Schiele M. Multi-person Tracking by Multicut and Deep Matching. *Computer Vision - {ECCV} 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part {II}* (2016) 9914:100–11. doi: 10.1007/978-3-319-48881-3_8
23. Teo TP, Ahmed SB, Kawalec P, Alayoubi N, Bruce N, Lyn E, et al. Feasibility of Predicting Tumor Motion Using Online Data Acquired During Treatment and a Generalized Neural Network Optimized With Offline Patient Tumor Trajectories. *Med Phys* (2018) 45:830–45. doi: 10.1002/mp.12731
24. Ernst F. *Compensating for Quasi-Periodic Motion in Robotic Radiosurgery*. Berlin, Germany: Springer Science & Business Media (2011).
25. Shumeng H, Shanda M, Wei W, Dongshan F. Lung Tumor Motion Tracking Method and Clinical Evaluation Based on Dual-Energy X-Ray Fluoroscopic Imaging. *J Tianjin Med Univ* (2020) 26:6. doi: 10.1109/TKDE.2016.2609424
26. Pohl M, Uesaka M, Demachi K, Chhatkuli RB. Prediction of the Motion of Chest Internal Points Using a Recurrent Neural Network Trained With Real-Time Recurrent Learning for Latency Compensation in Lung Cancer Radiotherapy. *Computer Med Imaging Graphics* (2021) 91:101941. doi: 10.1016/j.compmedimag.2021.101941
27. Oza NC, Russell SJ. 2005 IEEE International Conference on Systems, Man and Cybernetics. *Online Bagging and Boosting* (2001) 3:2340–5. doi: 10.1109/ICSMC.2005.1571498
28. Perais A, Seznec A. Eole: Combining Static and Dynamic Scheduling Through Value Prediction to Reduce Complexity and Increase Performance. *ACM Trans Comput Syst (TOCS)* (2016) 34:1–33. doi: 10.1145/2870632
29. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and non-Stationary Time Series Analysis. *Proc R Soc London. Ser A: Mathe Phys Eng Sci* (1998) 454:903–95. doi: 10.1098/rspa.1998.0193
30. Hu J, Shen L, Sun G. Squeeze-And-Excitation Networks. In: *2018 {IEEE} Conference on Computer Vision and Pattern Recognition, (CVPR) 2018, Salt Lake City, UT, USA, June 18-22, 2018*. Computer Vision Foundation / (IEEE) Computer Society (2018) 7132–41. doi: 10.1109/CVPR.2018.00745
31. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. In: *(IEEE) Conference on Computer Vision and Pattern Recognition, {CVPR} 2015, Boston, MA, USA, June 7-12, 2015*. (IEEE) Computer Society (2015) 3431–40. doi: 10.1109/CVPR.2015.7298965
32. Bai S, Kolter JZ, Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* (2018) abs/1803.01271. doi: 10.48550/arXiv.1803.01271

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 | Frontiers in Oncology

ORIGINAL RESEARCH
published: 01 June 2022
doi: 10.3389/fonc.2022.873268

A Sequential Machine Learning-cum-Attention Mechanism for Effective Segmentation of Brain Tumor

Tahir Mohammad Ali[1], Ali Nawaz[1], Attique Ur Rehman[2], Rana Zeeshan Ahmad[3], Abdul Rehman Javed[4], Thippa Reddy Gadekallu[5], Chin-Ling Chen[6,7,8]* and Chih-Ming Wu[9]*

[1] Department of Computer Science, GULF University for Science and Technology, Mishref, Kuwait, [2] Department of Software Engineering, University of Sialkot, Sialkot, Pakistan, [3] Department of Information Technology, University of Sialkot, Sialkot, Pakistan, [4] Department of Cyber Security, Air University, Islamabad, Pakistan, [5] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India, [6] School of Information Engineering, Changchun Sci-Tech University, Changchun, China, [7] School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China, [8] Department of Computer Science and Information Engineering, Chaoyang University of Technology, Taichung, Taiwan, [9] School of Civil Engineering and Architecture, Xiamen University of Technology, Xiamen, China

OPEN ACCESS

Edited by:
Wei Wei,
Xi'an University of Technology, China

Reviewed by:
Mohammad Hamghalam,
Shenzhen University, China
Chennareddy Vijay Simha Reddy,
Middlesex University, United Kingdom
Praveen Kumar Donta,
Vienna University of Technology, Austria
Keping Yu,
Waseda University, Japan

***Correspondence:**
Abdul Rehman Javed
abdulrehman.cs@au.edu.pk
Chin-Ling Chen
clc@mail.cyut.edu.tw

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

Received: *10 February 2022*
Accepted: *18 April 2022*
Published: *01 June 2022*

Citation:
Ali TM, Nawaz A, Ur Rehman A, Ahmad RZ, Javed AR, Gadekallu TR, Chen C-L and Wu C-M (2022) A Sequential Machine Learning-cum-Attention Mechanism for Effective Segmentation of Brain Tumor. Front. Oncol. 12:873268. doi: 10.3389/fonc.2022.873268

Magnetic resonance imaging is the most generally utilized imaging methodology that permits radiologists to look inside the cerebrum using radio waves and magnets for tumor identification. However, it is tedious and complex to identify the tumorous and nontumorous regions due to the complexity in the tumorous region. Therefore, reliable and automatic segmentation and prediction are necessary for the segmentation of brain tumors. This paper proposes a reliable and efficient neural network variant, i.e., an attention-based convolutional neural network for brain tumor segmentation. Specifically, an encoder part of the UNET is a pre-trained VGG19 network followed by the adjacent decoder parts with an attention gate for segmentation noise induction and a denoising mechanism for avoiding overfitting. The dataset we are using for segmentation is BRATS'20, which comprises four different MRI modalities and one target mask file. The abovementioned algorithm resulted in a dice similarity coefficient of 0.83, 0.86, and 0.90 for enhancing, core, and whole tumors, respectively.

Keywords: VGG19, UNET, attention mechanism, brain tumor segmentation, MRI, BRATS

INTRODUCTION

Glioma is the most common type of tumor that is difficult to detect, with the lowest survival rate of 22% and constituting about 33% of all brain tumors (1–3). Some brain tumors are noncancerous, called benign, with a high survival rate, and some brain tumors are cancerous, known as malignant, with a low survival rate. There are also two types of brain tumors based on origin. The first is a primary brain tumor because it originates in the brain and occurs due to abnormal brain cells; it is also known as mutations. As cells mutate, they grow to multiply uncontrollably, forming a mass or tumor. A brain tumor is among the leading cause of death. Conversely, tumors that have spread to the brain from other locations in the body are known as brain metastasis, or secondary brain tumors (4). According to a 2019 report from the London Institute of Cancer and World Health

Organization (WHO),[1] there are approximately eighteen million registered cancer cases worldwide. Of these, 286,000 cases are brain tumors, and the highest cases of brain tumors are reported in Asia, with 156,000 cases. According to the same report, approximately 9 million deaths are due to global cancer. Out of which, 241 deaths are due to a brain tumor, and the highest mortality rate was observed in Asia with 129 cases.

Brain tumor segmentation aims to detect the extension and location of tumor regions (5). These regions are necrotic, edema, and active tissues, usually achieved by identifying abnormal areas compared to normal tissue. As glioma is the most common type of brain tumor and is hard to detect manually due to confusing boundaries, more than one MRI modality for detection was utilized (6). The different forms of MRI modalities are T1-weighted images, T2-weighted images, and fluid attenuation inversion recovery (FLAIR)-weighted images (7). These images were distinguished based on repetition time (TR) and time to echo (TE). T1-weighted images are generated using short TR and TE. T2-weighted images are generated using longer TR and TE than T1-weighted images, and FLAIR-weighted images are generated using longer TR and TE than T2-weighted images.

Previous brain tumor segmentation methods used hand-designed features. Those methods were based on the classical machine learning approach in which features are first extracted by applying statistical approaches, and then machine learning algorithms were applied for brain tumor segmentation (8, 9). In these techniques, the nature of the features did not affect the training procedure of the classifier. An alternative approach to this is automatically extracting the features used for brain tumor segmentation. This approach is most recently used and is known as deep learning. Deep learning is the study of deep neural networks (DNN), and DNN automatically learns the hierarchy of complex features directly from available data (10). Specifically, we use a pre-trained Convolutional Neural Network (CNN) (11, 12), i.e., VGG19, for brain tumor segmentation. CNN is the most widely used DNN for computer vision tasks. Similar to DNN, the standard CNN comprises the input, hidden, and output layers. The different hidden layers are convolutional, pooling, and fully connected. The working of CNN is simple: it compares the image pixels. These pixels are also known as the features of the image.

To summarize, a pre-trained CNN learned the pixels of the image by passing through different hidden layers. Therefore, in this research, we apply CNN to automatically learn feature hierarchy and utilize it for brain tumor segmentation. Subsequently, the binary classification of tumors and nontumorous regions are performed, and their results are utilized to classify all types of tumors. An overview of the whole sequential research methodology is presented in **Figure 1**. Specifically, we will propose a fully automatic, efficient encoder-decoder architecture by using BRATS'20 datasets.

The main contributions of this research article are summarized as;

- An attention-based mechanism reduces computational complexity problems and improves brain tumor segmentation results. Specifically, an image processing and attention mechanism are applied to extract the specified area of the image, followed by a pre-trained encoder part to extract the minimum but valuable features for further improving the results with efficiency.
- The implementation of the proposed framework in PYTHON using state-of-the-art libraries. The complete code is available on the GitHub repository. https://github.com/alinawazT/Brain-Tumor-Segmentation
- The validation of the proposed method was performed on the BraTS'20 and improved the Dice Similarity Coefficient (DSC) of enhancing, whole, and core tumors with 0.83, 0.90, and 0.86, respectively.

The rest of the paper is organized as follows: Section 2 highlights the previous work related to the brain tumor and addresses the research gaps. The proposed methodology is presented in Section 3. The comparison of the results with the state-of-the-art methods is presented in Section 4. Finally, Section 5 concludes the research paper with expected future research.

RELATED WORKS

A brain tumor is hard to detect manually due to nonuniform shapes and confusing boundaries (13). Therefore, deep learning and image processing play an essential role in early brain tumor diagnosis. Different intelligent techniques were proposed for automatic early diagnosis and segmentation of the tumor region. Among them, CNN and Ensemble learning are the most widely used techniques. A short review of some of the prominent and latest techniques is presented below.

Zeldin et al. (14) apply different pre-trained deep learning architectures for fully automatic segmentation of brain tumors. Specifically, different CNN models such as dense convolutional network (DenseNet), residual neural network (ResNet), and NASNet were utilized as encoders. Like conventional U-NET, an encoder is a CNN responsible for feature extraction followed by separate decoder parts to achieve the semantic probability map. The evaluation of the proposed method was performed on BRATS'19 datasets and achieved a DSC of 0.839, 0.837, 0.839, and 0.835 on Xception, VGGNet, DenseNet, and MobileNet encoders, respectively.

Pei et al. (13) proposed a context-aware deep neural network (CANet) framework for brain tumor segmentation. In addition to U-NET's encoder and decoder parts, it has a context encoding module that computes scaling factors of all classes. This scaling factor learns the global representation of all tumor classes. The validation of the proposed method was performed on the BRATS'19 and BRATS'20 datasets, and the evaluation metric used in the experimentation was DSC. The DSC on the test set was 0.821, 0.895, and 0.835 for enhancing tumor (ET), whole tumor (WT), and core tumor (TC), respectively.

Ghosh et al. (15) proposed a pre-trained deep learning architecture for brain tumor segmentation. The proposed architecture is similar to standard UNET except the encoder part is pre-trained VGG16, which consists of 13 convolutional layers, five pooling layers, and three fully connected layers; therefore, the

[1] https://www.who.int/news-room/fact-sheets/detail/cancer

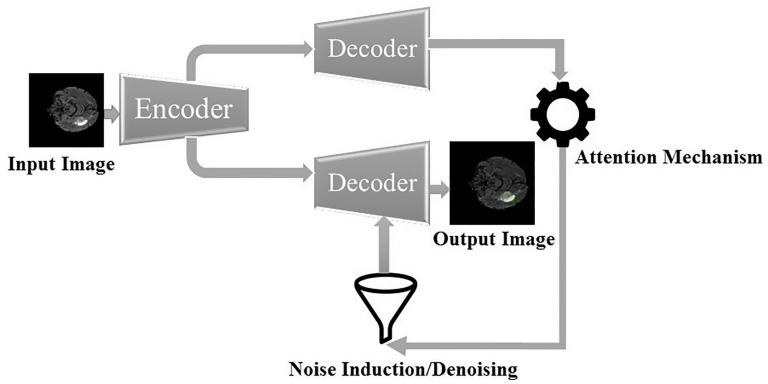


FIGURE 1 | Overview of the sequential framework.

decoder also has 13 convolutional layers, five upsampling layers, and three fully connected layers. The validation of the proposed method was performed on The Cancer Imaging Archive (TCIA); an evaluation was performed on different metrics such as accuracy, DSC, and intersection over union (IoU). The proposed method's accuracy, DSC, and intersection over Union (IoU) are 0.998, 0.93, and 0.83, respectively.

Alqazzaz et al. (16) trained a variant of Segnet for brain tumor segmentation. Specifically, four different SegNets were trained on T1, Flair, T1ce, and T2-weighted images. Four SegNets are then combined, and feature extraction is performed. Finally, a Decision Tree is applied to the extracted features to generate the predicted segmentation mask of the tumor region. The datasets used in the experimentation were BRATS'17, and the evaluation metrics were precision, recall, and F-measure. They achieved an F-measure of 0.85, 0.81, and 0.79 on the whole, enhancing, and core tumors, respectively.

Karak et al. (17) proposed an encoder-decoder deep neural network for multi-class brain tumor segmentation. The proposed architecture is called TwoPath U-NET because it learns both local and global features by using local and global feature extraction paths in the down-sampling path of the deep neural network. The validation of the proposed method was performed on BRATS'19, and DSC was the evaluation metric used in the experimentation. The DSC of the proposed method was 0.76, 0.64, and 0.58 for the whole, enhancing, and core tumors, respectively.

Silva et al. (18) presented a deep multicascade fully connected neural network for brain tumor segmentation. Specifically, the proposed architecture is composed of three deep layer aggregation neural networks, i.e., basic convolutional block, convolutional block, and aggregation block. The proposed method was evaluated using BRATS'20 datasets, and the evaluation metrics used in the experimentation were DSC and Hausdorff distance. The DSC was 0.88, 0.82, and 0.79 for the whole, enhancing, and core tumors, respectively, while the Hausdorff distance was 5.32, 22.32, and 20.44 mm for whole, core, and enhanced tumors, respectively. Murugesan et al. (19) presented a multidimensional and multiresolution ensemble neural network for brain tumor segmentation and trained a traditional machine learning model for survival prediction. Specifically, an ensemble of pre-trained neural networks such as DenseNET-169, SERESNEXT-101, and SENet-154 was utilized to segment whole, core, and enhanced tumors. The segmentation map was then produced by combining the segmentation of an ensemble of pretrained deep neural networks. The datasets used in the experimentation were BRATS'19 and achieved a DSC of 0.89, 0.78, and 0.779 for the whole, core, and enhancing tumors, respectively, and survival prediction accuracy was 34%.

Specifically, the proposed architecture extracts the multistake information by combining the 3D convolutional neural network information in the residual inception block and utilizing hyperdense inception 3D UNET. Qamar et al. (20) trained a3D UNET to classify the whole, enhancing, and core tumor classes. The validation of the proposed method was performed on BRATS 2020 datasets and achieved a DSC of 0.79, 0.87, and 0.83 for enhancing, whole, and core tumors, respectively. Zhao et al. (21) performed integration of a fully connected neural network (FCNN) and conditional random field (CRF) for brain tumor segmentation. After basic preprocessing, FCNN was applied to predict the class label probability of each pixel then the prediction output was passed to the CRF-RNN for global optimization and spatial consistency of segmentation results. The validation of the proposed architecture was performed on BRATS'13, BRATS'15, and BRATS'16 datasets. The DSC of the proposed method was between 0.79 and 0.85 for the whole tumor, 0.65 and 75 for the core tumor, and 0.75 and 0.80 for the enhancing tumor, respectively.

Zhu et al. (22) presented a holistically nested neural network for brain tumor segmentation. The multiscale and multilevel hierarchical features of the brain MRI were learned by the holistically nested neural network, which is the extension of CNN to generate the prediction map of test images of brain MRI. The evaluation of the proposed method was performed on BRATS'13 datasets, and the evaluation metrics used in the experimentation were DSC and sensitivity. The results show that the presented method outperformed the previous method

with DSC and sensitivity of 0.83 and 0.85, respectively. Cui et al. (23) proposed a cascaded convolutional neural network for brain tumor segmentation. The proposed architecture is composed of two subnetworks. The first network is called the tumor localization network (TCN), and it is used to detect the tumor region from an MRI scan. The second network was called as intratumor classification network (ITCN), which was used to label the defined tumorous region into subregions. The proposed architecture was validated on BRATS'15 datasets, and DSC, sensitivity, and positive predicted value (PPV) are the evaluation metrics used in the experimentation, achieving a DSC of 0.89, 077, and 0.80 for the whole, core, and enhancing tumors, respectively.

Hoseini et al. (24) proposed a Deep Convolutional Neural Network (DCNN) for brain tumor segmentation. The proposed architecture is composed of two parts. The architecture part was used to design the network model and was composed of five convolutional layers, one fully connected layer, and a softmax layer, while the second was used to optimize the learning parameters of the network during the training phase. The evaluation metric used in the experimentation was DSC and achieved a DSC of 0.9, 0.85, and 0.84 for the whole, core, and enhancing tumors on BRATS'16 datasets. Wang et al. (25) presented a Fully Connected Convolutional Neural Network for individual segmentation of WT, ET, and TC, respectively. The first step is the segmentation of WT by proposing WNet. The segmented output is used to segment ET by proposing ENet. The output was then used to segment TC by proposing CNet. The presented methods were validated on BRATS'17 and achieved a DSC of 0.78, 0.90, and 0.83 on enhancing, whole, and core tumors, respectively.

Kamnitsas et al. (26) proposed an Ensemble of Multiple Model and Architecture (EMMA) for efficient brain tumor segmentation to determine the influence of metaparameters on individual models while reducing the risk of overfitting. Specifically, the proposed architecture is the ensemble of two 3D multiscale CNNs called DeepMedic, Fully Connected Network (FCN), and UNET. The validation of the proposed architecture was performed on BRATS'17, consisting of 215 high-grade glioma (HGG) images and 75 low-grade glioma (LGG) images. The DSC of 72.9, 88.6, and 78.5 were obtained for enhancing, whole, and core tumors, respectively. Colmeiro et al. (27) proposed a fast and straightforward 3D UNET method for automatic segmentation of brain tumors. Specifically, a two-stage 3D Deep Convolutional Network was proposed. In the first step, the whole tumor was segmented from the low-resolution volume, and then the second step was the segmented delicate tissues. The proposed method was evaluated on BRATS 2017 datasets, and DSC sensitivity, specificity, and Hausdorff distance were evaluation metrics used in the experimentation. The maximum DSC, sensitivity, specificity, and Hausdorff distance mean on unseen datasets were 0.86, 0.997, and 14.0 for the whole, enhanced, and core regions, respectively. Myronenko et al. (28) proposed an automatic 3D brain tumor semantic segmentation using encoder–decoder architecture from MRI. Specifically, a variational autoencoder was used to construct input images, and a decoder was used to impose constraints on its layer. The encoder is a pre-trained ResNet, which is followed by the respective decoder. The proposed method was evaluated on BRATS 2018, and the maximum DSC values for the enhancing, whole, and core tumors were 0.82, 0.91, and 0.86, respectively. Similarly, Hausdorff distances for the enhancing, whole, and core tumors were 8.0, 10.0, and 5.9, respectively.

Hamghalam et al. (29, 30) proposed generative techniques for brain tumor segmentation. Specifically, the proposed technique uses the Cycle-Gan as an attention mechanism for improving the contrast of the tumorous region. A model is then applied to the contrasted image for final segmentation. The performance of the proposed architecture is validated on BRATS18 datasets and achieved a DSC of 0.8%, 0.6%, and 0.5%, respectively, on the WT, TC, and ET.

Most researchers focus on improving the result of segmentation while ignoring the efficiency of the task. Therefore, the prime thing in any machine learning task is to extract the minimum but valuable features. In order to tackle this problem, we will use an attention-based mechanism that will extract the useful features from the whole MRI and further utilize it for segmentation. Similarly, to reduce the algorithmic and computational complexity of the task, we will use transfer learning compared to training complete NN from scratch. This technique helps us improve the performance of the segmentation while preserving the accuracy of the task.

METHODOLOGY

The framework for the segmentation of brain tumors is presented in **Figure 2**. Like the standard UNET (31) system, the proposed system contains the encoder and decoder parts. The encoder part is standard VGG19 (32)[2] with the convolutional unit, which is used to extricate features from MR images, while the decoder part utilizes the output of the encoder VGG19 with an attention mechanism to segment the image by upsampling the element maps. The figure likewise shows that different colors address various hidden layers. The convolutional layers are represented by the blue color, the pooling layer by the yellow color, the upsampling layers by the pink color, and lastly, the SoftMax layer by the red color. Input of size (224 × 224) is given at the encoder part. In the wake of going through various hidden layers, a binary segmented image is received as the first output of the decoder part. Additionally, an attention mechanism and an overfitting reduction mechanism are applied to extract the specified segmented image and the final multiclass segmentation of the tumorous region. At first, there are 144 million boundaries of VGG19 that are diminished to 36.1 million boundaries by disposing of fully connected layers. Output is passed to the SoftMax layer to group pixels autonomously into "K" classes. K is a number of classes and is equivalent to four since we have classes with (0, 1, 2, 3) marks. 0 for nontumorous, 1 for CT, 2 for WT, and 3 for ET.

The encoder network performs convolution with a filter to produce feature maps. The rectified linear unit (ReLU) then transforms the nonlinear output into a linear output. The output

[2]https://neurohive.io/en/popular-networks/vgg19/

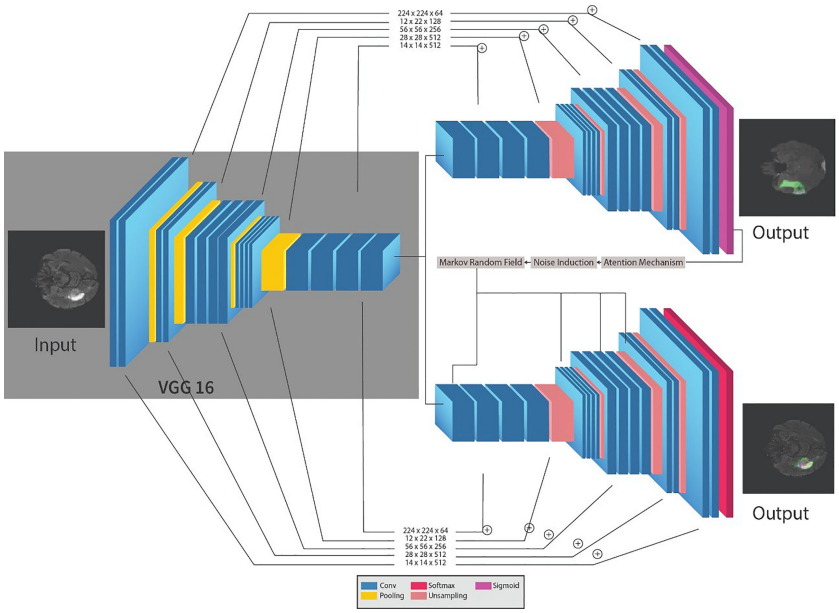


FIGURE 2 | Proposed sequential framework for segmentation of brain tumor.

is then batch normalized. Also, the max-pooling layer with 22 windows and stride 2 is performed to reduce the dimension of the image. We discard the FC layer to reduce the parameters learned from 144 to 36.1 million. The decoder, which corresponds to the initial encoder (front to the input image), generates a multichannel feature map. Similarly, the input feature map is upsampled by the decoder network. The following process represents the high-dimensional feature at the output of the last decoder. It is fed to a SoftMax classifier, which is trainable, and its output is a K channel image of probabilities. K represents the number of classes.

Transfer Learning

Transfer learning moves the information gained by resolving one dilemma to another related issue. A model built and trained in machine learning for one dataset or recognition issue is repeatedly used as the preliminary step for the following database (33). It is difficult in practice to train a network from scratch using random initialization due to data limitations. Therefore, using pre-trained network weights as initializations or a fixed extractor of features helps solve most problems. Since pretrained models are computationally costly, it can also take a couple of days or even weeks to learn correctly from the beginning, and it also helps accelerate the training cycle, which helps in solving most of the problems at hand (25).

VGG19

CNN is a feed-forward ANN and inspired by the human visual cortex. The neurons of CNN followed a similar connectivity pattern (2). The visual cortex is the area of the cerebral cortex that is responsible for processing visual information. Visual input is provided from the eyes and reaches the visual cortex *via* the lateral geniculate nucleus in the thalamus. The state-of-the-art artificial neural network is employed in image processing and machine vision tasks such as segmentation, classification, and recognition. The standard CNN comprises input layers, hidden layers, and output layers. The hidden layers are usually convolutional, pooling, and fully connected. The working of CNN is simple by comparing the pixels of the images (34). The pixels are also called features of the image. So, in short, CNN works by learning the features of the images, and CNN learns these features by passing through different hidden layers, and hidden layers in CNN are usually filters of different sizes. VGG19 is the commonly used CNN, composed of nineteen layers. Out of these 19 layers, 16 are convolutional layers, 5 are max-pool layers, 3 are completely connected layers, and 1 is a SoftMax layer. The architecture of VGG19 (15) is simple and follows a six-step process;

- First, the image is fed into the architecture as input; usually, the shape (224, 224, 3) is provided.
- The kernel of size (3, 3) was then applied to discover the underlying patterns of the image.
- Padding was used to preserve image resolution.
- Pooling was applied to reduce the dimension of the image.
- The output of the layers is usually linear. Therefore, a fully connected layer was applied to transform the linear output into the nonlinear output.
- Finally, the SoftMax layer is applied to predict the probability distribution of the multiple classes.

The training of VGG19 from scratch is a tedious and complex task; therefore, nowadays, a pre-trained VGG19 is often used. A

pre-trained VGG19 is usually trained on larger datasets, i.e., ImageNet; thus, learning new and complex patterns becomes efficient and straightforward. The architecture of VGG19 is shown in **Figure 3**.

Attention Mechanism

The attention mechanism is a model for medical image examination that naturally figures out how to focus on the targeted image of changing shapes and sizes (35–40). An attention mechanism helps the decoder focus on the area of interest. Subsequently, with the attention mechanism, we will classify the pixel by the hidden state of the decoder. Hence, we partition the image into *n* parts; then, at that point, at the *i*th area of the image, we utilize the hidden region of the decoder part. The hidden region is then utilized as the setting to choose the interest area of the image. The *zi* is the output of the attention mechanism. Models prepared with attention mechanisms certainly figure out how to smother unessential regions in an input image while featuring remarkable features helpful for a particular task, which empowers us to eliminate the need to utilize express outside tissue/organ localization modules when using CNN. **Table 1** describes the important symbols and variables used in the equation. The mathematical representation of the attention mechanism with overfitting reduction is as follows.

Algorithm 1: Pseudocode for the proposed approach.

1. $i_{jt}=f_A(o_{t-1},e_j)$
2. $f_A = V_A^T * tanh(U_A * i_j + W_A * o_t)$
3. $C_t = \sum_{j=1}^{T} alpha_{ji} * e_j \quad such \quad that \sum_{j=1}^{T_x} \alpha_{jt} = 1$
4. where $\alpha_{ij} \geqq 0$
5. $o_t=CNN(s_{t=1}.[e(y_{t=1})\cdot c_t])$
6. $A(x,y)=o_t+N(x,y)$
7 where $N(x,y) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(z-)}{2\pi^2}}$
8. select the pixels with the lowest energy $-E_{total}(x,y) = \sum_{i,j}^{N} x_{i,j} \cdot x_{i+1,j+1} - \eta \sum_{i,j=1}^{N} x_{i,j} \cdot y_{i,j}$

- In the algorithm, $i_{j\ t}$ represents the location of *j*th pixel in the input image
- o_t is the is the output of the decoder part
- e_j is the current state of the encoder part

For a linear transformation of input, a simple feed-forward neural network f_A is applied and then a nonlinearity (*tanh*) and transformation function V_A^T (scalar quantity). The fourth line of the pseudocode shows that, as now, we know the input, we need to feed the weighted sum combination of input to the decoder. In the following line,

- $e(y_{t-1})$ is the previously predicted label of binary classification
- C_t is the context vector, i.e., the weighted sum of the input

After that, noise $N(x,y)$ is combined with the output of decoder o_t where $N(x,y)$ is the Gaussian noise function. Finally, we apply the MRF by looping over the pixels of image $A(x,y)$ and computing the energies of the current pixels of $A(x,y)$ by applying the formula given in the eighth line.

Markov Random Field

The initial outcomes of the proposed models result in overfitting. Therefore, we introduce noise in the generated image of the decoder. Specifically, we used the Gaussian noise function to introduce noise. We add 20% noise to the image. The MRF algorithm is then applied to denoising the resultant image. Compared to the Bayesian network, the connection between nodes in MRF are undirected and cyclic. The MRF is defined in terms of energy. When pixels of both images match, we say that the energy of both images is low and high otherwise. The algorithm moves on to the pixels, either moving through them in some predetermined order or choosing a random pixel at each step, running through the set of pixels until their values stop changing. The equation in line 15 represents the energy of the where η and ζ

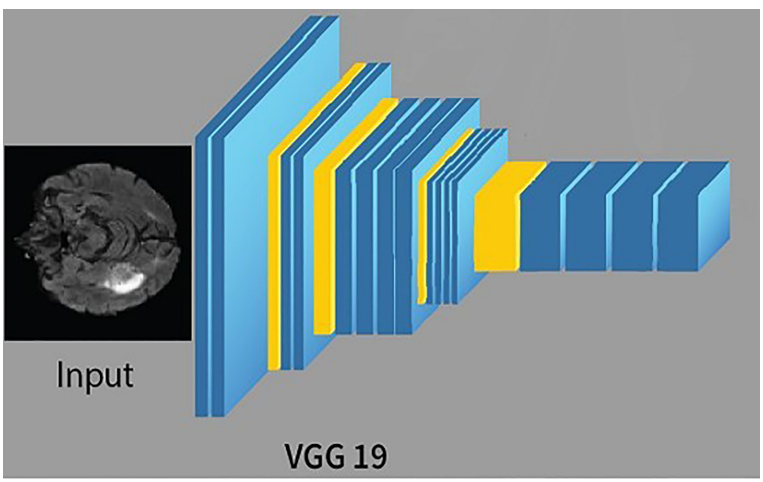


FIGURE 3 | The architecture of VGG19.

TABLE 1 | Symbols with description.

Serial number	Symbol	Description
1.	f_A	Feed-forward neural network
2.	V_A^T	Transformation function
3.	W_A	Attention function
4.	C_t	Context vector
5.	η	Learning rate
6.	$\eth$	Standard deviation

are positive constant, energy of the "output pixels" and here we set η=15, $\eth$=1.5 .

A reliable and efficient variant of a pre-trained neural network, i.e., an attention-based recurrent convolutional neural network for brain tumor segmentation, is proposed in the proposed framework. Specifically, the encoder part of the UNET is a pre-trained recurrent VGG19 network followed by the adjacent recurrent decoder part with an attention gate.

RESULTS

Evaluation Metrics

In this section, the qualitative and quantitative results of the proposed framework are presented along with evaluation metrics.

Generally, two types of segmentation of brain tumors are used, i.e., manual segmentation and automatic segmentation. Firstly, the manual segmentation is performed by MRI experts, which is a tedious and complex task, but accurate, while the accurate and straightforward software does automatic segmentation due to developments in artificial intelligence. It is also worth mentioning that the MRI experts first label the datasets used for automatic segmentation (41). The evaluation metrics used for brain tumor segmentation are DSC, accuracy, sensitivity, and precision (42). Similarly, true negative (TN) refers to the negative tuple correctly labeled by the classifier. False negative (FN) refers to the classifier's tuple to the positive tuple incorrectly labeled. Similarly, false positive (FP) refers to the negative tuple incorrectly labeled by the classifier.

DSC is the commonly used evaluation metric for image segmentation and segmentation of brain tumors. DSC is the measure of overlapping area between two images (23). For example, in **Figure 4**, there are two circle images labeled "A" and "B." The DSC of the figure is then illustrated in **Equation 1**, which shows that DSC is equal to two times the overlapped area in the general area of the image element of both images. It can also be illustrated as two times the true positive (TP) divided by total TP, FP, and FN as represented in **Equation 2**.

$$\mathrm{DSC} = 2|A \cap B|/(|A| + |B|) \tag{1}$$

$$\mathrm{DSC} = 2TP/2TP + FN + FP \tag{2}$$

Qualitative Results

For the evaluation of the segmentation task, the BRATS'20 (43) was used, consisting of 371 image files, and each file is composed of five subfiles, out of which four files are MRI modalities of the individual patients, and one file is the target mask of the individual patient. T1, T2, T2*, and attenuated inversion recovery (FLAIR)-weighted mages are the most common modalities of MRI utilized in this dataset. A different clinical protocol was acquired for each modality, and multiple scanners from several institutions and each modality have been segmented manually by one to four raters. All the modalities are available as NIFTI files with the extension (.nii.gz). A NIFTI file is the most common file format for neuroimaging. The available datasets are imbalanced; therefore, in the data preprocessing step, a patch-wise training procedure is applied (44).

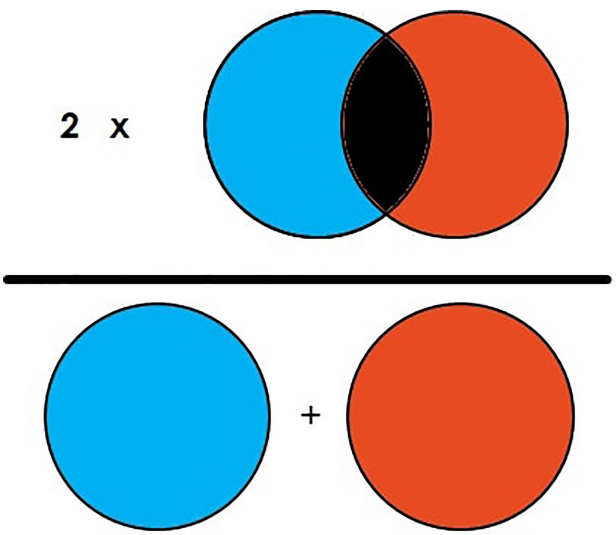


FIGURE 4 | Illustration of Dice Similarity Coefficient (DSC).

Figure 5 shows the segmentation results of the proposed model on the BRATS'20 dataset. The first column is the tumor segmentation of all tumor classes, followed by the individual segmentation of core, whole, and enhancing tumors in columns fourth, fifth, and sixth, respectively.

Quantitative Results

The quantitative results of the proposed model are presented in **Table 2** with a sensitivity, specificity, accuracy, and precision of 0.98, 0.981, 0.99, and 0.993, respectively. Similarly, the comparison of the achieved results with the primary method is presented in **Table 3**, which reveals that the proposed framework outperformed the state-of-the-art methods. The comparison is performed based on the DSC score of ET, WT, and TC, respectively.

TABLE 2 | Quantitative results of the proposed model.

Metrics	Results
Sensitivity	0.98
Specificity	0.981
Precision	0.993
Accuracy	0.99
DSC of ET	0.861
DSC of WT	0.90
DSC of TC	0.83

TABLE 3 | Comparison of results of brain tumor segmentation.

Methods	ET	WT	TC
Ghaffari et al. (45)	0.78	0.90	0.82
Ballester et al. (46)	0.67	0.85	0.78
Colman et al. (27)	0.75	0.86	0.79
Proposed method	0.83	0.90	0.86

CONCLUSION

In conclusion, a pre-trained VGG19 neural network with an attention mechanism and an image processing technique is trained for brain tumor segmentation. Applying the attention mechanism aims to suppress irrelevant regions in an input image while highlighting essential features useful for a specific task. The

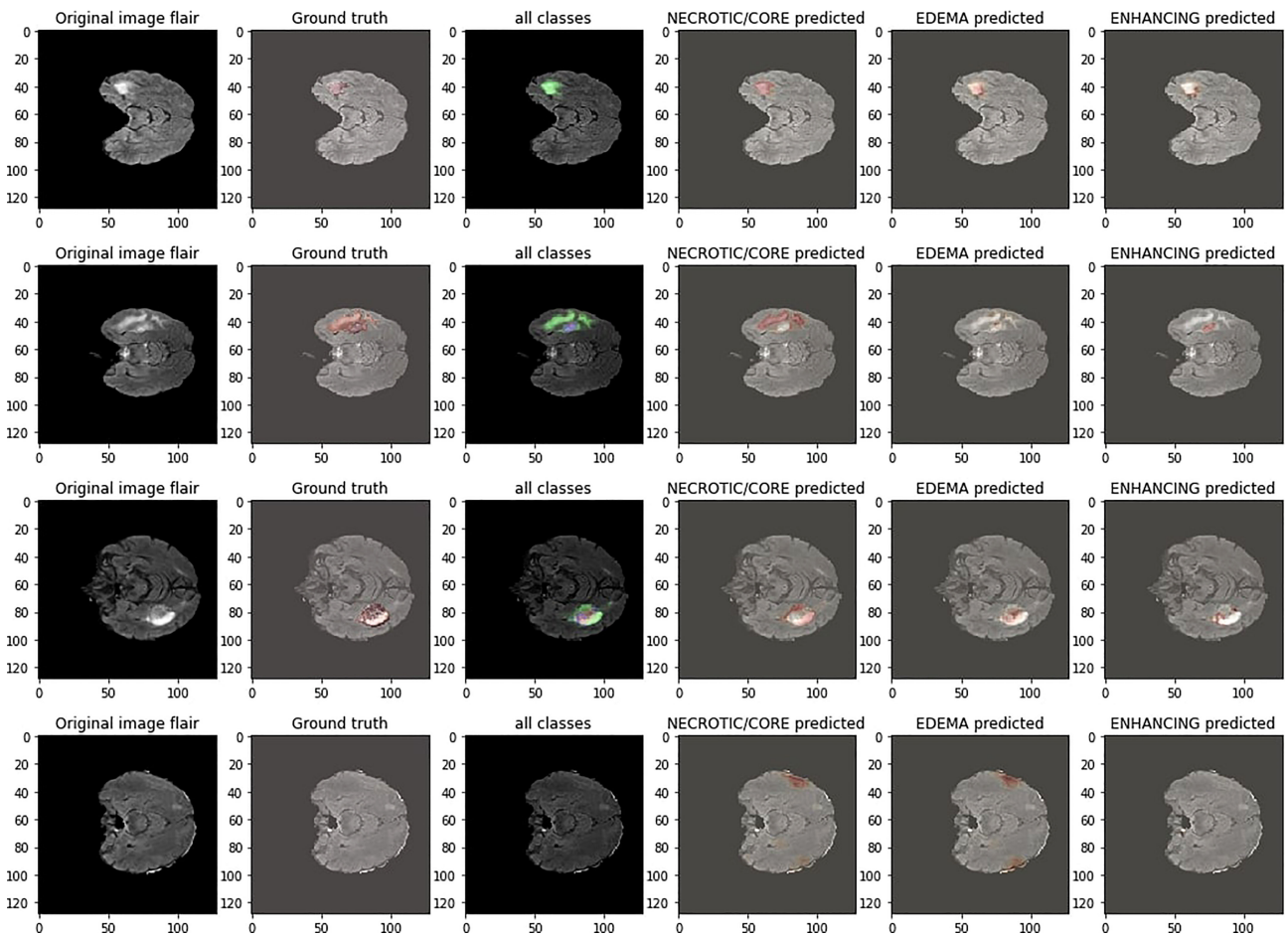


FIGURE 5 | Qualitative results of the proposed framework.

proposed model's evaluation is carried out on BRATS20, and evaluation metrics used in the segmentation method are accuracy, sensitivity, specificity, precision, and DSC. The obtained results show that the proposed model produces more accurate and better outputs than the previous method for enhancing, whole, and core tumors with dice similarity coefficient scores of 0.83, 0.9, and 0.86, respectively. The proposed segmentation methods enable the efficient and effective diagnosis of brain tumors. In the future, an ensemble attention mechanism will be proposed to extract the more important features and increase the segmentation results.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

This research specifies below the individual contributions. Conceptualization: TA and AN. Data curation: AN. Formal analysis: AJ. Funding acquisition: TA. Investigation: AJ and AR. Methodology: AJ. Project administration: C-LC, RA, and TA. Resources: RA, TG, and TA. Software: AJ. Supervision: C-LC, AJ, C-LC, and C-MW. Validation: AN, RA, and TG. Visualization: TG and C-LC. Writing—review and editing: AR, C-MW, TA, AR, and C-MW.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (No. 51808474) and the Ministry of Science and Technology in Taiwan (No. MOST 110-2218-E-305-MBK and MOST 110-2410-H-324-004-MY2). We also want to thank the "Gulf University for Science and Technology" for supporting this research with grant number "223565." This paper is an advancement of the article "VGG-UNET for Brain Tumor Segmentation and Ensemble Model for Survival Prediction," which was already published at the ICRAI conference (47).

REFERENCES

1. Rizwan M, Shabbir A, Javed AR, Shabbr M, Baker T, Obe DAJ. Brain Tumor and Glioma Grade Classification Using Gaussian Convolutional Neural Network. *IEEE Access* (2022) 10:29731–40. doi: 10.1109/ACCESS.2022.3153108
2. Abiwinanda , N, Muhammad Tafwida HS. H, Astri H, and Tati RM. "Brain Tumor Classification Using Convolutional Neural Network." In World Congress on Medical Physics and Biomedical Engineering 2018. Singapore: Springer (2019). p. 183-9.
3. Forst DA, Nahed BV, Loeffler JS, Batchelor TT. Low-Grade Gliomas. *Oncol* (2014) 19:403–13. doi: 10.1634/theoncologist.2013-0345
4. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (Brats). *IEEE Trans Med Imaging* (2014) 34:1993–2024. doi: 10.1109/TMI.2014.2377694
5. Lather M, Singh P. Investigating Brain Tumor Segmentation and Detection Techniques. *Proc Comput Sci* (2020) 167:121–30. doi: 10.1016/j.procs.2020.03.189
6. Hussain S, Anwar SM, Majid M. Segmentation of Glioma Tumors in Brain Using Deep Convolutional Neural Network. *Neurocomputing* (2018) 282:248–61. doi: 10.1016/j.neucom.2017.12.032
7. Pereira S, Pinto A, Alves V, Silva CA. Brain Tumor Segmentation Using Convolutional Neural Networks in Mri Images. *IEEE Trans Med Imaging* (2016) 35:1240–51. doi: 10.1109/TMI.2016.2538465
8. Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, et al. Deep Learning and Process Understanding for Data-Driven Earth System Science. *Nature* (2019) 566:195–204. doi: 10.1038/s41586-019-0912-1
9. Yu K, Tan L, Lin L, Cheng X, Yi Z, Sato T. Deep-Learning-Empowered Breast Cancer Auxiliary Diagnosis for 5gb Remote E-Health. *IEEE Wireless Commun* (2021) 28:54–61. doi: 10.1109/MWC.001.2000374
10. Sun L, Zhang S, Chen H, Luo L. Brain Tumor Segmentation and Survival Prediction Using Multimodal Mri Scans With Deep Learning. *Front Neurosci* (2019) 810. doi: 10.3389/fnins.2019.00810
11. Gadekallu TR, Alazab M, Kaluri R, Maddikunta PKR, Bhattacharya S, Lakshmanna K, et al. Hand Gesture Classification Using a Novel Cnn-Crow Search Algorithm. *Complex Intell Syst* (2021) 7:1855–68. doi: 10.1007/s40747-021-00324-x
12. Erden B, Gamboa N, Wood S. 3d Convolutional Neural Network for Brain Tumor Segmentation. *Comput Sci Stanf Univ USA Tech Rep* (2017).
13. Pei L, Vidyaratne L, Rahman MM, Iftekharuddin KM. Context Aware Deep Learning for Brain Tumor Segmentation, Subtype Classification, and Survival Prediction Using Radiology Images. *Sci Rep* (2020) 10:1–11. doi: 10.1038/s41598-020-74419-9
14. Zeineldin RA, Karar ME, Coburger J, Wirtz CR, Burgert O. Deepseg: Deep Neural Network Framework for Automatic Brain Tumor Segmentation Using Magnetic Resonance Flair Images. *Int J Comput Assist Radiol Surg* (2020) 15:909–20. doi: 10.1007/s11548-020-02186-z
15. Ghosh S, Chaki A, Santosh K. Improved U-Net Architecture With Vgg-16 for Brain Tumor Segmentation. *Phys Eng Sci Med* (2021) 44:703–12. doi: 10.1007/s13246-021-01019-w
16. Alqazzaz S, Sun X, Yang X, Nokes L. Automated Brain Tumor Segmentation on Multi-Modal Mr Image Using Segnet. *Comput Visual Med* (2019) 5:209–19. doi: 10.1007/s41095-019-0139-y
17. Crimi A, Bakas S. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction With MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I*, Vol. *12658 (Spring Nat)* (2021). doi: 10.1007/978-3-030-72087-2
18. Silva CA, Pinto A, Pereira S, Lopes A. Multi-Stage Deep Layer Aggregation for Brain Tumor Segmentation. *Int MICCAI Brainles Worksh (Spring)* (2020) 179–88. doi: 10.1007/978-3-030-72087-2_16
19. Murugesan GK, Nalawade S, Ganesh C, Wagner B, Yu FF, Fei B, et al. Multidimensional and Multiresolution Ensemble Networks for Brain Tumor Segmentation. *Int MICCAI Brainles Worksh (Spring)* (2019) 148–57. doi: 10.1101/760124
20. Qamar S, Ahmad P, Shen L. Hi-Net: Hyperdense Inception 3d Unet for Brain Tumor Segmentation. *Int MICCAI Brainles Worksh (Spring)* (2020) 50–7.
21. Zhao X, Wu Y, Song G, Li Z, Zhang Y, Fan Y. A Deep Learning Model Integrating Fcnns and Crfs for Brain Tumor Segmentation. *Med Imag Anal* (2018) 43:98–111. doi: 10.1016/j.media.2017.10.002
22. Zhuge Y, Krauze AV, Ning H, Cheng JY, Arora BC, Camphausen K, et al. Brain Tumor Segmentation Using Holistically Nested Neural Networks in Mri Images. *Med Phys* (2017) 44:5234–43. doi: 10.1002/mp.12481
23. Cui S, Mao L, Jiang J, Liu C, Xiong S. Automatic Semantic Segmentation of Brain Gliomas From Mri Images Using a Deep Cascaded Neural Network. *J Healthcare Eng* (2018) 2018:1–15. doi: 10.1155/2018/4940593
24. Hoseini F, Shahbahrami A, Bayat P. An Efficient Implementation of Deep Convolutional Neural Networks for Mri Segmentation. *J Digit Imaging* (2018) 31:738–47. doi: 10.1007/s10278-018-0062-2

25. Wang G, Zhang G, Choi KS, Lam KM, Lu J. Output Based Transfer Learning With Least Squares Support Vector Machine and its Application in Bladder Cancer Prognosis. *Neurocomputing* (2020) 387:279–92. doi: 10.1016/j.neucom.2019.11.010
26. Kamnitsas K, Bai W, Ferrante E, McDonagh S, Sinclair M, Pawlowski N, et al. Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation. *Int MICCAI Brainles Worksh (Spring)* (2017) 450–62.
27. Colman J, Zhang L, Duan W, Ye X. Dr-Unet104 for Multimodal Mri Brain Tumor Segmentation. *Int MICCAI Brainles Worksh (Spring)* (2020) 410–9.
28. Myronenko A. 3d Mri Brain Tumor Segmentation Using Autoencoder Regularization. *Int MICCAI Brainles Worksh (Spring)* (2018) 311–20.
29. Hamghalam M, Wang T, Lei B. High Tissue Contrast Image Synthesis *via* Multistage Attention-Gan: Application to Segmenting Brain Mr Scans. *Neural Networks* (2020) 132:43–52. doi: 10.1016/j.neunet.2020.08.014
30. Hamghalam M, Lei B, Wang T. High Tissue Contrast Mri Synthesis Using Multi-Stage Attention-Gan for Segmentation. *Proc AAAI Conf Artif Intell* (2020) 34:4067–74. doi: 10.1609/aaai.v34i04.5825
31. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Int Conf Med Imag Comput Comput-assist Interven (Spring)* (2015) 234–41. doi: 10.1007/978-3-319-24574-4_28
32. Pravitasari AA, Iriawan N, Almuhayar M, Azmi T, Fithriasari K, Purnami SW, et al. Unet-Vgg16 With Transfer Learning for Mri-Based Brain Tumor Segmentation. *Telkomnika* (2020) 18:1310–8. doi: 10.12928/telkomnika.v18i3.14753
33. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A Comprehensive Survey on Transfer Learning. *Proc IEEE* (2020) 109:43–76. doi: 10.1109/JPROC.2020.3004555
34. Toğaçar M, Ergen B, Cömert Z. Brainmrnet: Brain Tumor Detection Using Magnetic Resonance Images With a Novel Convolutional Neural Network Model. *Med Hypotheses* (2020) 134:109531. doi: 10.1016/j.mehy.2019.109531
35. Sun Y, Liu J, Yu K, Alazab M, Lin K. Pmrss: Privacy-Preserving Medical Record Searching Scheme for Intelligent Diagnosis in Iot Healthcare. *IEEE Trans Ind Inf* (2021) 18:1981–90. doi: 10.1109/TII.2021.3070544
36. Nagarajan G, Babu LD. Missing Data Imputation on Biomedical Data Using Deeply Learned Clustering and L2 Regularized Regression Based on Symmetric Uncertainty. *Artif Intell Med* (2022) 123:102214. doi: 10.1016/j.artmed.2021.102214
37. Nagarajan G, Babu LD. A Hybrid of Whale Optimization and Late Acceptance Hill Climbing Based Imputation to Enhance Classification Performance in Electronic Health Records. *J Biomed Inf* (2019) 94:103190. doi: 10.1016/j.jbi.2019.103190
38. Noori M, Bahri A, Mohammadi K. Attention-Guided Version of 2d Unet for Automatic Brain Tumor Segmentation. In: *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*. Rawalpindi, Islamabad: IEEE (2019). p. 269–75.
39. Pandya S, Gadekallu TR, Reddy PK, Wang W, Alazab M. Infusedheart: A Novel Knowledge-Infused Learning Framework for Diagnosis of Cardiovascular Events. *IEEE Trans Comput Soc Syst* (2022). doi: 10.1109/TCSS.2022.3151643
40. Arikumar K, Prathiba SB, Alazab M, Gadekallu TR, Pandya S, Khan JM, et al. Fl-Pmi: Federated Learning-Based Person Movement Identification Through Wearable Devices in Smart Healthcare Systems. *Sensors* (2022) 22:1377. doi: 10.3390/s22041377
41. Hamwood J, Schmutz B, Collins MJ, Allenby MC, Alonso-Caneiro D. A Deep Learning Method for Automatic Segmentation of the Bony Orbit in Mri and Ct Images. *Sci Rep* (2021) 11:1–12. doi: 10.1038/s41598-021-93227-3
42. Bahadure NB, Ray AK, Thethi HP. Comparative Approach of Mri-Based Brain Tumor Segmentation and Classification Using Genetic Algorithm. *J Digit Imaging* (2018) 31:477–89. doi: 10.1007/s10278-018-0050-6
43. Alex V, Safwan M, Krishnamurthi G. Automatic Segmentation and Overall Survival Prediction in Gliomas Using Fully Convolutional Neural Network and Texture Analysis. *Int MICCAI Brainles Worksh (Spring)* (2017) 216–25.
44. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain Tumor Segmentation With Deep Neural Networks. *Med Imag Anal* (2017) 35:18–31. doi: 10.1016/j.media.2016.05.004
45. Ghaffari M, Sowmya A, Oliver R. Brain Tumour Segmentation Using Cascaded 3d Densely-Connected U-Net. *ArXiv Prepr ArXiv:2009.07563* (2020). doi: 10.1007/978-3-030-72084-1_43
46. Ballestar LM, Vilaplana V. Brain Tumor Segmentation Using 3d-Cnns With Uncertainty Estimation. *ArXiv Prepr ArXiv:2009.12188* (2020).
47. Nawaz A, Akram U, Salam AA, Ali AR, Rehman AU, Zeb J. Vgg-Unet for Brain Tumor Segmentation and Ensemble Model for Survival Prediction. In: *2021 International Conference on Robotics and Automation in Industry (ICRAI)*. Mashhad, Iran: IEEE (2021). 1–6.

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 | Frontiers in Oncology

ORIGINAL RESEARCH
published: 02 June 2022
doi: 10.3389/fonc.2022.894970

AX-Unet: A Deep Learning Framework for Image Segmentation to Assist Pancreatic Tumor Diagnosis

Minqiang Yang[1], Yuhong Zhang[1], Haoning Chen[2], Wei Wang[3], Haixu Ni[4], Xinlong Chen[5], Zhuoheng Li[1] and Chengsheng Mao[6]**

[1] School of Information Science Engineering, Lanzhou University, Lanzhou, China, [2] School of Statistics and Data Science, Nankai University, Tianjin, China, [3] School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, China, [4] Department of General Surgery, First Hospital of Lanzhou University, Lanzhou, China, [5] First Clinical Medical College, Lanzhou University, Lanzhou, China, [6] Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Image segmentation plays an essential role in medical imaging analysis such as tumor boundary extraction. Recently, deep learning techniques have dramatically improved performance for image segmentation. However, an important factor preventing deep neural networks from going further is the information loss during the information propagation process. In this article, we present AX-Unet, a deep learning framework incorporating a modified atrous spatial pyramid pooling module to learn the location information and to extract multi-level contextual information to reduce information loss during downsampling. We also introduce a special group convolution operation on the feature map at each level to achieve information decoupling between channels. In addition, we propose an explicit boundary-aware loss function to tackle the blurry boundary problem. We evaluate our model on two public Pancreas-CT datasets, NIH Pancreas-CT dataset, and the pancreas part in medical segmentation decathlon (MSD) medical dataset. The experimental results validate that our model can outperform the state-of-the-art methods in pancreas CT image segmentation. By comparing the extracted feature output of our model, we find that the pancreatic region of normal people and patients with pancreatic tumors shows significant differences. This could provide a promising and reliable way to assist physicians for the screening of pancreatic tumors.

Keywords: atrous spatial pyramid pooling, boundary-aware loss function, pancreas CT, image segmentation, group convolution

OPEN ACCESS

Edited by:
Shahid Mumtaz, Instituto de Telecomunicações, Portugal

Reviewed by:
Jakub Nalepa, Silesian University of Technology, Poland
Xinhai Chen, Argonne National Laboratory (DOE), United States
Ahmedin Ahmed, Florida International University, United States

***Correspondence:**
Haixu Ni
nihx@lzu.edu.cn
Chengsheng Mao
chengsheng.mao@northwestern.edu

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

Received: *12 March 2022*
Accepted: *19 April 2022*
Published: *02 June 2022*

Citation:
Yang M, Zhang Y, Chen H, Wang W, Ni H, Chen X, Li Z and Mao C (2022) AX-Unet: A Deep Learning Framework for Image Segmentation to Assist Pancreatic Tumor Diagnosis. Front. Oncol. 12:894970. doi: 10.3389/fonc.2022.894970

1 INTRODUCTION

According to the Report on Cancer from National Cancer Institute in 2021, pancreatic cancer is the third leading cause of cancer-related death in the United States (1). The identification and analysis of pancreatic region play an important role in the diagnosis of pancreatic tumors. As an important and challenging problem in medical image analysis, pancreas is one of the most challenging organs for automated segmentation, which aim to assign semantic class labels to different tomography image regions in a data-driven learning fashion. Usually, such a learning problem encounters numerous difficulties such as severe class imbalance, background clutter with confusing distractions, and variable location and geometric features. According to statistical analysis, pancreas occupies less than 0.5% fraction of entire CT volume (2), which has a visually blurry inter-class boundary with respect to other tissues.

In this article, we combine the advantages of deepLabV series, Unet, and Xception networks to present a novel deep learning framework AX-Unet for pancreas CT image segmentation to assist physicians for the screening of pancreatic tumors. The whole AX-Unet still preserves the encoder-decoder structure of Unet. In our framework, we incorporate a modified atrous spatial pyramid pooling (ASPP) module to learn the location information. The modified ASPP can also extract multi-level contextual information to reduce information loss during downsampling. We also introduce a special group convolution operation on the feature map at each level to decouple the information between channels, achieving more complete information extraction. Finally, we employ an explicit boundary-aware loss function to tackle the blurry boundary problem. The experimental results on two public datasets validated the superiority of the proposed AX-Unet model to the states-of-the-art methods.

In summary, we propose a novel deep learning framework AX-Unet for pancreas CT image segmentation. Our framework has several advantages as follows.

1. In our framework, we introduce a special group convolution, depth-wise separable convolution, to decouple the two types of information based on the assumption that inter-channel and intra-channel information are not correlated. This design can achieve better performance with even less computation than the normal convolution.
2. We restructure the ASPP module, and the extraction and fusion of multi-level global contextual features is achieved by multi-scale dilate convolution, which enables a better handling of the large scale variance of the objects without introducing additional operations. The efficacy of the restructured ASPP is validated in our ablation studies on foreground target localization.
3. We propose a loss function that can explicitly perceive the boundary of the target and combine the focal loss and generalized dice loss (GDL) to solve the problem of category imbalance. The weighted sum of the above parts is used as our final loss function, which can explicitly perceive the boundary of the target.
4. We segment a large number of external unlabeled pancreas images using our trained model. The analysis of the imagomics features of the pancreatic region shows a significant difference between patients with pancreatic tumors and normal people ($p \leq 0.05$), which may provide a promising and reliable way to assist physicians for the screening of pancreatic tumors.

2 RELATED WORK

We are developing an artificial intelligence (AI) method for medical application in this paper. In this section, we review some previous works related to our work. We first make a brief review of AI methods in medicine. Then, we focus on the research of the AI task involved in this paper (i.e., image segmentation) and review the related methods. Finally, most related to our study, we review a few representative studies that applied AI methods to medical image segmentation, especially, pancreas segmentation, and compare them with our methods.

2.1 Artificial Intelligence in Medicine

In recent years, with the popularization of AI technology in various fields, it has also made great progresses in the medical field. The development of AI techniques has been promoting the development of medicine, from the earliest AI methods, such as expert systems (3, 4), to more advanced statistic machine learning methods, such as support vector machine (5, 6), non-negative matrix factorization (7–9), and local classification methods (10–12). Recently, the deep learning techniques that have achieved great success in computer vision and natural language processing played an important role in the development of medicine and got great development over the past few years. Xu et al. (13) used an attention-based multilevel co-occurrence graph convolutional long short-term memory (LSTM) to enhance multilevel feature learning for action recognition. Fang et al. (14) proposed a dual-channel neural network to reduce the high noise and disturbance, which generally resides in the signal collected by wearable devices, improving the accuracy of action recognition in the process of surgical assistance and patient monitoring. Mao et al. (15–17) also employed GCN and deep generative classifiers for disease identification from chest x-rays and medication recommendation. The diagnosis of tumors based on morphological features has also found some applications, applying the morphological operators get the legion part that is possible for doctors to detect accurately where the tumor is located. Hu et al. (18) proposed an emotion-aware cognitive system. A novel undisturbed mental state assessment prototype was proposed by Giddwani et al. (19). The recent pre-trained language models are also employed for disease early prediction (20) and clinical records classification (21).

2.2 Image Segmentation

For the segmentation problem, many breakthroughs have been made in recent years. He et al. (22) proposed spatial pyramid pooling (SPP) to solve the fixed input size caused by the fully

connected layer and proposed the parallel extraction of multi-level features of SPP layer, which makes different size inputs have output with fixed dimension. PSPNet (23) applied multi-level feature extraction to the field of semantic segmentation. In its design of pyramid pooling module, four different sizes of pooling are fused and then stitched by a bilinear interpolation and a 1×1 convolution. This structure is designed to aggregate contextual information from different regions, thus improving the ability to obtain global information. The DeepLabV series (24) proposed by Google later introduced ASPP in later versions, which used dilate convolution with different dilate factors to expand the receptive field without losing resolution and to fuse multi-scale context information. In addition, a 1×1 convolution and a global pooling are added in parallel. In the latest deeplabV3+ (25), the upsampling has been further refined, and better results have been achieved in boundary segmentation. In addition, in this version, Xception (26) was introduced as the backbone to perform feature extraction. This model performs channel-by-channel convolution by the assumption that the channel correlation is decoupled. Isensee et al. (27) developed nnUnet, a method that automatically configures preprocessing, network architecture, training, and post-processing for any new task, rendering state-of-the-art segmentation accessible to a broad audience by requiring neither expert knowledge nor computing resources beyond standard network training.

2.3 Medical Image Segmentation

Since Unet was proposed in 2015 (28), it has undergone many versions of evolution, and its performance has been continuously improved (29). Inspired by the successful application of Unet architecture and its variants to various medical image segmentations, Li et al. (30) proposed a novel hybrid densely connected UNet for liver and tumor segmentation. Yu et al. (31) used a salience transformation module repeatedly to convert the segmentation probability map for small organ segmentation. The above methods mainly use general segmentation approaches for medical image segmentation, ignoring domain-specific challenges. In the field of pancreatic segmentation, many methods have also been proposed. Farag et al. (32) used a convolutional neural network (CNN) model with dropout to conduct a classification on pixel level. Cai et al. (33) added a convolutional LSTM network to the output layer of CNN to compute the segmentation on two-dimensional (2D) slices of the pancreas. However, all of these methods merge the information between 2D slices of CT images for segmentation, which may miss some spatial information across slices. Man et al. (34) proposed a coarse-to-fine classifier on image patches and regions *via* CNN. Zhang et al. (35) proposed a new efficient SegNet network, which is composed of basic encoder, slim decoder, and efficient context block. Although these methods integrate spatial information to a certain extent, there is still room for improvement in boundary segmentation decisions. Ribalta Lorenzo et al. (36) proposed a two-step multi-modal Unet–based architecture with unsupervised pre-training and surface loss component for brain tumor segmentation which allows model to seamlessly benefit from all magnetic resonance modalities during the delineation. Shi et al. (37) presented a new semi-supervised segmentation model CoraNet based on uncertainty estimation and separate self-training strategy. The definition of uncertainty directly relies on the classification output without requiring any predefined boundary-aware assumption. Different from previous methods, our framework extracts more complete spatial and channel features, introduces multi-level and multi-scale feature extraction, and explicitly evaluates the segmentation loss of boundaries, achieving excellent results on multiple public datasets.

3 METHODS

In this article, we propose an improved version of Unet-based backbone network, AX-Unet, incorporating a restructured ASPP module, depth-wise convolutions, and residual blocks. We also propose a hybrid loss function that is explicitly aware of the boundary.

3.1 Architecture

As shown in **Figure 1**, our model adopts a U-shaped encoder-decoder structure, which improves the basic Unet architecture in several ways. First, we replace the normal convolutions in the encoder and decoder except the first layer with group convolution, so that in the encoding process of each level, the inter-channel and intra-channel correlation information is independently extracted (38, 39). On the basis of this structure, the overlay of adjacent slices containing the foreground is used as the input of our model; in this way, we can independently extract the detailed differences between adjacent slices, which is helpful for more accurate segmentation. Therefore, in essence, the channels should be treated differently; it is better not to map them together. Second, we have added a residual structure (40) between adjacent convolution blocks, which can reduce the semantic information loss in downsampling. Third, after the encoding stage, we set up a bottleneck layer using ASPP (41), which plays an important role in extracting multi-level contextual information to reduce information loss during downsampling. By performing convolution operations on the feature maps obtained in the encoding stage in parallel with different dilated rates, the context of the image is captured at multiple scales to obtain more accurate foreground position information (42).

Because the pancreas has a small area in computed tomography images which is flexible and changeable, traditional methods may fail to find the presence of the pancreas when receiving a challenging input. The extraction of multi-level contextual semantic information is important for small and changeable target. In the decoding phase, we restore the feature maps to the original resolution of the input image layer by layer through group deconvolution (43) and reduce the number of feature maps to 2 through 1×1 convolution.

3.2 Depth-Wise Separable Convolution

We use a special group convolution, depth-wise separable convolution, instead of the normal convolution in the encoder. The normal convolution operation is a joint mapping of channel correlation information and spatial information in the channel (44). These two kinds of information are coupled, but the two

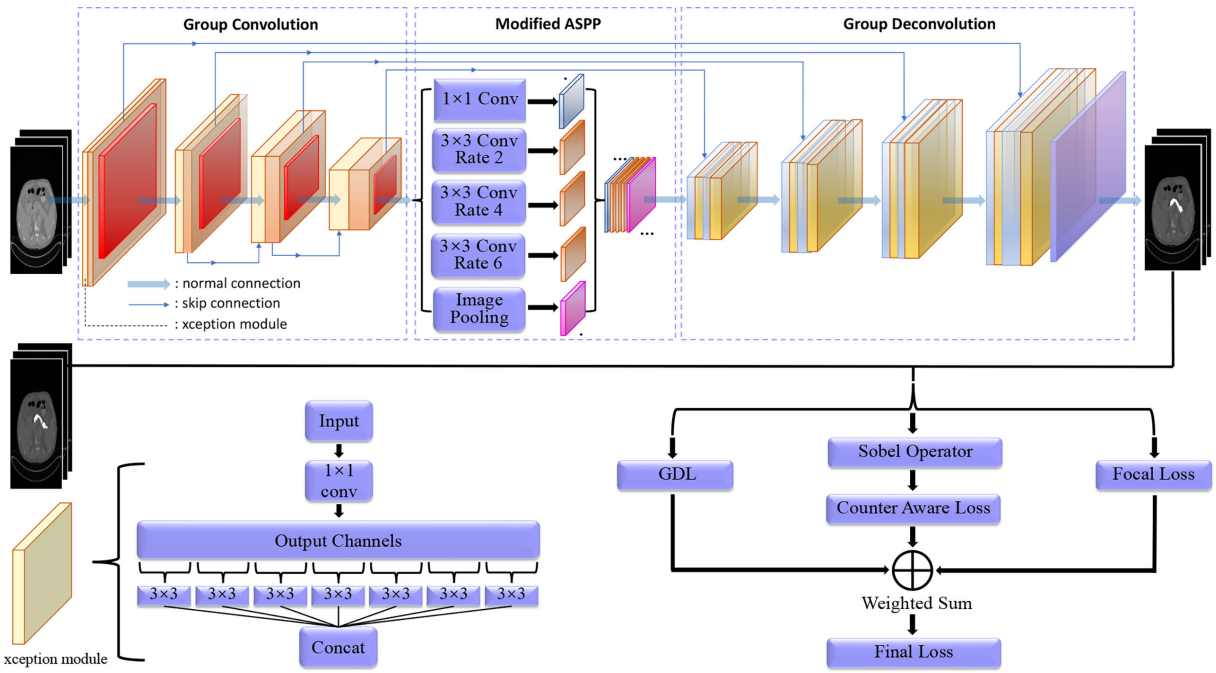


FIGURE 1 | An overview of our framework. When an original image is input, first, it goes through an encoder with four downsampling blocks consisting of a group convolution (xception) and a max pooling, yielding a small feature map; then the small feature map goes through our modified ASPP module with dilated convolution to achieve multi-scale feature parallel extraction; and then the multi-scaled feature maps go through a decoder with group decovolution operations to achieve a feature map with the same shape of the original input image for pixel-wise classification. Finally, in the test process, the output feature map implies a segmentation for evaluation; in the training process, comparing the output feature map and the original input annotation to calculated the loss by the designed hybrid loss function for model training.

correlations are decoupled in Inception by depth-wise convolution (45, 46). In the assumption of Inception, the two correlations are independent (47), mapping them separately can achieve better results. Because our input is in the form of numerous of slices, the independent mapping of information between channels is more reasonable. We use the extreme case of Inception, i.e., Xception in our framework, that is, the number of groups in the group convolution is equal to the number of input channels, which means inter-channel correlation and intra-channel spatial correlation are completely decoupled. The input feature map is linearly transformed channel by channel through a 1×1 convolution; the obtained feature map is fed to a number of 3×3 convolutions. Because the number of groups in our grouped convolution is equal to the number of input channels, all filters in this convolution process have a convolution kernel of 3×3, i.e., each channel of input feature map is only convolved by one kernel with size of 3×3×1. The outputs of these filters are stacked to construct the output feature map.

In terms of parameter comparison, assuming the number of input feature map is M, the number of output feature map is N, and the normal convolution kernel size is 3, the normal convolution has the number of parameters $N_n = 3\times3\times M\times N$, and the depth-wise separable convolution has the number of parameters from two parts, i.e., $N_g = N_{depth-wise} + N_{point-wise} = 3 \times 3\times M + 1 \times 1 \times M \times N$. Compared the depth-wise separable convolution with the normal convolution, the amount of parameters in our framework is reduced (48, 49), and the expressive ability of the network has been improved. In our framework, we use double convolutions for dowmsampling, in every double convolution block, we replace the first normal convolution with depth-wise separable structure Xception shown in **Figure 1**. Therefore, in each downsampling process, the convolution kernels with the same number of input channels are used to achieve information decoupling, and then, a normal convolution is used to double the number of feature maps. After calculation, if ordinary convolution is used completely, a total of 1,040,768 3×3 convolution kernels are needed in the entire downsampling process, whereas our improved structure only needs 700,544 3×3 convolution kernels.

3.3 ASPP Module

The pancreas images usually have blurry boundaries and are easy to be confused with surrounding soft tissues, especially, it occupies a relatively small region in a CT image with complicated background and usually less than 1.5% in a 2D image. This makes it even hard to decide whether the pancreas exists in the image. Most existing models cannot extract enough information about the position of the pancreas, which is largely related to the global context of the image. In our framework, we

use an ASPP module that contains atrous convolution to improve the information extraction ability. The ASPP module is inspired by the spatial pyramid and uses multiple parallel atrous convolution layers with different sampling rates. The context in the feature map is captured at multiple scales at the same time. In the scenario where the medical image itself does not contain complex background, noise and other information, we believe that the deep and shallow features of the medical image are all important, so the fusion of different levels of features can achieve better decision-making.

As illustrated in **Figure 1**, the ASPP module that we use mainly includes the following parts:

1. A 1×1 convolutional layer and three 3×3 atrous convolutions. When the dilated rate is close to the feature map size, filters will no longer capture the global context and will be degenerated into a simple 1×1 convolution with only the filter center working. Hence, here, we scale the dilated ratio of the original module to (2, 4, 6).
2. A global average pooling layer obtains the image-level feature, and then sends it to a 1×1 convolution layer (output with 256 channels); the output is bilinearly interpolated to be the same shape with the input.
3. The four kinds of feature maps from the above two steps are concatenated together in the channel dimension and then are sent to a 1×1 convolution for fusion to obtain a new feature map with 256 channels.

To a certain extent, the ASPP module solves the defect that the traditional Unet may have in characterizing information, can better extract multi-level position information, and has stronger characterization and learning capabilities to detect and locate the pancreas. In addition, if the dilate rate is close to or even exceeds the size of the input feature map, then it will degenerate into 1×1 convolution, and a too large dilate rate will not be conducive to pixel-level output, so we use a smaller dilate rate of (2, 4, 6).

3.4 Hybird Loss Function

Because the region to be segmented only occupies a small part of the entire image, this imbalance of foreground and background will cause sub-optimal performance (50). In addition, the pancreas as a soft tissue, the shape is variable. On the basis of the above characteristics, we proposed a hybrid loss function to update model parameters for the pancreas study tasks where category imbalance, boundary perception, and shape perception commonly exist. Our loss function consists of the following three parts.

• Generalized dice loss:

The use of ordinary dice loss is very unfavorable for small targets. The model will be overfiting (the output is all background) because once the small target has a part pixel prediction errors, it will result in large changes in dice coefficient, which will lead to dramatic changes in gradients. Therefore, GDL imposes a weight in each segmented category so as to balance the contribution of various target areas (including background) to loss.

$$Loss(GDL) = 1 - \frac{1}{m}\frac{2\sum_{j=1}^{m} w_i \sum_{i=1}^{N} y_{ij} y_{ij}^{pred}}{\sum_{j=1}^{m} w_j \sum_{i=1}^{N} \left({}_{y}ij + y_{ij}^{pred} \right)} \quad (1)$$

where w_i is valued by

$$w_i = \frac{1}{\left(\sum_{i=1}^{N} y_{ij}\right)^2} \quad (2)$$

• Focal loss:

Focal loss is designed to solve the serious imbalance in the proportion of positive and negative samples in target detection. Focal loss is optimized on the basis of the cross-entropy loss as Equation (3), where $y > 0$ reduces the loss of easy-to-classify samples ($y^{pred} \to 0$ or $y^{pred} \to 1$) and pays more attention to difficult, misclassified samples (y^{pred} around 0.5). In addition, the balance factor α is added to balance the uneven ratio of positive and negative samples. Here, we go to set α to 0.25, that is, we think negative samples are easier to distinguish.

$$Focal \quad Loss = \begin{cases} -\alpha\left(1 - y^{pred}\right)^{\gamma} \log y^{pred} & \text{for } y = 1 \\ -(1-\alpha)(y^{pred})^{\gamma} \log\left(1 - y^{pred}\right) & \text{for } y = 0 \end{cases} \quad (3)$$

• Counter-aware loss (CAL):

Pixels located at the boundary between background and foreground are so ambiguous that it is difficult to determine their labels even for experienced people. From the perspective of features, these vectors extracted from motley image pixels fall near the hyperplanes, acting as hard examples. As general networks only apply pixel-wise binary classification, target boundaries and interior pixels are processed indiscriminately using the cross-entropy loss function, so they usually predict broad outline of target objects, inferior in precision. Here, we designed a loss function based on a fixed edge extraction filter operator. The result of each iteration and the label are convolved separately. After processing, MSSS-IM (Multi-Scale-Structural Similarity Index), which measures the similarity of the image structure, is used as a loss function. This kind of explicit boundary extraction solves the problem of fuzzy boundary information and can better return the loss of boundary information.

There are many operators in edge extraction, such as Prewitt operator, Sobel operator, and Prewitt operator. They have different emphases and tendencies in boundary extraction. For example, Sobel operator detects edges according to the phenomenon of reaching extreme values at edges, which has a smoothing effect on noise. The effect of Roberts operator in detecting horizontal and vertical edges is better than that of oblique edges, and the positioning accuracy is high, but it is sensitive to noise. We choose to use the Sobel operator, which contains two sets of 3×3 matrices, which are horizontal and vertical templates, so that they can do plane convolution with our original label and segmentation output at the same time, and

then, the horizontal and vertical brightness difference approximations can be obtained, respectively.

The specific two convolution operator parameters are shown in the following matrix:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}_{3\times3}$$

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}_{3\times3}$$

Through the calculation of convolution and gradient, we get the edge of the predicted label and the original label, respectively, and then calculate loss through the cross-entropy loss function as part of the previous loss.

Our final loss function is the weighted sum of the above three loss functions as in Equation (4).w_1, w_2, and w_3 are tuned for different segmentation tasks. For all the pixels that are truly located in the pancreas region, we believe that the pixel values at the border are more indistinguishable, under this scene, we tune the weights of the three loss functions through grid search in range [0.2, 0.8] with step 0.2, try different combinations of weights, and finally find that, when a relatively large weight is given to CAL, the value of distance decreases significantly and dice score has also been improved to a certain extent, which proves the effectiveness of the perceptual boundary method we designed. However, when too large weight is given to the CAL, there will be many samples' target foreground cannot be found. We think this is caused by the fact that CAL itself cannot handle the problem of extreme class imbalance of samples, so focal loss and Dice loss are still required to a certain extent. Finally, we determined through experiments that GDL, focal loss, and CAL were given 0.2, 0.2, and 0.6, respectively, based on the validation performance.

$$Final\quad Loss = w_1 \times CAL + w_2 \times Focal\quad loss + w_3 \times GDL \tag{4}$$

where w_1, w_2, and w_3 represent the weights of the three loss functions.

4 EXPERIMENTS AND RESULTS

4.1 Datasets

Following previous work of pancreas segmentation, two different abdominal CT datasets are used:

- As one of the largest and most authoritative Open Source Dataset in pancreas segmentation, the NIH pancreas segmentation dataset sourced from TCIA (The Cancer Imaging Archive) provides an easy and fair way for method comparisons (51). The dataset contains 82 contrast-enhanced abdominal CT volumes. The resolution of each CT scan is 512 × 512 × L, where L have a range of 181 to 466 which is the number of sampling slices along the long axis of the body. The dataset contains a total of 19,327 slices from the 82 subjects, and the slice thickness varies from 0.5 to 1.0 mm. Only the CT slices containing the pancreas are used as input to the system. We followed the standard four-fold cross-validation, where the dataset is split to four folds, each fold contains images of 20 subjects, and the proposed model was trained on 3 folds and tested on the remaining fold.
- The Medical Segmentation Decathlon (52) is a challenge to test the generalizability of machine learning algorithms when applied to 10 different semantic segmentation tasks. In addition, we use the pancreas part in modality of portal venous phase CT from Memorial Sloan Kettering Cancer Center. We used the official training-test splits where 281 subjects are in training set and 139 subjects are in test set.

4.2 Evaluation Metric

The performance of our approach on pancreas segmentation was evaluated in terms of dice similarity coefficient (DSC)

$$DSC(Z,Y) = \frac{2 \times |Z \cap Y|}{|Z| + |Y|} \tag{5}$$

where Z is the predicted segmentation and Y is the ground truth. We reported the maximum, minimum, and average values of DSC score over all testing cases in the NIH dataset and MSD dataset (52).

Moreover, we also use Jaccard coefficient, recall, and precision as auxiliary metric:

$$Jaccard(U,V) = \frac{|U \cap V|}{|U \cup V|} \tag{6}$$

Where U and V represent the real pancreatic area and the predicted pancreatic area (pixel level), respectively.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

In addition, for the metric of the segmentation problem, although Dice and others can well reflect the difference between the segmentation effect and the actual situation, its defect is insensitivity to differences in target boundaries, and the focus is mainly on the inside of the mask, while the Hausdorff distance (HD) as a measure of shape similarity, can be a good complement to Dice. In a 2D plane, HD refers to the maximum of all distances from one set to the nearest point between another set. Given two finite set of points $A = \{\alpha_1,\ldots\alpha_p\}$ and $B = \{b_1,\ldots b_p\}$, the HD between them is defined as follows:

$$H(A,B) = max\{h(A,B),\ h(B,A)\} \tag{9}$$

where $\mathbf{h(A,B)} = \max a \in A \max b \in B \in a - b \in$, $\mathbf{h(B,A)} = \max b \in B \max a \in A \in b - a \in$, $\| \ \|$ is a distance norm defined on point set A and point set B. We use the Euclidean distance representation directly.

4.3 Implementation Details

We implement our approach base on PaddlePaddle platform on a server equipped with V100 Tesla GPU with 32-GB memory. We use four-fold cross-validation for training and use min max normalization to scale the pixel values of the original image to [0, 1] and performed independently on the training and test sets. We found that RMS optimizer has a faster convergence speed than the Adam optimizer. Although adaptively reducing the learning rate, RMS optimizer can still get convergence on a smaller number of iterations. Thus, we used RMS as our optimizer. Our complete source code is available at Github https://github.com/zhangyuhong02/AX-Unet.git. We list our hyperparameters and system settings in **Table 1**.

Because the method that we proposed achieves a variety of improvements in multiple levels of the network structure such as loss function, deep supervision and the form of deep supervision, we compare with the state-of-the-art methods in terms of multiple improvement direction control variables and the combined effects of each improvement structure.

We performed some basic processing on the original image. We performed 2.2 times contrast enhancement (the best performance can be obtained through hyperparameter grid search). **Figure 2** shows our comparative data enhancement effect.

TABLE 1 | Hyperparameters and device parameters.

Parameter	Value
Initial learning rate	0.001
Batch size	32
Epochs	150
Optimizer	RMS
Learning rate decay	fixed size
convolution kernel size	3×3
PaddlePaddle	2.1.2+cu101
CUDA	10.1
python	3.7
GPU	*TeslaV*100 × 4
RAM	128GB

4.4 Results

In this section, we compare our proposed method with the state-of-the-art methods for image segmentation. **Table 2** shows the segmentation performance on NIH and MSD datasets in terms of DSC, Jaccard, precision, and recall. From **Table 2**, our framework can outperform the other state-of-the-art methods by a wide margin in terms of DSC, Jaccard, precision, and recall. The mean HD between out segmentation and the ground truth is 4.68, with a standard deviation 1.76. **Figure 3** shows three examples of our segmentation results. We initialized different training parameters and conducted 15 independent repeated experiments on the NIH dataset and recorded the dice score for each trained model. The mean dice score is 87.67, and the standard deviation is 3.8. We compared our results on NIH dataset with state-of-the-art methods through one sample t test, as shown in **Table 3**. From **Table 3**, our proposed method has statistically significant improvements ($p < 0.0001$) compared with other methods.

4.4.1 Ablation Experiment

To demonstrate the effectiveness of our group convolution and other structures, we conducted an ablation experiment to evaluate the effects of each part in our framework, residual structure, depth-separable convolution module, and ASPP module on the segmentation results. We conduct experiments using separate additional structures or different combinations of the proposed structures and perform the four-fold cross-validation on the same NIH dataset, and we repeated the experiments with different initializations for 10 times. The results are shown in **Figure 4** and **Table 4**.

It can be seen that the depth-wise separable convolution achieves the greatest performance improvement when using

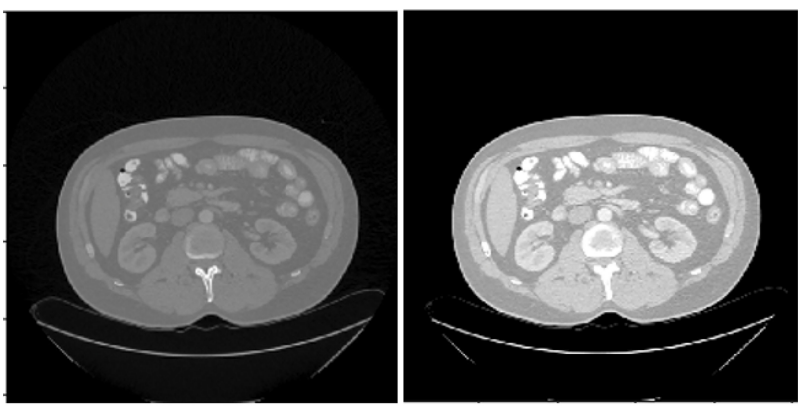


FIGURE 2 | Original image and contrast-enhanced image.

TABLE 2 | The average four-fold performance on two public dataset (the performance of our method is described by *mean* ± *std*).

Method	DSC (%)	Jaccard (%)	Recall (%)	Precision (%)
NIH dataset				
Bottom-up (32)	70.7	57.9	71.6	74.4
Fixed-point (53)	82.4	–	–	–
3D Coarse-to-Fine (54)	84.6	–	–	–
Holistically nested (55)	81.3	68.9	–	–
RSTN (31)	84.5	–	–	–
Recurrent Contextual Learning (39)	83.3	71.8	84.5	82.8
Vnet (56)	80.1	–	–	–
Attention Unet (57)	83.1	–	–	–
DenseASPP (40)	85.4	–	–	–
(46)	84.10	72.86	85.3	83.6
Cascaded FCN (23)	85.9	75.7	85.2	87.6
AX-Unet (Ours)	**87.7 ± 3.8**	**78.2 ± 5.3**	**90.9 ± 2.2**	**92.9 ± 6.1**
MSD dataset				
Unet-64	70.7	–	–	–
Unet-16	67.1	–	–	–
Attention Unet (57)	66.0	–	–	–
MoNet (58)	74.0	68.9	–	–
nn-Unet (27)	80.0	–	–	–
AX-Unet (Ours)	**85.9 ± 5.1**	**77.9 ± 3.4**	**86.3 ± 5.1**	**93.1 ± 6.9**

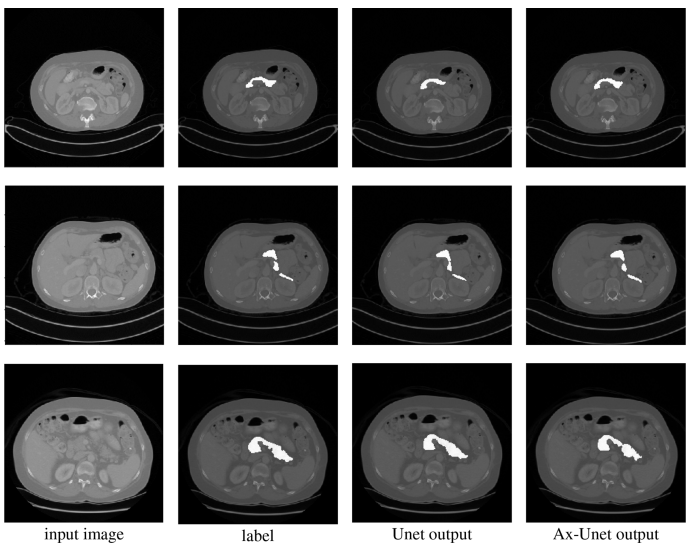


FIGURE 3 | Comparison of segmentation for three examples by the baseline model (Unet) and the AX-Unet, along with the original image and ground truth. In each row, from left to right, the images correspond to the original image, ground-truth segmentation, the baseline segmentation by Unet, and segmentation by our AX-Unet model, respectively. It can be clearly observed that the proposed model has better segmentation effect of the boundary than the baseline.

TABLE 3 | t-value and p-value for our method by one sample t-test.

Methods	t-value	p-value
RSTN (31)	9.2338	4.02×10^{-7}
3D Coarse-to-Fine (54)	8.9403	8.92×10^{-7}
DenseASPP (40)	6.5921	1.28×10^{-5}
Cascaded FCN (23)	5.1245	0.0001

only a single part, which validates the effectiveness depth-wise on the two correlation decoupling operations. Although the introduction of ASPP module alone did not achieve better results, the combination with depth-wise separable convolution achieved very good results. Combining all the proposed modules can achieve the best performance.

4.4.2 3D Rebuilding

To better demonstrate our segmentation effect, besides the segmentation results in **Figure 3**, we also show an example of the 3D rebuilding results based on our segmentation in **Figure 5**. From **Figure 5**, the rebuilding results based on our segmentation are similar with that from the ground truth, which validates the efficacy of our model.

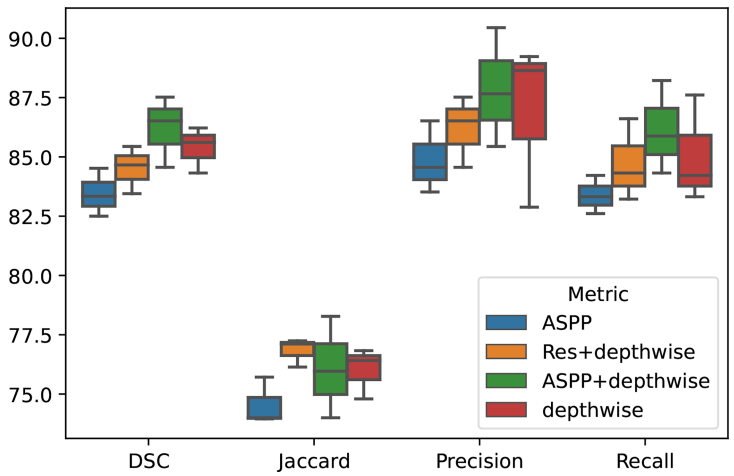


FIGURE 4 | Ablation experiment on different group of module proposed in our paper.

TABLE 4 | Results of ablation studies with different components.

Method	Jaccard (%)
Residual block	69.7 ± 8.9**
ASPP module (2,4,6)	76.5 ± 4.9**
Resuidual+ASPP(2,4,6)	76.8 ± 6.4**
depth-separable conv	77.4 ± 4.3*
Residual block+Depth-separable conv	76.7 ± 6.2*
Depth-separable conv+ASPP(2,4,6)	77.8 ± 3.2*
all	78.2 ± 5.3

The performance of different substructures is described by **mean ± std**; *the t-test was used for significance analysis, in which the all group containing all structures was the control group; ** indicated extremely significant difference (*$p < 0.01$*);* indicated significant difference (*$p < 0.05$*).*

4.5 Activation Map

Besides giving the segmentation results, the network can also output the activation maps of each layer, which could show a clear decision making process and give a clear medical evidence. Analyzing the activation map in the forward propagation process of the neural network can help to understand the decision making process of the model, thereby helping clinicians to achieve procedural diagnosis and more accurate treatment selection.

We extract the feature maps after each pooling in the downsampling process, take the average and maximum values of the feature maps in different levels in the channel dimension, and convert them into activation maps for visualization.

As shown in **Figure 6**, we extract the activate map after the pooling operations in two ways. The first row represents the activate map obtained by averaging the corresponding pixel values of each channel of the feature map of the specified level. The second row represents the activate map obtained by taking the maximum value of the corresponding pixel value of each channel. It can be clearly seen that the high-level feature maps have low resolution but strong semantics during downsampling, whereas the low-level feature maps have high resolution and rich details. This illustrates the necessity of our fusion of feature maps at different levels.

5 PATHOLOGICAL ANALYSIS OF PANCREATIC TUMORS WITH OUR MODEL

As we introduced before, the diagnosis of tumors based on morphological features has been used in brain tumors and other fields. To test the segmentation performance of our model in more complex scenarios and broaden its application scenarios, we use the proposed model to extract imagomics

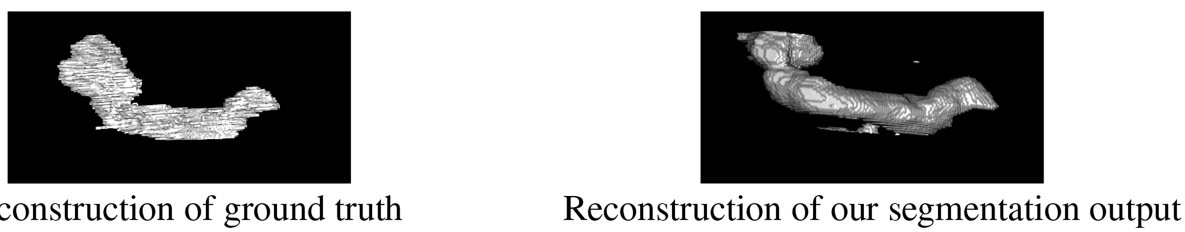


FIGURE 5 | The results of 3D rebuilding. The left picture is the reconstruction of ground truth, and the right picture is the reconstruction of the segmentation output of our model.

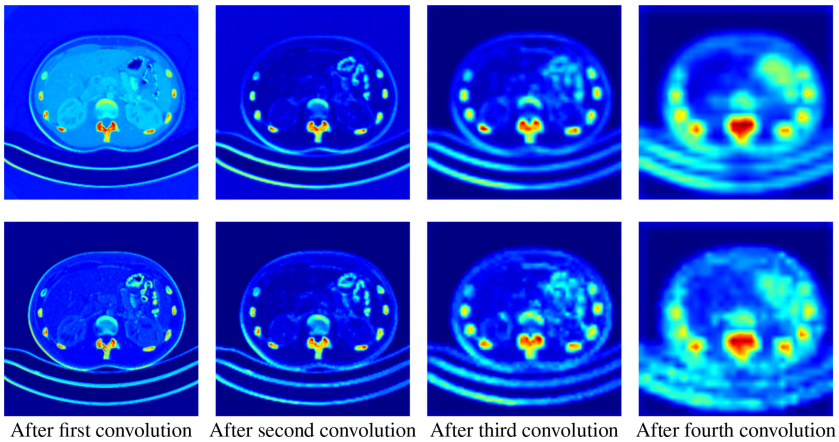


FIGURE 6 | Activation maps transformed from feature maps of different levels. The upper row is the average activation maps over channels, and the lower row is the max activated maps over channels. From left to right, the activation maps are from the output of the first to the fourth downsampling block, respectively.

features for analysis. To further explore the relationship between pancreatic tumors and imagomics features and to verify the robustness of our model, we collected a large number of unlabeled data and used our pre-trained model for few-shot learning to identify pancreatic regions, followed by imagomics feature extraction and significant difference analysis.

5.1 Data Collection and Processing Methods

We collected pancreas image data from 49 patients from The First Hospital of Lanzhou University, which contains 31 pancreatic tumor patients and 13 normal subjects. The ages ranged from 18 to 76 years with a mean (std) of 46.8 (16.7). The CT scans have resolutions of 512 × 512 with pixels. The slice thickness is between 1.5 and 2.5 mm. The CT imaging was created using Somatom Sensation scanner with the following parameters: craniocaudal abdominal scan (120-kVp tube voltage). We manually annotated pancreas images of five individuals for the fine-tuned task and used the best performing model on the NIH Dataset as our pre-trained model. A medical student manually performed slice-by-slice segmentation of the pancreas as ground truth, and these were verified by an experienced radiologist.

5.2 Ethical Approval

Institutional Review Board (IRB) approval was obtained prior to the collection of the dataset. The institutional review board of the first hospital of Lanzhou university approved this study and waived the need for informed consent.

5.3 Transfer Learning and Feature Extraction

Through transfer learning, we fine-tuned the model trained on the public dataset on a small number of labeled samples from our dataset dataset. Then, we segmented the unlabeled data and extract 10 representative texture features from the segmentation results for pathological analysis of tumors. The features we extract are entropy (10), energy (11), homogeneity of the gray level co-occurrence matrix (glcm) (12), glcm dissimilarity (13), edge sharpness (Acu) (14), contrast (15), gray mean (59), glcm contrast (GC), glcm mean, and glcm std (60).

Contrast reflects the definition of graphics and the depth of texture, which can measure the distribution of pixel values and the amount of local changes in the image. Energy is a measure of the stability of image texture gray changes, which reflects the uniformity of image gray distribution and texture thickness. Entropy is used to measure the randomness (i.e., intensity distribution) of image texture and characterize the complexity of the image. In addition, other features are calculated based on the gray level co-occurrence matrix, which can reflect the comprehensive information of image gray level about direction, adjacent interval, change amplitude, etc. The local model of the image and the arrangement rules of the pixels are used for analysis.

In Equations (10) to (15), S, E, GH, GD, Acu, and C represent entropy, energy, homogeneity and dissimilarity of gray-level co-occurrence matrix, sharpness of image edges, entropy, and contrast, respectively, and P_{ij} stands for the position of the current pixel.

Then, we checked the correlation of the extracted features themselves and screened out the irrelevant features with comparison differences. After comparative analysis, we eliminated the energy and glcm dissimilarity that were highly correlated with other features. As shown in **Figure 7**, we use the Pearson correlation coefficient to measure the correlation between variables and find that energy and glcm dissimilarity are highly correlated with other features.

$$S = \sum_{i,j=0}^{N-1} P_{i,j}\left(-lnP_{i,j}\right) \tag{10}$$

$$E = -\sum_{i} \sum_{i} P_{i,j}^{2} \tag{11}$$

$$GH = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i-j)^2} \tag{12}$$

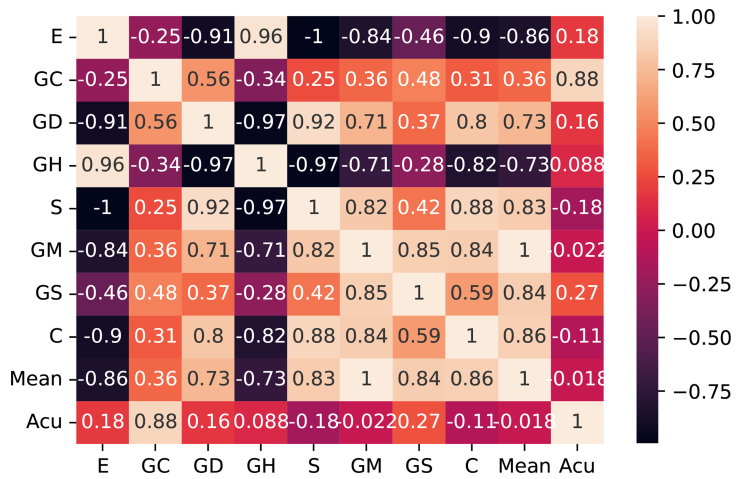


FIGURE 7 | Correlation matrix with Pearson correlation coefficient of the 10 features. E, entropy; GC, gray-level co-occurrence matrix contrast; GD, gray-level co-occurrence matrix dissimilarity; GH, gray-level co-occurrence matrix homogeneity; S, entropy; GM, gray mean; GS, gray standard deviation; C, contrast; Acu, sharpness of image edges.

$$GD = \sum_{i,j=0}^{N-1} P_{i,j} \quad |i-j| \tag{13}$$

$$Acu = \sum_{i}\sum_{j} \left[P_{i,j} - \mu\right]^2 \tag{14}$$

$$C = \sum_{i,j=0}^{N-1} P_{i,j} \quad (i-j)^2 \tag{15}$$

5.4 Results and Discussion

In this study, we have 31 pancreatic tumor patients and 13 normal subjects. After the features are extracted, we use the Shapiro–Wilk test to check how likely the extracted features follow a normal distribution. Feature distribution visualization and the results of the Shapiro–Wilk test are shown in **Figure 8**. Although most of the distributions have a p-value of the Shapiro–Wilk test more than 0.05, it can be found that most of the features' distribution is skewed to some extent, and it is safe to use a non-parametric test for significant difference analysis. We performed a Mann–Whitney U rank test to test whether a certain characteristic is significantly different between pancreatic

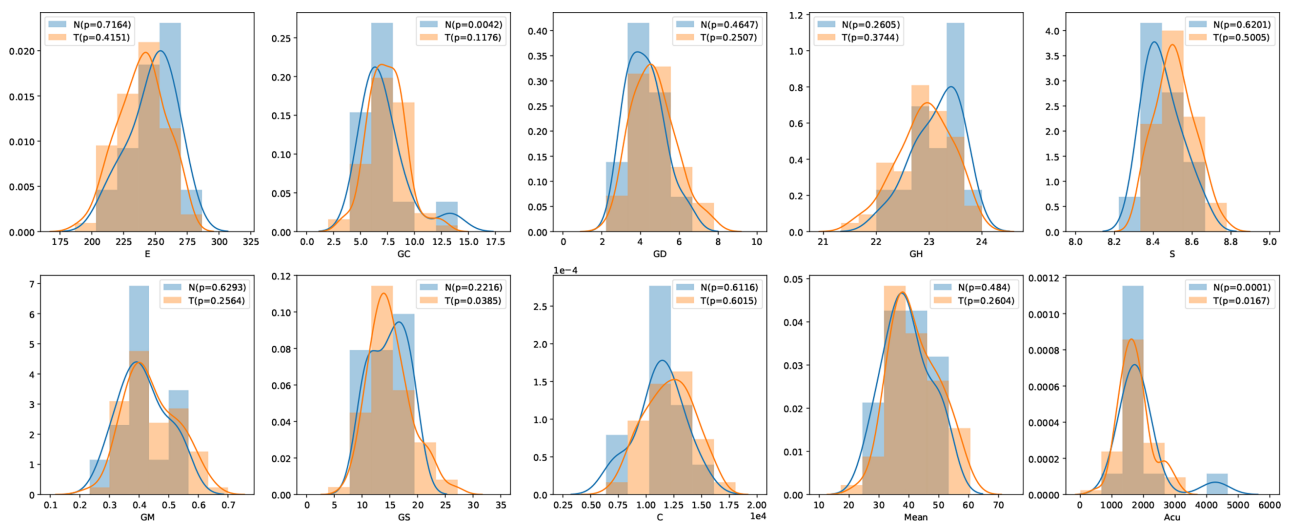


FIGURE 8 | Feature distribution visualization. N represents the group of normal subjects, and T represents the group of pancreatic tumor patients. p value is the results of Shapiro–Wilk test.

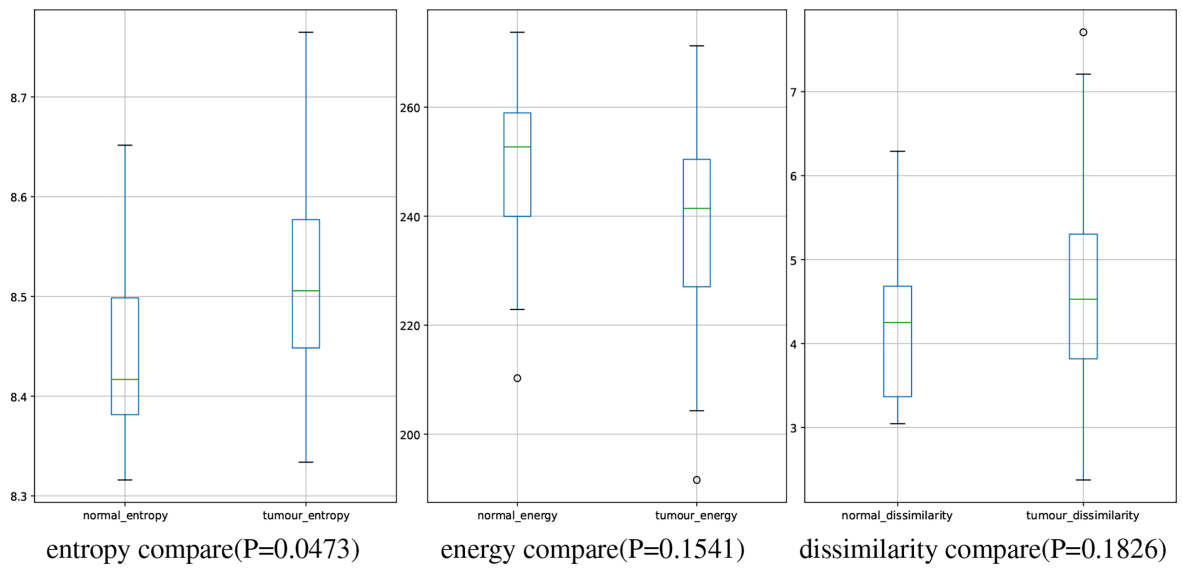


FIGURE 9 | Boxplot of numerical distributions with different features.

tumor patients and normal subjects. After our calculation, it was found that the entropy extracted from the segmented images was significantly different between pancreatic tumor patients and normal people ($P \leq 0.05$). The box plot of entropy, energy, and dissimilarity is shown in **Figure 9**. We believe that the feature entropy extracted from the output segmentation of the model is helpful for pancreas tumor diagnosis.

Entropy represents the feature of increased cellular heterogeneity during the differentiation of normal tissue into tumor tissue, which not only can reflect the difference in entropy between the two tissues on CT images but also can predict tumor recurrence and metastasis. For example, entropy can predict the pathological grade in pancreatic neuroendocrine tumors; while the entropy increases, the possibility of high-grade will increase. In addition, in related studies (61), image features of peritumoral tissue vary differently from pancreatic tumor, which may demonstrate the possibility of entropy for predicting recurrence of pancreas tumor and metastasis of small tumor from other organs.

By constructing such an interdisciplinary pancreas segmentation model, it can be applied to multiple topics in clinical research. It may be applied to the detection of small tumors and the relationship between pancreatic margins and pancreatic fibrosis and to explore the relationship between tumor or pancreatic tissue margins and important blood vessels, so as to make more reasonable treatment choices, implement the concept of precision surgery.

6 CONCLUSION

This paper proposes a novel deep learning framework AX-Unet for image segmentation for pancreas CT images. Facing the challenging scene of pancreatic segmentation, we analyzed the defects of the existing mainstream segmentation framework for medical images and proposed a more sophisticated network structure based on the encoder-decoder structure. We combine the ASPP module with multi-scale feature extraction capabilities and group convolutions that can decouple information. It can show excellent results when facing small targets that are blurred by the boundary of the pancreas and are easy to confuse the surrounding tissues. Finally, we used the proposed segmentation model to extract and analyze the radiomics features and found that there were significant differences in entropy between normal and pancreatic tumor patients, providing a promising and reliable way to assist physicians for the screening of pancreatic tumors.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the First Hospital of Lanzhou University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript.

FUNDING

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFA0706200), in part by the National Natural Science Foundation of China (Grant No.61632014, No.61627808).

REFERENCES

1. [Dataset] National Cancer Institute. *Cancer Stat Facts: Common Cancer Sites* (2021). Available at: https://seer.cancer.gov/statfacts/html/common.html (Accessed 2022-03-05).
2. Liu S, Yuan X, Hu R, Liang S, Feng S, Ai Y, et al. Automatic Pancreas Segmentation *via* Coarse Location and Ensemble Learning. *IEEE Access* (2019) 8:2906–14. doi: 10.1109/ACCESS.2019.2961125
3. Clancey WJ, Shortliffe EH. *Readings in Medical Artificial Intelligence: The First Decade*. New Jersey: Addison-Wesley Longman Publishing Co., Inc (1984).
4. Buchanan BG, Shortliffe EH. Rule-Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project. *Art Intellig* (1984) 26(3):364–6. doi: 10.1016/0004-3702(85)90067-0
5. Son YJ, Kim HG, Kim EH, Choi S, Lee SK. Application of Support Vector Machine for Prediction of Medication Adherence in Heart Failure Patients. *Healthcare Inf Res* (2010) 16:253–9. doi: 10.4258/hir.2010.16.4.253
6. Yu W, Liu T, Valdez R, Gwinn M, Khoury MJ. Application of Support Vector Machine Modeling for Prediction of Common Diseases: The Case of Diabetes and Pre-Diabetes. *BMC Med Inf Dec Mak* (2010) 10:1–7. doi: 10.1186/1472-6947-10-16
7. Zeng Z, Vo AH, Mao C, Clare SE, Khan SA, Luo Y. Cancer Classification and Pathway Discovery Using non-Negative Matrix Factorization. *J Biomed Inf* (2019) 96:103247. doi: 10.1016/j.jbi.2019.103247
8. Chao G, Mao C, Wang F, Zhao Y, Luo Y. Supervised Nonnegative Matrix Factorization to Predict Icu Mortality Risk, in: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (IEEE)*. Madrid, Spain: IEEE (2018) pp. 1189–94.
9. Luo Y, Mao C, Yang Y, Wang F, Ahmad FS, Arnett D, et al. Integrating Hypertension Phenotype and Genotype With Hybrid non-Negative Matrix Factorization. *Bioinf (Oxford England)* (2019) 35:1395–403. doi: 10.1093/bioinformatics/bty804
10. Mao C, Hu B, Wang M, Moore P. (2015). Learning From Neighborhood for Classification With Local Distribution Characteristics, in: *2015 International Joint Conference on Neural Networks (IJCNN) (IEEE)*. Killarney, Ireland: IEEE (2015) pp. 1-8.
11. Hu B, Mao C, Zhang X, Dai Y. (2015). Bayesian Classification With Local Probabilistic Model Assumption in Aiding Medical Diagnosis, in: *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (IEEE)*. Washington, DC, USA: IEEE (2015) pp. 691–4. pp. 691–4.
12. Mao C, Lu L, Hu B. Local Probabilistic Model for Bayesian Classification: A Generalized Local Classification Model. *Appl Soft Comput* (2020) 93:106379. doi: 10.1016/j.asoc.2020.106379
13. Xu S, Rao H, Peng H, Jiang X, Guo Y, Hu X, et al. Attention-Based Multilevel Co-Occurrence Graph Convolutional Lstm for 3-D Action Recognition. *IEEE Internet Thing J* (2020) 8:15990–6001. doi: 10.1109/JIOT.2020.3042986
14. Fang B, Chen J, Liu Y, Wang W, Wang K, Singh AK, et al. Dual-Channel Neural Network for Atrial Fibrillation Detection From a Single Lead Ecg Wave. *IEEE J Biomed Health Inf* (2021) 1. doi: 10.1109/JBHI.2021.3120890
15. Mao C, Yao L, Pan Y, Luo Y, Zeng Z. (2018). Deep Generative Classifiers for Thoracic Disease Diagnosis With Chest X-Ray Images, in: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (IEEE)*. Madrid, Spain: IEEE (2018) 1209–14.
16. Mao C, Yao L, Luo Y. Imagegcn: Multi-Relational Image Graph Convolutional Networks for Disease Identification With Chest X-Rays. *IEEE Trans Med Imaging* (2022) 1. doi: 10.1109/TMI.2022.3153322
17. Mao C, Yao L, Luo Y. Medgcn: Medication Recommendation and Lab Test Imputation *via* Graph Convolutional Networks. *J Biomed Inf* (2022) 127:104000. doi: 10.1016/j.jbi.2022.104000
18. Hu X, Cheng J, Zhou M, Hu B, Jiang X, Guo Y, et al. Emotion-Aware Cognitive System in Multi-Channel Cognitive Radio *Ad Hoc* Networks. *IEEE Commun Magazine* (2018) 56:180–7. doi: 10.1109/MCOM.2018.1700728
19. Giddwani B, Tekchandani H, Verma S. (2020). Deep Dilated V-Net for 3d Volume Segmentation of Pancreas in Ct Images, in: *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN) (IEEE)*. Noida, India: IEEE (2020) pp. 591–6.
20. Mao C, Yao L, Luo Y. (2020). A Pre-Trained Clinical Language Model for Acute Kidney Injury, in: *2020 IEEE International Conference on Healthcare Informatics (ICHI) (IEEE)*. Oldenburg, Germany: IEEE (2020) pp. 1–2.
21. Yao L, Jin Z, Mao C, Zhang Y, Luo Y. Traditional Chinese Medicine Clinical Records Classification With Bert and Domain Specific Corpora. *J Am Med Inf Assoc* (2019) 26:1632–6. doi: 10.1093/jamia/ocz164
22. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE (2016) pp. 770–8.
23. Xue J, He K, Nie D, Adeli E, Shi Z, Lee SW, et al. Cascaded Multitask 3-D Fully Convolutional Networks for Pancreas Segmentation. *IEEE Trans Cybernet* (2019) 51:2153–65. doi: 10.1109/TCYB.2019.2955178
24. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. (2016). Rethinking the Inception Architecture for Computer Vision, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE (2016) pp. 2818–26.
25. Murugesan B, Sarveswaran K, Shankaranarayana SM, Ram K, Joseph J, Sivaprakasam M. (2019). Psi-Net: Shape and Boundary Aware Joint Multi-Task Deep Network for Medical Image Segmentation, in: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE)*. Berlin, Germany: IEEE (2019) pp. 7223–6.
26. Chollet F. (2017). Xception: Deep Learning With Depthwise Separable Convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Jeju, South Korea: IEEE (2017) pp. 1251–8.
27. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. Nnu-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nat Methods* (2021) 18:203–11. doi: 10.1038/s41592-020-01008-z
28. Ronneberger O, Fischer P, Brox T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer)*. Cham; Springer International Publishing (2015) pp. 234–41.
29. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. (2018). Unet++: A Nested U-Net Architecture for Medical Image Segmentation, in: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (Springer)*. Cham: Springer International Publishing (2018) pp. 3–11.
30. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-Denseunet: Hybrid Densely Connected Unet for Liver and Tumor Segmentation From Ct Volumes. *IEEE Trans Med Imaging* (2018) 37:2663–74. doi: 10.1109/TMI.2018.2845918
31. Yu Q, Xie L, Wang Y, Zhou Y, Fishman EK, Yuille AL. Recurrent Saliency Transformation Network: Incorporating Multi-Stage Visual Cues for Small Organ Segmentation. *Proc IEEE Conf Comput Vision Pattern Recog* (2018) 8280–9. doi: 10.1109/CVPR.2018.00864
32. Farag A, Lu L, Roth HR, Liu J, Turkbey E, Summers RM. A Bottom-Up Approach for Pancreas Segmentation Using Cascaded Superpixels and (Deep) Image Patch Labeling. *IEEE Trans Imag Process* (2016) 26:386–99. doi: 10.1109/TIP.2016.2624198

33. Cai J, Lu L, Xing F, Yang L. Pancreas Segmentation in Ct and Mri Images *via* Domain Specific Network Designing and Recurrent Neural Contextual Learning. *ArXiv Preprint ArXiv* (2018) 1803:11303. doi: 10.1109/TIP.2016.2624198
34. Man Y, Huang Y, Feng J, Li X, Wu F. Deep Q Learning Driven Ct Pancreas Segmentation With Geometry-Aware U-Net. *IEEE Trans Med Imaging* (2019) 38:1971–80. doi: 10.1109/TMI.2019.2911588
35. Zhang F, Wang Y, Yang H. Efficient Context-Aware Network for Abdominal Multi-Organ Segmentation. *ArXiv Preprint ArXiv* (2021) 2109:10601. doi: 10.48550/arXiv.2109.10601
36. Ribalta Lorenzo P, Marcinkiewicz M, Nalepa J. Multi-Modal U-Nets With Boundary Loss and Pre-Training for Brain Tumor Segmentation. *Int MICCAI Brainlesion Workshop (Springer)* (2019) 11993: 135–47. doi: 10.1007/978-3-030-46643-5_13
37. Shi Y, Zhang J, Ling T, Lu J, Zheng Y, Yu Q, et al. Inconsistency-Aware Uncertainty Estimation for Semi-Supervised Medical Image Segmentation. *IEEE Trans Med Imaging* (2021) 41(3):608–20. doi: 10.1109/TMI.2021.3117888
38. Yang Z, Peng X, Yin Z. (2020). Deeplab_v3_plus-Net for Image Semantic Segmentation With Channel Compression, in: *2020 IEEE 20th International Conference on Communication Technology (ICCT) (IEEE)*. pp. 1320–4.
39. Cai J, Lu L, Xie Y, Xing F, Yang L. Improving Deep Pancreas Segmentation in Ct and Mri Images *via* Recurrent Neural Contextual Learning and Direct Loss Function. *ArXiv Preprint ArXiv* (2017) 1707:04912. doi: 10.48550/arXiv.1707.04912
40. Hu P, Li X, Tian Y, Tang T, Zhou T, Bai X, et al. Automatic Pancreas Segmentation in Ct Images With Distance-Based Saliency-Aware Denseaspp Network. *IEEE J Biomed Health Inf* (2020) 25:1601–11. doi: 10.1109/JBHI.2020.3023462
41. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic Image Segmentation With Deep Convolutional Nets, Atrous Convolution, and Fully Connected Crfs. *IEEE Trans Pattern Anal Mach Intell* (2017) 40:834–48. doi: 10.1109/TPAMI.2017.2699184
42. Roth H, Oda M, Shimizu N, Oda H, Hayashi Y, Kitasaka T, et al. Towards Dense Volumetric Pancreas Segmentation in Ct Using 3d Fully Convolutional Networks. *Med Imaging 2018: Imag Process (International Soc Optic Photonics)* (2018) 10574:105740B.
43. Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, et al. (2020). Unet 3+: A Full-Scale Connected Unet for Medical Image Segmentation, in: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE)*, . pp. 1055–9. doi: 10.1117/12.2293499
44. Alhichri H, Alswayed AS, Bazi Y, Ammour N, Alajlan NA. Classification of Remote Sensing Images Using Efficientnet-B3 Cnn Model With Attention. *IEEE Access* (2021) 9:14078–94. doi: 10.1109/ACCESS.2021.3051085
45. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *ArXiv Preprint ArXiv* (2017) 1704:04861. doi: 10.48550/arXiv.1704.04861
46. Roth HR, Lu L, Farag A, Shin HC, Liu J, Turkbey EB, et al. (2015). Deeporgan: Multi-Level Deep Convolutional Networks for Automated Pancreas Segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer)*, . pp. 556–64.
47. Cai J, Lu L, Zhang Z, Xing F, Yang L, Yin Q. (2016). Pancreas Segmentation in Mri Using Graph-Based Decision Fusion on Convolutional Neural Networks, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer)*, . pp. 442–50.
48. Tan M, Le Q. (2019). Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks, in: *International Conference On Machine Learning (PMLR)*, . pp. 6105–14.
49. Zhang Y, Wu J, Wang S, Liu Y, Chen Y, EX Wu, et al. (2020). Liver Guided Pancreas Segmentation, in: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI) (IEEE)*, . pp. 1201–4.
50. Xu W, Liu H, Wang X, Ouyang H, Qian Y. (2020). Counet: An End-to-End Colonoscopy Lesion Image Segmentation and Classification Framework, in: *2020 The 4th International Conference on Video and Image Processing*, . pp. 81–7.
51. Roth HR, Farag A, Turkbey EB, Lu L, Liu J, Summers RM. *Nih Pancreas-Ct Dataset*. (2017).
52. Antonelli M, Reinke A, Bakas S, Farahani K, Landman BA, Litjens G, et al. The Medical Segmentation Decathlon. *ArXiv Preprint ArXiv* (2021) 2106:05735. doi: 10.48550/arXiv.2106.05735
53. Zhou Y, Xie L, Shen W, Wang Y, Fishman EK, Yuille AL. (2017). A Fixed-Point Model for Pancreas Segmentation in Abdominal Ct Scans, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer)*, . pp. 693–701.
54. Zhu Z, Xia Y, Shen W, Fishman EK, Yuille AL. A 3d Coarse-to-Fine Framework for Automatic Pancreas Segmentation. *ArXiv Preprint ArXiv* (2017) 1712:00201. doi: 10.48550/arXiv.1712.00201
55. Roth HR, Lu L, Lay N, Harrison AP, Farag A, Sohn A, et al. Spatial Aggregation of Holistically-Nested Convolutional Neural Networks for Automated Pancreas Localization and Segmentation. *Med Imag Anal* (2018) 45:94–107. doi: 10.1016/j.media.2018.01.006
56. Abdollahi A, Pradhan B, Alamri A. Vnet: An End-to-End Fully Convolutional Neural Network for Road Extraction From High-Resolution Remote Sensing Data. *IEEE Access* (2020) 8:179424–36. doi: 10.1109/ACCESS.2020.3026658
57. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: Learning Where to Look for the Pancreas. *ArXiv Preprint ArXiv* (2018) 1804:03999. doi: 10.48550/arXiv.1804.03999
58. Knolle M, Kaissis G, Jungmann F, Ziegelmayer S, Sasse D, Makowski M, et al. Efficient, High-Performance Semantic Segmentation Using Multi-Scale Feature Extraction. *PloS One* (2021) 16:e0255397. doi: 10.1371/journal.pone.0255397
59. Schurink NW, van Kranen SR, Berbee M, van Elmpt W, Bakers FC, Roberti S, et al. Studying Local Tumour Heterogeneity on Mri and Fdg-Pet/Ct to Predict Response to Neoadjuvant Chemoradiotherapy in Rectal Cancer. *Eur Radiol* (2021) 31:7031–8. doi: 10.1007/s00330-021-07724-0
60. Chee CG, Kim YH, Lee KH, Lee YJ, Park JH, Lee HS, et al. Ct Texture Analysis in Patients With Locally Advanced Rectal Cancer Treated With Neoadjuvant Chemoradiotherapy: A Potential Imaging Biomarker for Treatment Response and Prognosis. *PloS One* (2017) 12:e0182883. doi: 10.1371/journal.pone.0182883
61. Fiz F, Costa G, Gennaro N, la Bella L, Boichuk A, Sollini M, et al. Contrast Administration Impacts Ct-Based Radiomics of Colorectal Liver Metastases and non-Tumoral Liver Parenchyma Revealing the "Radiological" Tumour Microenvironment. *Diagnostics* (2021) 11:1162. doi: 10.3390/diagnostics11071162

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 | Frontiers in Oncology

ORIGINAL RESEARCH
published: 17 June 2022
doi: 10.3389/fonc.2022.886739

Early Prediction of Lung Cancers Using Deep Saliency Capsule and Pre-Trained Deep Learning Frameworks

*Kadiyala Ramana[1], Madapuri Rudra Kumar[2], K. Sreenivasulu[2], Thippa Reddy Gadekallu[3], Surbhi Bhatia[4], Parul Agarwal[5] and Sheikh Mohammad Idrees[6]**

[1] Department of Information Technology (IT), Chaitanya Bharathi Institute of Technology, Hyderabad, India, [2] Department of Computer Science and Engineering (CSE), G. Pullaiah College of Engineering and Technology, Kurnool, India, [3] Department of Information Technology, Vellore Institute of Technology, Vellore, India, [4] Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Al Hasa, Saudi Arabia, [5] Department of Computer Science and Engineering (CSE), Jamia Hamdard, India, [6] Department of Computer Science Institutt for datateknologi og informatikk (IDI), Norwegian University of Science and Technology, Gjøvik, Norway

OPEN ACCESS

Edited by:
Shahid Mumtaz, Instituto de Telecomunicações, Portugal

Reviewed by:
Delphin Raj, Kookmin University, South Korea
Chennareddy Vijay Simha Reddy, Middlesex University, United Kingdom
Lucas Lima, University of São Paulo, Brazil

***Correspondence:**
Sheikh Mohammad Idrees
sheikh.m.idrees@ntnu.no

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

Received: *28 February 2022*
Accepted: *13 May 2022*
Published: *17 June 2022*

Citation:
Ramana K, Kumar MR, Sreenivasulu K, Gadekallu TR, Bhatia S, Agarwal P and Idrees SM (2022) Early Prediction of Lung Cancers Using Deep Saliency Capsule and Pre-Trained Deep Learning Frameworks. Front. Oncol. 12:886739. doi: 10.3389/fonc.2022.886739

Lung cancer is the cellular fission of abnormal cells inside the lungs that leads to 72% of total deaths worldwide. Lung cancer are also recognized to be one of the leading causes of mortality, with a chance of survival of only 19%. Tumors can be diagnosed using a variety of procedures, including X-rays, CT scans, biopsies, and PET-CT scans. From the above techniques, Computer Tomography (CT) scan technique is considered to be one of the most powerful tools for an early diagnosis of lung cancers. Recently, machine and deep learning algorithms have picked up peak energy, and this aids in building a strong diagnosis and prediction system using CT scan images. But achieving the best performances in diagnosis still remains on the darker side of the research. To solve this problem, this paper proposes novel saliency-based capsule networks for better segmentation and employs the optimized pre-trained transfer learning for the better prediction of lung cancers from the input CT images. The integration of capsule-based saliency segmentation leads to the reduction and eventually reduces the risk of computational complexity and overfitting problem. Additionally, hyperparameters of pretrained networks are tuned by the whale optimization algorithm to improve the prediction accuracy by sacrificing the complexity. The extensive experimentation carried out using the LUNA-16 and LIDC Lung Image datasets and various performance metrics such as accuracy, precision, recall, specificity, and F1-score are evaluated and analyzed. Experimental results demonstrate that the proposed framework has achieved the peak performance of 98.5% accuracy, 99.0% precision, 98.8% recall, and 99.1% F1-score and outperformed the DenseNet, AlexNet, Resnets-50, Resnets-100, VGG-16, and Inception models.

Keywords: computer tomography (CT) scan images, saliency segmentation, pre-trained models, whale optimization, DenseNet, VGG-16, inception models

1 INTRODUCTION

Lung tumor (LT) is the most lethal cancer on the planet. As a result, numerous countries are working on early detection measures for lung disease. The NLST experiment (1) found that screening high-risk participants three times a year with low-dose computed tomography (CT) reduces death rates significantly (2). As a result of these procedures, a radiologist will have to examine a large number of CT scan images. Because lesions are difficult to identify, even for qualified clinicians, the strain on radiologists grows exponentially as the quantity of CT scans to review grows.

Lung cancer is the second most prevalent cause of cancer death in people. Cancers of the bladder, breast, colon, cervix and, prostate have 5-year survival rates of over 80%. Thus, early identification of lung cancer is critical to reducing mortality or facilitating full care. Due to their thin cell layers (0.2-1*mm*) and lack of symptoms, early lung malignancies and precancers such as dysplasia and carcinoma *in situ* (CIS) are difficult to identify visually using traditional diagnostic procedures such as medical imaging. In clinical practice, roughly 80% of cases are advanced when initially diagnosed and verified, losing the best chance for surgical therapy. Clearly, early detection of lung cancer is clinically significant.

With the predicted rise in the number of preventive/early-detection measures, scientists are developing automated solutions to assist doctors in decreasing their workload, improving diagnostic precision by minimizing subjectivity, speeding up analysis, and lowering medical costs. Specific traits must be detected and assessed to identify the cancerous cells in the lung region. Cancer risk can be determined by the observed features and their combination. Even for an experienced medical expert, this work is challenging because nodule existence and a positive cancer diagnosis are not easily linked. Volume, shape, subtlety, firmness, spiculation, sphericity, and other previously described properties are used in common computer-assisted diagnostic (CAD) techniques.

Machine learning (ML) techniques like Support Vector Machine (SVM) are utilized to identify the nodules as benign or cancerous. Despite the fact that numerous works employ comparable machine learning frameworks (3–10), the limitation of this technique is that in order for the system to function properly, different variables must be customized, making it difficult to repeat results. Furthermore, the lack of uniformity among CT scans and screening parameters makes these systems vulnerable. The development of deep training in CAD systems might do end-to-end identification by acquiring the most essential factors during training. The network is resistant to variations since it gathers tumor features in multiple CT scans with repeated modes. By adopting a training set that is rich in variability, the system may be able to learn invariant properties from malignant nodules intrinsically and enable higher performances (11, 12). Since no characteristics are generated, the system may be able to understand the relationship between traits and disease using the data provided on its own. Once trained, the network should be able to generalize its training and recognize cancerous lesions (or malignancy at the clinical bedside) on cases reported that have never been observed before (13, 14). **Figure 1** shows the normal and abnormal CT lung images. Early classification and classification of lung cancers play a critical role in designing an intelligent and accurate diagnosis system (15). With the advent of machine and deep learning algorithms, the design of early diagnosis systems has reached new heights. Machine learning algorithms such as artificial neural networks (ANN), Support Vector Machines (SVM), Naïve Bayes Classifiers (NB), and Ensemble classifiers (EC) are primarily used for an early diagnosis of lung cancers (16). Also, deep learning is considered to be the most promising field which can enhance the performance of various medical imaging and diagnosis systems (17).

However, handling the images with different imaging protocols remains a real challenge to train the learning modes for greater performance. To compensate for the above drawback of learning models, this paper proposes the novel hybrid intelligent diagnosis framework Deep Fused Features Based Reliable Optimized Networks (DFF-RON), which fuses the saliency maps and convolutional layers for better segmentation and feature extraction that are used to train the ant-lion optimized single feedforward networks. To the best of our knowledge, this is the first work that has integrated the fused

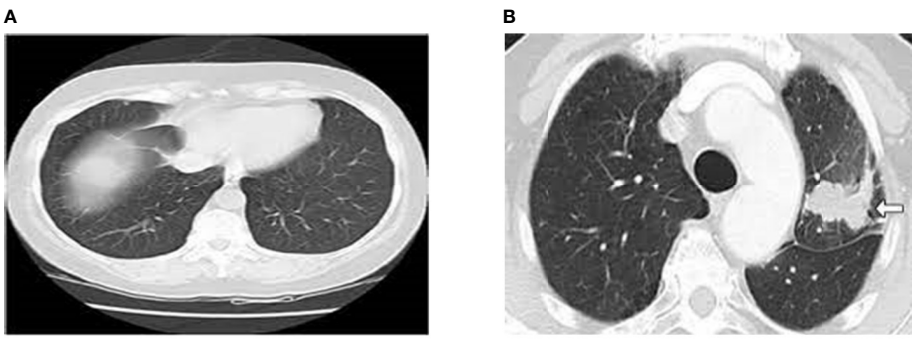


FIGURE 1 | **(A)** Normal CT Lung Image **(B)** Abnormal CT Lung Image (Cancer Image).

features and optimized learning networks to design an efficient and high-performance CT-based lung cancer diagnosis system.

1.1 Contribution of the Research Work

1. A novel hybrid deep learning based model is proposed for the early detection of lung cancer using CT scan images. The proposed architecture has been trained with LIDC datasets and performance metrics have been calculated and compared with other existing models.
2. The proposed architecture introduces the capsule network's better segmentation and transfer learning for feature extraction. Also, the proposed fusion algorithm can increase the high diagnosis rate.
3. The whale optimization algorithm is proposed for training the features obtained from the hybrid fusion of saliency maps and capsule networks. The feed-forward layers are designed based on the principle of Extreme Learning Machines (ELM).

The rest of the paper is organized as follows: Section-II presents the related works proposed by more than one author. The working mechanism of the saliency maps, CNN layers, ant lion optimization, and feedforward networks are presented in Section-III. The dataset descriptions, experimentations, results, findings, and analysis are presented in Section-IV. Finally, the paper is concluded in Section-V with future enhancements.

2 RELATED WORKS

In De Bruijne (18), the presented framework looked at the most up-to-date lung cancer detection and diagnosis methods. Using standardized databases LIDC-IDRI, LUNA 16, and Super Bowl Dataset 2016, the newest lesion detection, identification, and detectors are acquainted with labeled models. According to the author Jindal et al. (19), these are the most common and typical threshold CT data considered for diagnosing. The authors in Nalepa and Kawulok (20) developed the modified-CNN in order to recognize the tumor cells in the lung regions with the segmented images. The ACM method has been used for segregating the tumor region initially and identifying cancer or normal cells.

The label-free techniques do not injure cells or cause effects on cell structure or intrinsic features. To enhance cell identification using recorded optical profiles, this study combined advancements in optical coherence tomography with Prony methods. In Ganesan et al. (21), the framework finds signature genes by improving Tobacco Exposure Pattern (TEP) Prediction model and revealing their interaction connections at many biological levels. TTZ Kasinathan et al. (22) is a new way to extract core features and use them as an input variable in the TEP classification model. With two distinct LUAD datasets used to train and evaluate the TEP classification model, 34 genes were identified as nicotine-associated mutation signature genes, with an accuracy of 94.65% for training data and 91.85% for validation data.

The researcher examined tissue samples and devised a categorization method to discriminate between five types of pulmonary and colorectal tissues (two benign and three malignant). According to the observations, the suggested approach can detect tumor cells up to 96.33% of the time (23). The framework presented in Suzuki (24) described how to use computer-assisted diagnostics to assess EGFR mutation status, including gathering, evaluating, and merging multi-type interdependence characteristics. This research uses a new hybrid network model based on CNN-RNN architecture. CNN is used to extract image quantitative properties, and the link between different types of features is modeled. Their study indicate that multi-type dependency-based feature representations beat single-type feature representations (accuracy = 75%, AUC = 0.78) when compared to conventional features extracted.

The 3D_Alex Net unsupervised learning model (25) was introduced for lung cancer detection. The 3D CNN is a highly predictive architecture with an improved steepest descent input signal that increases the appearance of tumor tissues. The LUNA database is used to assess the proposed Alex Net detection technique to an existing 2D CNN training classifier. Due to a lack of testing data, the proposed model is unsuccessful, with just 10% of the training database being utilized.

Tajbakhsh and Suzuki (26) examined the performance of CNNs and MTANNs for detecting and classifying lung nodules. Achieving 100% sensitivity and 2.7 false positives per patient, MTANN exceeds the top performing CNN (AlexNet) in their testing. The MTANNs achieved an accuracy of 0.88 in classifying nodules as benign or malignant.

Gu et al. (27) suggested a unique 3D-CNN CAD system for lung nodule detection. They used a multiscale technique to improve the system's detection of nodules of varying sizes. The suggested CAD system considers preprocessing, which is common in standalone CAD systems. It uses volume segmentation to create ROI cubes for 3D-CNN classification. After categorization, DBSCAN was used to blend adjacent regions that could be from the same nodule. Larger scale cubes have lesser sensitivity (88%) but an average of one false positive per patient, according to the LUNA16 dataset.

The multi-section CNN model suggested by Sahu et al. (28) uses multiple view sampling to classify nodules and estimate malignancy. Their proposed model is faster than the widely utilized 3D-CNNs. To develop their system, they employed pre-trained MobileNet networks and sample slices extracted in various directions. On the LUNA2016 dataset, the suggested model had a sensitivity of 96% and an AUC of 98%. They estimate the class likelihood of malignancy using a logistic regression model. It estimated malignancy with 93.79% accuracy. Because it is so light, it can be used on smaller devices like phones and tablets.

Deep3DSCan was proposed by Bansal et al. (29). To do so, they applied a deep 3D segmentation technique on CTs. The ResNet-based model was trained using a combination of deep fine-tuned residual network and morphological features. The LUNA16 dataset was utilized for training and testing. The proposed architecture achieved an F1 score of 0.88 in segmentation and classification tasks.

In Jothi et al. (30), the framework designed a controlled CNN classifier for patients with lung cancer to detect potential adenocarcinoma (ADC) and squamous cell carcinoma (SCC). CNN has already been verified using authentic Non-SCLC patient information from preliminary phase afflicted subjects collected at

Massachusetts General Hospital (31). In the record, there are 311 data phases that have been collected. The created CNN, which is a VGG system training predictor, only had a 71% AUC predictive performance, which was insufficient. The VGG CNN model's flaw is that it hasn't been preprocessed for background subtraction or image reconstruction fragmentation, which increases the predictive accuracy. In Kasinathan and Jayakumar (32), the new cloud-based tumor recognition model was developed. The author analyzed various standard dataset "CT-scans and PET-scans" for segmenting the ROC and for recognizing the tumor. In Jakimovski and Davcev (33), the framework proposes a novel deep learning method based on binary particle swarm optimization with a decision tree (BPSO-DT) and CNN to identify the malignant or normal cells in the lung region using the genetic features (34).

3 PROPOSED METHODOLOGY

3.1 System Overview

Figure 2 shows the complete architecture for the proposed framework. The working mechanism of the proposed deep learning-based diagnosis and classification system is subdivided into three important phases. Image preprocessing and augmentation process, capsule based saliency segmentation, accurate feature extraction using the pretrained transfer learning, and finally trained by the whale optimized extreme learning networks.

3.2 Data Preprocessing and Augmentation

As the first step, CT scans are differentiated by using the Histogram Equalization (HOE) process. This pre-processing step is applied for adjusting the image intensities and contrast. The mathematical expression of HOE applied for image preprocessing is given by

$$I = T * N/P \tag{1}$$

Where T*N – Number of Pixels in N levels where N=0, 1, 2, 3……255

P - Total Number of Pixels.

After preprocessing, an adaptive median filter (AMF) is applied over the images for effective denoising. AMF is a category of bilateral images which renders "clean, crisp, and artifact-free edges," and improves the overall appearance.

In the second stage, the image augmentation process is used in the suggested architecture. Deep Neural Network (DNN) (35, 36) leads to overfitting problems where a limited quantity of labeled data is available. The most proficient and efficient method to tackle this problem is data augmentation. During the data augmentation phase, each image undergoes a series of transformations, producing a huge amount of newly corrected training image samples. As discussed in Pei and Hsiao (37), an affine transformation is employed for efficient data augmentation. The offline transformation techniques, such as conversion, ascending, and spins are used. Inputs are correlated with the augmentation step which is extracted before the training phase and the correlated values are utilized to avoid the over-fitting issue. **Figure 3** shows the different lung images obtained after applying the offline transformations.

3.3 Capsule Saliency Segmentation

Segmentation is a technique of partitioning the images with different magnitudes of patterns and pixels. For segregating the images, various techniques have been established. Capsule saliency maps are a structured technique that has been presented here. It subdivides the images into compacted and diverse parts. There will be a reduction in the number of unnecessary elements in the images.

To build the saliency models (38), color difference and spatial difference is applied in a pixel-based processing in which each pixel is represented as a block. To achieve this, pixels 'X' of images and then disintegrated into non-overlapping blocks with size nxn where n=8 and 16, respectively. Hence the saliency maps S(k) are calculated by using the mathematical expressions given by

$$S(k) = \sum_{n=8,16} X(n) * S`(k) \tag{2}$$

Since the location, dimensions, and shape of cancer cells are the same in their adjacent slices, the finishing saliency maps are

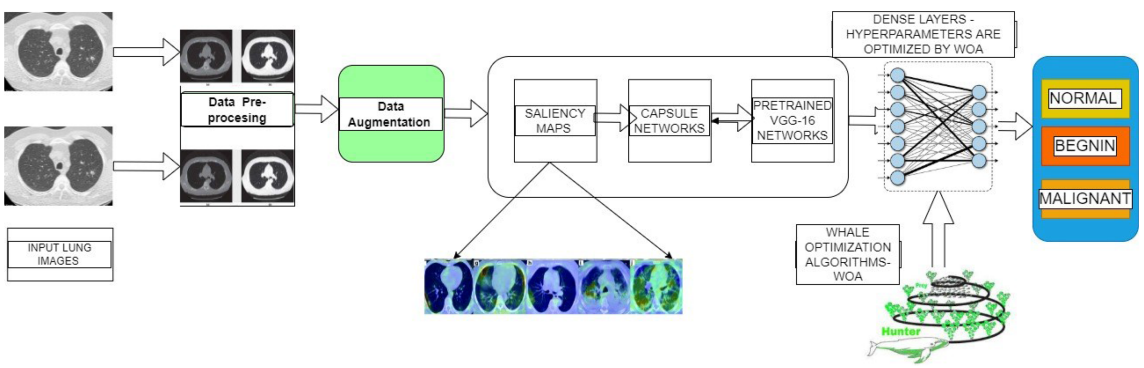


FIGURE 2 | Overall Working Flow Diagram for the Proposed Architectures.

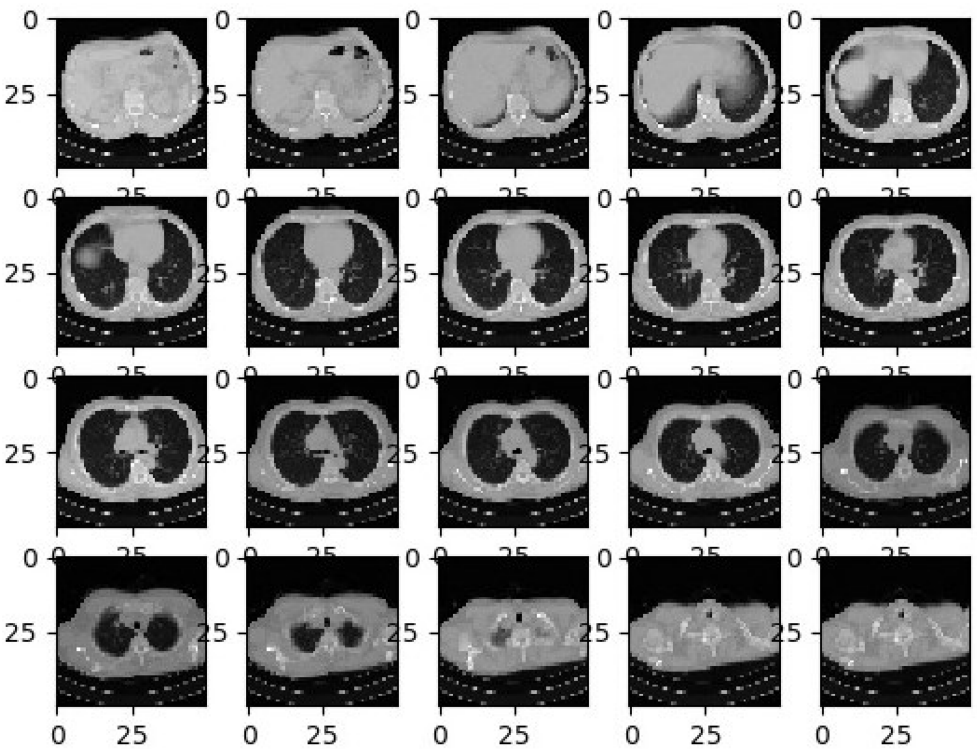


FIGURE 3 | Sample CT-Lung Images after Augmentation Process.

calculated as the biased sum of the authentic (S(m), preceding (S (m1) and next blocks(S(m2)) color and spatial saliency as mentioned in Banerjee et al. (38)

$$S(m) = w1 * S(m1) + w2 * S(m) + w * S(m2) \quad (3)$$

After calculating the saliency maps, post-processing techniques need to be adopted for refinement of segmentation images. Active contour methods [28] are used for the recognition of cancer cells in the most consecutive twin blocks. Also, accurate separation of cancer cells from the other parts of CT scan images is badly needed to give a precise output. Moreover, active contours are based on image intensity, which probably fails in differentiating the cancer cells. Additionally, these contour methods require higher computation time, which is considered to be a serious problem in handling larger datasets.

Motivated by this drawback, this paper introduces capsule networks with pretrained optimized models to obtain high performance and accurate detection of CT lung cancer images. Its main disadvantage seems to be that, in order to get such high standard findings, these techniques necessitate substantial fine-tuning and optimization that is clearly not feasible with massive datasets and has an impact on the recognition rates. But in this proposed system, training effort take reduced time and increase the efficiency and performance of the system.

3.4 Capsule Networks – An Overview

Capsule network (39) is the new and upcoming network that is replacing the prevailing models. The capsule network contains four layers: 1) convolutional layer, 2) hidden layer, 3) PrimaryCaps layer, and 4) DigitCaps layer. **Figure 4** shows the entire working structure for the given training model. The capsule networks provide more advantages in categorizing the distinct saliency maps in the images. The input preprocessed visual image is given as the input to the proposed capsule networks. Capsules are groupings of cells that encrypt the location data as well as the likelihood of an object being present in an image. In capsules networking, there is a shell for every object in an image that gives:

1. Probability of existence in entities
2. Entities' instantiation parameters

The combination of the matrices of the input variables with the weight matrix is computed to represent the essential spatial correlation between poor and large-scale features within the image.

$$Y(i,j) = W_{i,j}U(i,j) * S_j \quad (4)$$

Equation (5) estimates the total weight which is calculated to determine the updating of the current capsule values and the same id feedforward into the next level of capsule determination.

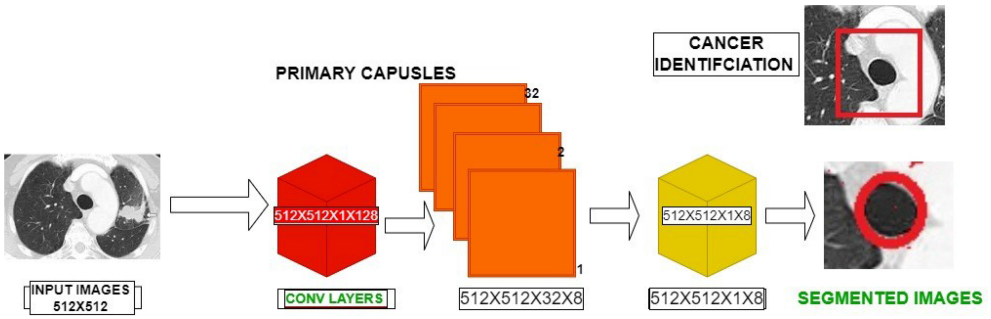


FIGURE 4 | Capsule Architecture for the Saliency Based Segmentation.

$$S(j) = \sum_{j} Y(i,j) * D(j) \tag{5}$$

At last, the squash task is used to apply non-linearity. The squashing utility translates a vector to an extreme length of one and the least length of zero while keeping its orientation.

$$G(j) = squash(S(j)) \tag{6}$$

3.5 Segmentation Process

Figure 4 shows the capsule architecture employed for the saliency segmentation. The preprocessed images are encoded using the equation (2) which involves the convolutional layers and capsule layers. The convolutional operations are first performed over the preprocessed images from conv layers to the capsules of the first capsule layers often followed by the higher capsule layers. The data transformations between the one capsule to other capsule layers are formulated by mathematical equations (3) and (4). Finally, the last layer produces the saliency segmented information which are used toward the categorization.

3.6 Transfer Learning for Feature Extractor

In this process, the transfer learning technique is adopted for better feature extraction and classification. Transfer learning approaches are considered as the pretrained convolutional neural networks that can be repurposed to solve different image classification problems. In this research, the Inception V3 module has been implemented due to its high accuracy and high flexibility. The custom Inception-V3 weights are pre-trained using ImageNet and it considers the reshaped size of 150×150×3 for all images.

3.7 Classification Layers

After the segmentation process, features extracted are then fed for training the networks. In the proposed architecture, traditional training networks are replaced with feedforward networks that are based on the principle of ELM. ELM is a category of neural network proposed by G.B. Huang (40). This kind of neural network utilizes the single hidden layers in which the hidden layers don't require the tuning mandatorily. Compared with the other learning algorithms such as "support vector machine (SVM) and Random Forest (RF), ELM exhibits the better performance," high speed, and less computational overhead.

The working procedure of single-layer network is illustrated with the mathematical formulation which is given below. Generally, the ML classifiers or predictors follow the feature extraction, weights formulation, and identifying the final score for the given problem. The algorithm itself generates the weights and bias factors for identifying the best final score without any backpropagation or stochastic gradient approach which minimizes the computation complexity. This is a major benefit of ELM compared to other networks (41). Due to this, ELM reduces the training error that achieves better results. Most of the categorization problem utilizes this single-layer network and many applications adopt this network for low-level data availability.

In the below statistical estimation, the extracted features are represented as "p" points with their objective function (i.e., sigmoid) where the final score is denoted as a linear graph. The concealed layer may include N-number of nodes which is not tuned mandatorily. The concealed layer's weights are assigned at random (counting the biàs loads). Nodes are not irrelevant; however, they do not need to be calibrated, and the concealed synapse characteristics might be created arbitrarily even in advance. That is, before dealing with the data from the training set. The system yield for a single-hidden layer ELM is given by equation (7)

$$\omega_s(p) = \sum_{i=1}^{s} \alpha_i ab_i(p) = ab(p)\mu \tag{7}$$

where, p→ input
"L" denotes the output weight vector and it is denoted as

$$\mu = [\mu_1, \mu_2, ..., \mu_s]^T \tag{8}$$

$$ab(p) = [ab_1(p), ab_2(p), ..., ab_s(p)] \tag{9}$$

$$ab = \begin{bmatrix} ab(p_1) \\ ab(p_2) \\ . \\ . \\ . \\ ab(p_s) \end{bmatrix} \tag{10}$$

The minimal non-linear least square method is used to denote the basic calculation of ELM that is represented by the below equation.

$$ab` = ab*L = ab^T(ab * ab^T)^{-1}L \quad (11)$$

Where ab* → inverse of "ab": Moore-Penrose generalized inverse.

$$ab` = ab^T(\frac{1}{D}ab * ab^T)^{-1}L \quad (12)$$

Hence the output function can be found by using the above equation

$$\omega_s(p) = ab(p)\alpha = ab(p)ab^T(\frac{1}{D}ab * ab^T)^{-1}L \quad (13)$$

Though extreme learning principle-based feedforward networks produce the best performance, non-optimal tuning of hyperparameters such as input weights, hidden neurons, and learning rate affects the accuracy of classification. Hence, optimization is required for tuning the hyperparameters for achieving the best performance. The next section discusses the proposed algorithm used for optimization of the extreme learning networks.

3.8 Optimized Extreme Learning Models

This section discusses whale optimization algorithm (WOA) and proposed optimized extreme classification layers.

3.8.1 Whale Optimization Algorithms

WOA, first proposed in Mukherjee et al. (42), has sparked renewed interest in recent years. This stochastic search technique is computed by the following simulation of humpback whale behavior and movements in their search for food and supplies. WOA was inspired by the bubble-net attacking method, in which whales target fish by forming tailspin bubbles surrounding them down to 12 meters below the surface, subsequently swimming back up to trap and grab their prey, as shown in **Figure 5**. The search phase in this method is characterized by a randomized hunt for food based on the spatial location of whales, which can be statistically interpreted by automatically updating responses rather than picking the appropriate ones by selecting random solutions.

In addition to this intriguing behavior, WOA differs from other optimization algorithms in that it only requires the adjustment of two parameters. These variables allow for a smooth transition between the exploitation and exploration phases.

Encircling prey: The search process initiated from starting point and circles the food around the nearby region in order to update their process to the best target. The working process is detailed with statistical formulations.

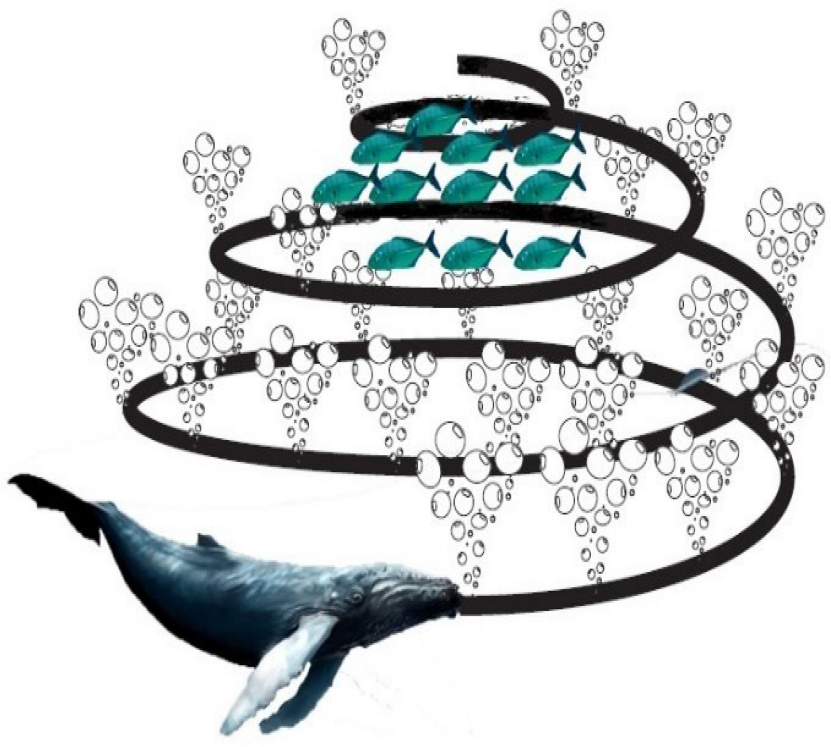


FIGURE 5 | WOA Basic Structure.

If (c < 0.5 and mod(k) < 1)

$$V = modu\{(k \,.\, V - V \,.\, q)\} \tag{14}$$

$$V(q+1) = [V(s) - \{c \,.\, q\}] \tag{15}$$

where c=0.1 (constant), "V(q+1)" represent the best solution and other attributes are estimated as per the below formulations.

$$M = mod\{2 * c * k - c\} \tag{16}$$

$$X_p = 2 * k \tag{17}$$

Where k denotes the arbitrary value within the range of {0 - 2}

Prey Searching: In the food searching process, the input "V" is denoted with "V_{random}" which is estimated using the below equation.

$$V = mod\{O \,.\, V_{rand} - V(q)\} \tag{18}$$

$$V(q+1) = [V_{rand}(q) - \{P \,.\, R\}] \tag{19}$$

During the search phase of the WOA approach, the target was encircled and spiral upgrade was performed. Equation (19) represents the quantitative phrase for updating a new position.

$$V(q+1) = R^1 * ex^{s1} - \cos\ (2\pi p) + V^{\star}(q) \tag{20}$$

"R" denotes the distance among the initial and updated position after each iteration and "s1" denotes the constant 0-1

3.8.2 Proposed Model

As discussed, the WOA model is utilized to enhance the weights of ELM networks. In this case, the whale's criteria for searching and fixing the prey are used as the main term to optimize the weights of ELM networks. Typically, the ELM channels are fed a randomized weight matrix and biased. The performance index is defined as the highest precision. The numerical simulations (14), (15), and (16) are used to determine input bias and weights for each repetition. These parameters are then fed into the ELM system, which generates the exponential function utilizing equations (9). If the output function equals the fitness value, the repetition will either come to a halt or continue. Whale adaptation has a slower convergence time than other meta-heuristic methods, but it takes less time to refine and improve response time. The whale optimized ELM is now used as the classification of lung cancer images. **Table 1** presents the optimized parameters used for training the network.

TABLE 1 | Optimized Parameters for Whale Optimized Extreme Learning Networks.

Sl.no	Parameters	Optimized Parameters
1	No. of Epochs	100
2	Learning Rate	100%
3	No. of batches	20
4	Optimization Iterations	19
5	No. of hidden nodes	78

4 PROPOSED FRAMEWORK VALIDATION

4.1 Datasets Descriptions

The experiments are carried out using lung CT images which are obtained from the cancer imaging archives (https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI). The collection contains 1018 lung CT scans from the National Cancer Institute, which were connected with proteomics and genetic experimental data. All trained radiographs are sorted into normal and cancerous tumors in this article. A benign lesion with a grade of less than 3 is called a normal nodule, while a malignancy lesion with a score of more than 3 is known as a malignant lesion. To eliminate ambiguity in lesion specimens, bronchial lesions with a value of 3 in malignancy are deleted. Separate software NBIA retriever is used for the conversion of tcia format data to DICOM image data which can be used for further processing. The detailed description of the datasets used for testing is presented in the tcia website (43).

4.2 Experiment Details

The whole experiment is carried out in the Intel I7CPU with 2GB NVIDIA GeForce K+10 GPU, 16GB RAM, 3.0 GHZ with 2TB HDD. The proposed architecture is implemented using Tensorflow 1.8 with Keras API. All the programs are implemented in the anaconda environment with python 3.8 programming.

4.3 Performance Metrics and Evaluation

The proposed architecture implements the six CNN layers for the better classification of cancer cells in lung images. **Table 2** depicts the partitioned datasets used for preparation and analysis the network.

Various metrics such as accuracy, sensitivity, specificity, recall, and f1-score are calculated. The following are the mathematical expressions for calculating the metrics used for evaluating the proposed architecture.

$$Accuracy = \frac{True\ \ Positive + True\ \ Negative}{True\ \ Positive + True\ \ Negative + False\ \ Positive + False\ \ Negative} \tag{21}$$

$$Sensitivity\ \ or\ \ Recall = \frac{True\ \ Positive}{True\ \ Positive + False\ \ Negative} * 100 \tag{22}$$

$$Specificity = \frac{True\ \ Negative}{True\ \ Negative + False\ \ Positive} \tag{23}$$

$$Precision = \frac{True\ \ Negative}{True\ \ Positive + False\ \ Positive} \tag{24}$$

TABLE 2 | Total Number of Datasets (After Augmentation).

Sl.no	Total Number of Images	Training Data (%)	Testing Data (%)
1	78090	80	20

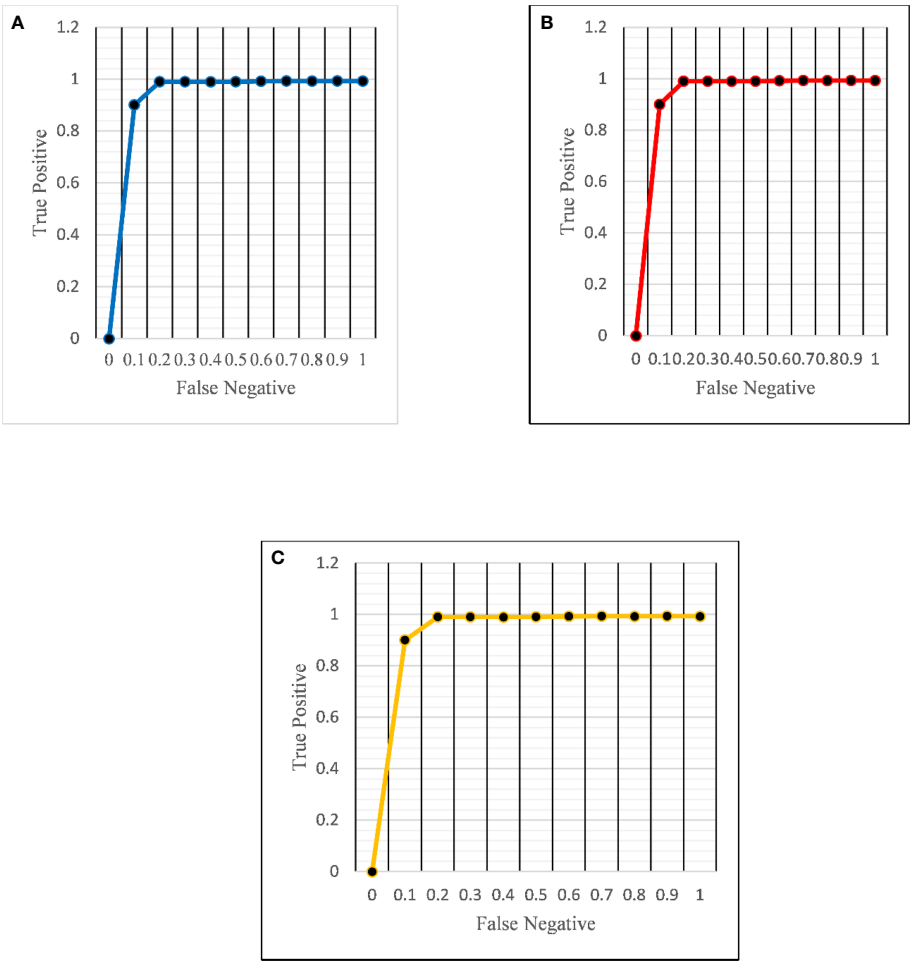


FIGURE 6 | ROC curves for the proposed architecture in detecting **(A)** normal **(B)** benign and **(C)** malignant images.

CLASSES	BEGNIN	MALIGNANT	NORMAL
BEGNIN	**295**	**01**	**04**
MALINGNANT	**01**	**296**	**03**
NORMAL	**01**	**03**	**296**

FIGURE 7 | Confusion matrix for the proposed architecture using 900 random tested images.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (25)$$

4.4 Results and Discussion

This section highlights the validation results obtained through proposed tumor predictor along with other depth networks. The validation testing data have been segregated into four distinct folds (i) confusion matrix and (ii) ROC for the first iteration. In the next fold, the projected design is compared with the other prevailing transfer learning models such as convolutional neural network (CNN), Resnets-100, Resnets-150, InceptionV3, Google-Net, Mobile-Net, and Densenet-169 by computing the diverse performance metrics as mentioned in **Table 4**. The proposed algorithm is tested with the random 900Lung CT (50% benign, 50% normal, and 50% malignant) scan images in order to overcome the imbalance problems.

The ROC curve (**Figure 6**) and the confusion matrix (**Figure 7**) of the proposed framework in detecting the categories of CT scan lung Images. **Tables 3–5** highlight the performance obtained through presented framework that is associated with other prevailing algorithms. From **Table 3**, it is found that the suggested algorithm has shown the accuracy of 98.95% with 98.85% sensitivity, 98.76% precision, and high f1score of 98.85% in detecting the normal, benign, and malignant CT images. A similar performance is found in **Table 4** in detecting images of malignancy. **Tables 3–5** show that fusion of saliency with capsule and optimized transfer learning optimized has shown the better detection ratio using the presented network than the traditional methods. **Figures 8–10**

TABLE 3 | Different deep learning architectures' performance such as accuracy, sensitivity, specificity, precision, and recall in predicting normal tissue in lung CT images.

Algorithm Details	Performance Metrics				
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)
CNN	80.2	78.5	0.0224	78.4	78.3
Resnets-100	81.5	80.5	0.0020	81.3	81.2
Resnets-150	86,2	86.0	0.0142	85.7	85
Inception V3	88.78	87.67	0.013	84.3	83.5
Mobile Nets	86.5	85.6	0.0015	85.2	85.0
Densenet-169	85.54	84.67	0.00167	84.5	84.6
SegCaps	91.0	90.8	0.0010	90.6	90.7
Proposed Framework	98.95	98.85	0.0010	98.75	98.85

TABLE 4 | Different deep learning architectures' performance such as accuracy, sensitivity, specificity, precision, and recall in predicting benign tissue in lung CT images.

Algorithm Details	Performance Metrics				
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)
CNN	78	77.6	0.0224	77	76.5
Resnets-100	81.44	81.45	0.0019	80.8	81.2
Resnets-150	86.21	85.0	0.0150	86.9	86.3
Inception V3	89.28	88.623	0.0127	88.4	87.9
Mobile Nets	85.32	84.5	0.00156	85.9	84.75
Google nets	86.57	85.8	0.00145	86.9	84.89
SegCaps	91.2	91.8	0.0090	91.3	90.67
Proposed Framework	98.95	98.85	0.0015	98.9	98.89

TABLE 5 | Different deep learning architectures' performance such as accuracy, sensitivity, specificity, precision, and recall in predicting malignant cancer in lung CT images.

Algorithm Details	Performance Metrics				
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)
CNN	80	78	0.0224	77.7	78.2
Resnets-100	81.5	80.5	0.0020	81.3	81.3
Resnets-150	86.32	86.0	0.0142	85.7	85
Inception V3	88.78	87.67	0.013	84.3	83.5
Mobile Nets	86.5	85.6	0.0015	85.2	85.0
Google nets	83.784	82.9	0.0018	81.9	81.2
SegCaps	92.0	92.83	0.0080	91.0	91.6
Proposed Framework	98.95	98.85	0.0010	98.75	98.85

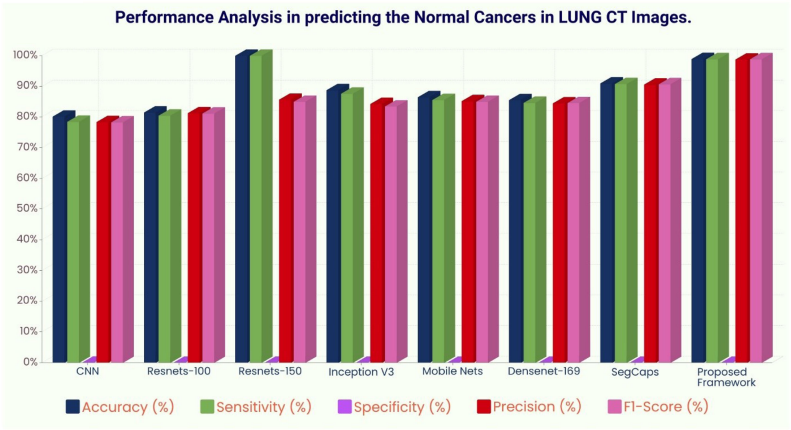


FIGURE 8 | Performance analysis in predicting normal tissue in lung CT images.

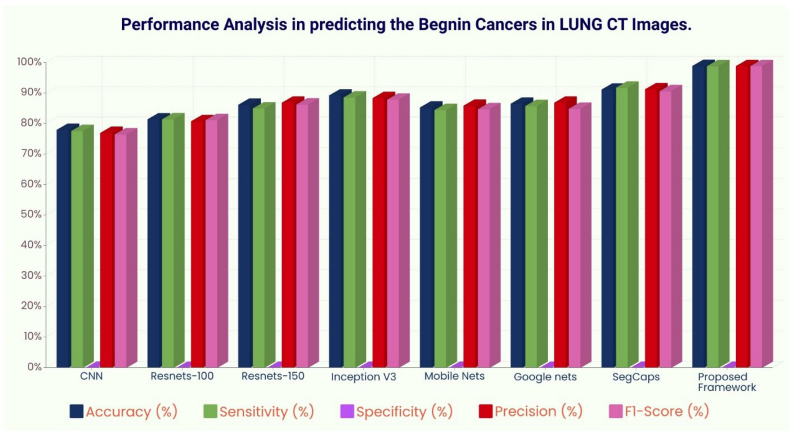


FIGURE 9 | Performance analysis in predicting benign tissue in lung CT images.

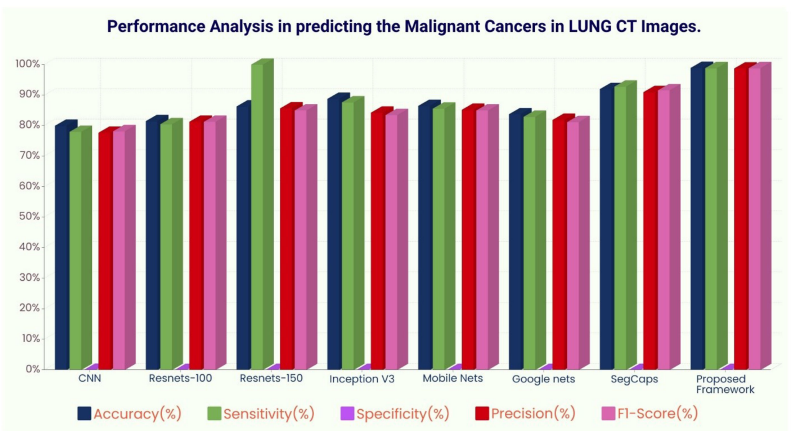


FIGURE 10 | Performance analysis in predicting malignant cancer in lung CT images.

represent the performance analysis in predicting the normal, benign, and malignant cancer in lung CT images.

5 CONCLUSION

This research goal is to detect and classify malignant and benign cancer cells using CT scan lung images. To detect the location of cancer cells, this work uses the capsule-based saliency segmentation and transfer learning-based feature extraction. Furthermore, the proposed architecture employs the whale-based classification layers to achieve better accuracy. Tensorflow 1.8 tool with Keras API has been used to evaluate the presented tumor detection approach, and various performance metrics such as accuracy, precision, recall, specificity, and f1-score are calculated and analyzed. The experimental results show that the proposed architecture has achieved the best results associated with other standard architectures and obtained the best peak results. In the future, more vigorous testing is required using larger real-time clinical datasets. Additionally, the proposed algorithm needs improvisation in terms of computational complexity which will play a significant role in the analysis and identification of tumor cells as per radiologists' perspective more accurately in future.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.886739/full#supplementary-material

REFERENCES

1. Bharati S, Podder P, Mondal R, Mahmood A, Raihan-Al-Masud M. Comparative Performance Analysis of Different Classification Algorithm for the Purpose of Prediction of Lung Cancer. In: *International Conference on Intelligent Systems Design and Applications*. Vellore, India: Springer (2018). p. 447–57.
2. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and Mutation Prediction From non–Small Cell Lung Cancer Histopathology Images Using Deep Learning. *Nat Med* (2018) 24:1559–67. doi: 10.1038/s41591-018-0177-5
3. Nie L, Wang M, Zhang L, Yan S, Zhang B, Chua TS. Disease Inference From Health-Related Questions via Sparse Deep Learning. In: *IEEE Transactions on Knowledge and Data Engineering* (2015). p. 2107–19.
4. Nie L, Zhang L, Yang Y, Wang M, Hong R, Chua TS. Beyond Doctors: Future Health Prediction From Multimedia and Multimodal Observations. In: *Proceedings of the 23rd ACM International Conference on Multimedia*. Brisbane, Australia: ACM (2015). p. 591–600.
5. Sun W, Zheng B, Qian W. Computer Aided Lung Cancer Diagnosis With Deep Learning Algorithms. *Med Imaging 2016: Computer-aided Diag (SPIE)* (2016) 9785:241–8. doi: 10.1117/12.2216307
6. Zhou ZH, Jiang Y, Yang YB, Chen SF. Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles. *Artif Intell Med* (2002) 24(1):25–36. doi: 10.1016/S0933-3657(01)00094-X
7. Dhaware BU, Pise AC. Lung Cancer Detection Using Bayasein Classifier and Fcm Segmentation. In: *International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*. Pune, India: IEEE (2016). p. 170–4.
8. da Silva GLF, Carvalho FA, AC S, Paiva A, Gattass M. Taxonomic Indexes for Differentiating Malignancy of Lung Nodules on Ct Images. *Res Biomed Eng* (2016) 32(3):263–72. doi: 10.1590/2446-4740.04615
9. Revathi A, Kaladevi R, Ramana K, Jhaveri RH, Rudra Kumar M, Sankara Prasanna Kumar M. Early Detection of Cognitive Decline Using Machine Learning Algorithm and Cognitive Ability Test. *Secur Communicat Networks* (2022) 2021:1–13. doi: 10.1155/2022/4190023
10. Reddy GT, Khare N. An Efficient System for Heart Disease Prediction Using Hybrid Ofbat With Rule-Based Fuzzy Logic Model. *J Circuit Syst Comput* (2017) 26(4):1750061. doi: 10.1142/S021812661750061X
11. Obulesu O, Kallam S, Dhiman G, Patan R, Kadiyala R, Raparthi Y, et al. Adaptive Diagnosis of Lung Cancer by Deep Learning Classification Using Wilcoxon Gain and Generator. *J Healthcare Eng* (2021) 2021:1–13. doi: 10.1155/2021/5912051
12. Deepa N, Prabadevi B, Maddikunta PK, Gadekallu TR, Baker T, Khan MA, et al. An Ai-Based Intelligent System for Healthcare Analysis Using Ridge-Adaline Stochastic Gradient Descent Classifier. *J Supercomputing* (2021) 27:1998–2017. doi: 10.1007/s11227-020-03347-2
13. Shitharth S, Mohammad GB, Ramana K, Bhaskar V. Prediction of Covid-19 Wide Spread in India Using Time Series Forecasting Techniques. (2021). doi: 10.21203/rs.3.rs-354432/v1
14. Mubashar A, Asghar K, Javed AR, Rizwan M, Srivastava G, Gadekallu TR, et al. Storage and Proximity Management for Centralized Personal Health Records Using an Ipfs-Based Optimization Algorithm. *J Circuit Syst Comput* (2022) 31(1):2250010. doi: 10.1142/S0218126622500104
15. Park SC, Tan J, Wang X, Lederman D, Leader JK, Kim SH, et al. Computer-Aided Detection of Early Interstitial Lung Diseases Using Low-Dose Ct Images. *Phys Med Biol* (2011) 56(4):1139. doi: 10.1088/0031-9155/56/4/016
16. Song Q, Zhao L, Luo X, Dou X. Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images. *J Healthcare Eng* (2017) 2017:1–7. doi: 10.1155/2017/8314740
17. Ignatious S, Joseph R. Computer Aided Lung Cancer Detection System. In: *Global Conference on Communication Technologies (GCCT)*. Thuckalay, India: IEEE (2015). p. 555–8.
18. De Bruijne M. Machine Learning Approaches in Medical Image Analysis: From Detection to Diagnosis. *Med Image Anal* (2016) 33:94–7. doi: 10.1016/j.media.2016.06.032
19. Jindal A, Aujla GS, Kumar N, Chaudhary R, Obaidat MS, You I. Sedative: Sdn-Enabled Deep Learning Architecture for Network Traffic Control in Vehicular Cyber-Physical Systems. In: IEEE network (2018). p. 66–73.
20. Nalepa J, Kawulok M. Selecting Training Sets for Support Vector Machines: A Review. *Artif Intell Rev* (2019) 52:857–900. doi: 10.1007/s10462-017-9611-1
21. Ganesan N, Venkatesh K, Rama M, Palani AM. Application of Neural Networks in Diagnosing Cancer Disease Using Demographic Data. *Int J Comput Appl* (2010) 1(26):76–85. doi: 10.5120/476-783
22. Kasinathan G, Jayakumar S, Gandomi AH, Ramachandran M, Fong SJ, Patan R. Automated 3-D Lung Tumor Detection and Classification by an Active Contour Model and Cnn Classifier. *Expert Syst Appl* (2019) 134:112–9. doi: 10.1016/j.eswa.2019.05.041
23. Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng* (2017) 19:221–48. doi: 10.1146/annurev-bioeng-071516-044442

24. Suzuki K. Overview of Deep Learning in Medical Imaging. *Radiol Phys Technol* (2017) 10(3):257–73. doi: 10.1007/s12194-017-0406-5
25. Bharati S, Podder P, Mondal MRH. Hybrid Deep Learning for Detecting Lung Diseases From X-Ray Images. *Inf Med Unlocked* (2020) 20:100391. doi: 10.1016/j.imu.2020.100391
26. Tajbakhsh N, Suzuki K. Comparing Two Classes of End-to-End Machine-Learning Models in Lung Nodule Detection and Classification: Mtanns vs. Cnns. *Pattern Recognit* (2017) 63:476–86. doi: 10.1016/j.patcog.2016.09.029
27. Gu Y, Lu X, Yang L, Zhang B, Yu D, Zhao Y, et al. Automatic Lung Nodule Detection Using a 3d Deep Convolutional Neural Network Combined With a Multi-Scale Prediction Strategy in Chest Cts. *Comput Biol Med* (2018) 103:220–31. doi: 10.1016/j.compbiomed.2018.10.011
28. Sahu P, Yu D, Dasari M, Hou F, Qin H. A Lightweight Multi-Section Cnn for Lung Nodule Classification and Malignancy Estimation. *IEEE J Biomed Health Inf* (2018) 23(3):960–8. doi: 10.1109/JBHI.2018.2879834
29. Bansal G, Chamola V, Narang P, Kumar S, Raman S. Deep3dscan: Deep Residual Network and Morphological Descriptor Based Framework for Lung Cancer Classification and 3d Segmentation. *IET Imag Process* (2020) 14 (7):1240–7. doi: 10.1049/iet-ipr.2019.1164
30. Jothi G, Inbarani HH. Soft Set Based Feature Selection Approach for Lung Cancer Images. *ArXiv Preprint ArXiv* (2012) 1212.5391.
31. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Trans Med Imaging* (2016) 35 (5):1207–16. doi: 10.1109/TMI.2016.2535865
32. Kasinathan G, Jayakumar S. Cloud-Based Lung Tumor Detection and Stage Classification Using Deep Learning Techniques. *BioMed Res Int* (2022) 2022:1–17. doi: 10.1155/2022/4185835
33. Jakimovski G, Davcev D. Using Double Convolution Neural Network for Lung Cancer Stage Detection. *Appl Sci* (2019) 9(3):427. doi: 10.3390/app9030427
34. Yu H, Zhou Z, Wang Q. Deep Learning Assisted Predict of Lung Cancer on Computed Tomography Images Using the Adaptive Hierarchical Heuristic Mathematical Model. In: . IEEE Access (2020). p. 86400–10.
35. Chauhan D, Jaiswal V. An Efficient Data Mining Classification Approach for Detecting Lung Cancer Disease. In: *International Conference on Communication and Electronics Systems (ICCES)*. Coimbatore, India: IEEE (2016). p. 1–8.
36. Shakeel PM, Burhanuddin M, Desa MI. Automatic Lung Cancer Detection From Ct Image Using Improved Deep Neural Network and Ensemble Classifier. *Neural Comput Appl* (2020) 34:1–14. doi: 10.1007/s00521-020-04842-6
37. Pei SC, Hsiao YZ. Spatial Affine Transformations of Images by Using Fractional Shift Fourier Transform. *IEEE Int Symposium Circuit Syst (ISCAS) (IEEE)* (2015) 1586–9. doi: 10.1109/ISCAS.2015.7168951
38. Banerjee S, Mitra S, Shankar BU, Hayashi Y. A Novel Gbm Saliency Detection Model Using Multi-Channel Mri. *PLoS One* (2016) 11(1):e0146388. doi: 10.1371/journal.pone.0146388
39. Takács P, Manno-Kovacs A. Mri Brain Tumor Segmentation Combining Saliency and Convolutional Network Features. In: *International Conference on Content-Based Multimedia Indexing (CBMI)*. La Rochelle, France: IEEE (2018). p. 1–6.
40. Huang GB, Zhu QY, Siew CK. Extreme Learning Machine: Theory and Applications. *Neurocomputing* (2006) 70(1–3):489–501. doi: 10.1016/j.neucom.2005.12.126
41. Wang B, Huang S, Qiu J, Liu Y, Wang G. Parallel Online Sequential Extreme Learning Machine Based on Mapreduce. *Neurocomputing* (2015) 149:224–32. doi: 10.1016/j.neucom.2014.03.076
42. Mukherjee A, Chakraborty N, Das BK. Whale Optimization Algorithm: An Implementation to Design Low-Pass Fir Filter. In: *Innovations in Power and Advanced Computing Technologies (I-PACT)*. Vellore, India: IEEE (2017). p. 1–5.
43. *Dataset. Lidc-Idri - the Cancer Imaging Archive (Tcia) Public Access - Cancer Imaging Archive Wiki* (2021). Available at: https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI (Accessed on 02/12/2021).

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 | Frontiers in Oncology

ORIGINAL RESEARCH
published: 07 July 2022
doi: 10.3389/fonc.2022.894978

High-Accuracy Oral Squamous Cell Carcinoma Auxiliary Diagnosis System Based on EfficientNet

Ziang Xu[1†], *Jiakuan Peng*[1†], *Xin Zeng*[1*], *Hao Xu*[1*] *and Qianming Chen*[2]

[1] State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, Chinese Academy of Medical Sciences Research Unit of Oral Carcinogenesis and Management, West China Hospital of Stomatology, Sichuan University, Chengdu, China, [2] Key Laboratory of Oral Biomedical Research of Zhejiang Province, Affiliated Stomatology Hospital, Zhejiang University School of Stomatology, Hangzhou, China

It is important to diagnose the grade of oral squamous cell carcinoma (OSCC), but the current evaluation of the biopsy slide still mainly depends on the manual operation of pathologists. The workload of manual evaluation is large, and the results are greatly affected by the subjectivity of the pathologists. In recent years, with the development and application of deep learning, automatic evaluation of biopsy slides is gradually being applied to medical diagnoses, and it has shown good results. Therefore, a new OSCC auxiliary diagnostic system was proposed to automatically and accurately evaluate the patients' tissue slides. This is the first study that compared the effects of different resolutions on the results. The OSCC tissue slides from The Cancer Genome Atlas (TCGA, n=697) and our independent datasets (n=337) were used for model training and verification. In the test dataset of tiles, the accuracy was 93.1% at 20x resolution (n=306,134), which was higher than that at 10x (n=154,148, accuracy=90.9%) and at 40x (n=890,681, accuracy=89.3%). The accuracy of the new system based on EfficientNet, which was used to evaluate the tumor grade of the biopsy slide, reached 98.1% [95% confidence interval (CI): 97.1% to 99.1%], and the area under the receiver operating characteristic curve (AUROC) reached 0.998 (95%CI: 0.995 to 1.000) in the TCGA dataset. When verifying the model on the independent image dataset, the accuracy still reached 91.4% (95% CI: 88.4% to 94.4%, at 20x) and the AUROC reached 0.992 (95%CI: 0.982 to 1.000). It may benefit oral pathologists by reducing certain repetitive and time-consuming tasks, improving the efficiency of diagnosis, and facilitating the further development of computational histopathology.

Keywords: oral squamous cell carcinoma, computational histopathology, deep learning, EfficientNet, auxiliary diagnosis

OPEN ACCESS

Edited by:
Wei Wei,
Xi'an University of Technology, China

Reviewed by:
Jakub Nalepa,
Silesian University of Technology, Poland
Shankargouda Patil,
Jazan University, Saudi Arabia

****Correspondence:***
Xin Zeng
zengxin22@163.com
Hao Xu
hao.xu@scu.edu.cn

[†]These authors have contributed equally to this work

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

Received: *12 March 2022*
Accepted: *07 June 2022*
Published: *07 July 2022*

Citation:
Xu Z, Peng J, Zeng X, Xu H and Chen Q (2022) High-Accuracy Oral Squamous Cell Carcinoma Auxiliary Diagnosis System Based on EfficientNet. Front. Oncol. 12:894978. doi: 10.3389/fonc.2022.894978

INTRODUCTION

Oral squamous cell carcinoma (OSCC) accounted for more than 377,713 new cancers and 177,757 deaths in 2020. The 5-year survival rate of patients in the earlier stage is about 55%–60%, while that of patients in advanced stages drops to 30%–40% (1, 2). The histological 'grade' of a malignant tumor is an index to describe its malignant degree. The current WHO classification of head and

neck tumors is based on the simple grading system of the Broders standard (3), which is divided into three types: well-differentiated, moderately-differentiated, and poorly-differentiated. Later, more complex grading systems were suggested by Jakobsson et al. (4) and Anneroth et al. (5). The current way of diagnosing grades is still relying on the manual reading of slides by pathologists, which is a heavy workload, and the subjectivity of the pathologists greatly affects the diagnosis results, so it is valuable to develop an automatic auxiliary diagnosis system (6, 7).

Deep learning (DL) refers to the class of machine learning methods. It allows computers to learn complex concepts through relatively simple concepts (8). Since DL performs well in image interpretation and classification problems (9, 10), it has been widely used in medical image analysis tasks, especially in survival prediction and computational histopathology (11, 12), as well as classification of histological phenotypes (13).

Meanwhile, there have been several studies about the application of deep learning on the diagnosis of OSCC (10). For example, one study could judge whether the tissue is malignant (14, 15), but it could not determine the severity of the tumor tissue. Das et al. used only the images of the epithelial part to judge the grade of the tissue while the accuracy was not high enough (16). These studies have confirmed the application of deep learning in the field of OSCC, but there are still some imperfections, such as the lack of accuracy. Therefore, we carried out an automatic OSCC auxiliary diagnosis system, which was called EfficientNet-based Computational Histopathology of OSCC (ECHO). In this study, The Cancer Genome Atlas Program (TCGA) dataset was used to train and test the model (17). By comparing the performance of different convolutional neural networks (CNNs), the best performing one was be selected, and the performance was verified by using our independent dataset.

MATERIALS AND METHODS

The workflow of this study is shown in **Figure 1**. First the slides were cut into tiles for training and testing (18), then the dataset was balanced by decomposing a multiclass imbalanced dataset into a binary problem (19), and the tiles with a blank area more than 50% were removed to ensure that each tile contains valid information. Then the preprocessed dataset was divided into training set, validation set, and test set.

Secondly, three CNNs, EfficientNet b0 (20), ShuffleNetV2 (21), and ResNeXt_18 (22), were trained at different resolutions, and the most accurate CNN with the best performing resolution was selected for the further analysis.

Finally, we tested the performance of ECHO on the external dataset, the OSCC tissue microarrays (TMA). There are differences in the image forms between TCGA and TMA, but both contain valid information, so we used TMA for external validation to prove that the model has high accuracy when dealing with various types of images.

The complete and detailed workflow is described below:

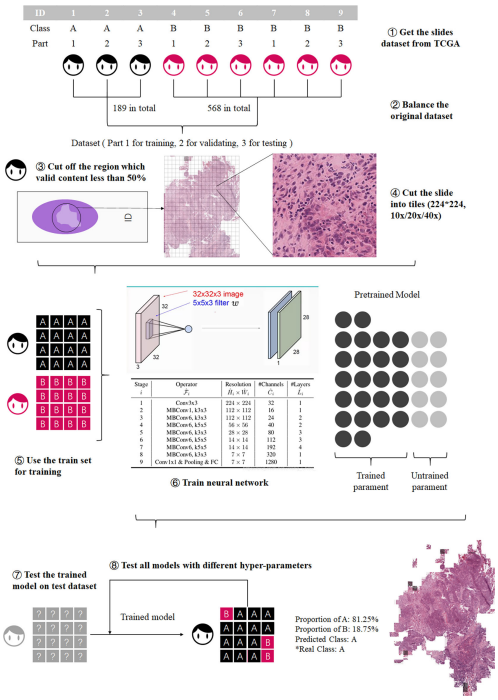


FIGURE 1 | Process of model training. The training process was divided into three steps: preprocessing the datasets, model training, and model testing.

Data Resource and Data Preprocessing

The image datasets include the TCGA OSCC image dataset and our independent TMA images dataset. We downloaded 757 whole-slide images of OSCC from the official website of TCGA in 2019 as the original dataset of TCGA. The TCGA dataset was used for model training and testing, and the TMA dataset was used for external verification. TCGA classifies OSCC into grade I (G1, well-differentiated), grade II (G2, moderately differentiated), grade III (G3, poorly differentiated), and grade IV (G4, undifferentiated or anaplastic) (23).

For the TCGA dataset, considering that the number of G4 slides was too small, and the imbalance of the dataset would seriously affect the training result, the G1 and G2 were combined as the well-differentiated group; G3 and G4 were combined as the poorly-differentiated group. There are 757 slides in the TCGA dataset, 568 slides in the G1-G2 group, and 189 slides in the G3-G4 group. These slides were cut into 224*224 pixel tiles (18, 20) at 10x, 20x, and 40x resolution, respectively, and the tiles with a blank area more than 50% were removed. The number of slides in the G3-G4 group and the G1-G2 group was quite different, which would adversely affect the results (19). We used the number of slides of the minimum class as the standard number, N0. Then calculated the ratio of N0 to the number of slides of each other class Nk. The ratio, Rk, was used to balance the tiles dataset. The tiles set of each class was multiplied by a coefficient Rk, as the final number of tiles for each class. The tiles of major classes were randomly removed until the numbers of

tiles reached the final number. Then they were used as the preprocessed TCGA dataset.

The preprocessed data set was divided into three datasets: training set, validation set, and test set, which account for 60%, 20%, and 20% (24). These datasets were used for training and testing CNNs, and each CNN had two outputs: the possibility of G1-G2 and the possibility of G3-G4.

Additionally, we collected the TMAs dataset from the West China Hospital of Stomatology (Chengdu, China) and this study was approved by the ethics committee of the West China Hospital of Stomatology. The TMAs included 337 available slides of patients recruited from 2004 to 2014 who had received informed consent in this study. In the TMA dataset, according to the degree of tumor differentiation, the histological grades were divided into 1 (high differentiation), 2 (moderate differentiation), and 3 (low differentiation). Grades 1 and 2 were combined as the well-differentiated group (corresponding to G1-G2), and grade 3 was considered as the poorly-differentiated group (corresponding to G3-G4). The TMAs dataset was used for the external verification of the best performing CNN model chosen by above training steps. Moreover, they were used to verify the generalizability of the model.

CNNs and Resolutions

For the consideration of training speed, training accuracy, and estimated time, we used three CNNs: EfficientNet b0, ShuffleNetV2, and ResNeXt_18. The performance of EfficientNet has shown great advantages since its inception, The accuracy and operation speed of EfficientNet is much faster than other networks (20), and it is often used as a comparison standard by the newly proposed CNNs (25, 26). ResNet is a classic neural network that is widely used in many fields and has good performance (22), so we chose ResNet as a benchmark to compare other CNNs. ShuffleNet is lightweight and computationally can be used on mobile devices (21). The reasons for choosing three models for this study was not only to select a better performing CNN, but also to try out the practicality of lightweight models.

These CNNs were trained on three different resolutions, and we selected the best CNN by comparing their accuracy and AUC on the tiles in test sets. In order to compare the effects of different resolutions on model training time and model accuracy, we decided to use 10x, 20x, and 40x resolution to train the three models separately. Finally, the model was used to evaluate the slides at the corresponding resolution. The resolution with better performance was selected.

Model Training and Selection

In order to compare and select the best model more efficiently, all slides were cut into tiles which were used to train and test models. The label of a tile was determined by the slide which it came from. When the training was completed, the accuracy and AUC of the model on the tile dataset was used to evaluate the performance of the model, and in this way, the best model for the next study was selected.

Three networks were trained on the 10x resolution (154,148 tiles in all, 74,977 in G1-G2 group, 78,171 in G3-G4 group), 20x resolution (306,134 tiles in all, 149,826 in G1-G2 group, 156,308 in G3-G4 group), and 40x resolution (890,681 tiles in all, 469,751 in G1-G2 group, 420,930 in G3-G4 group). **Table 1** shows the information of the datasets used for training and testing.

During the training process, we observed the accuracy of each model at each epoch and drew an epoch-accuracy curve. When the epoch was low, the accuracy would also be low due to insufficient image feature extraction; when the epoch was high, the accuracy of the model would decrease due to over-fitting. When the epoch was around 60-70, the accuracy of the model would be high and stable (27). An epoch of about 60-70 would make the accuracy of the model high and stable, so the epoch was set to 80 and the accuracy of models was compared at each epoch.

Other hyperparameters are as follows: batch size: 80, learning rate: 0.0005, optimization algorithm: Adam, activation function: Swish.

The Construction of ECHO

The best model and resolution selected in the above process was used to construct ECHO. Different from the above test process, the dataset here was composed of all slides. The accuracy on the slide dataset was used to evaluate the application value of ECHO.

The purpose of ECHO is to give the differentiation level of the input slide. The workflow mainly included two steps. First, ECHO cut the input slide into 224*224-pixel tiles and used the best model to give each tile a label of G1-G2 or G3-G4. In the second step, ECHO counted the tags of all tiles and used tags that account for more than 50% as the result of the slide. If the results given were consistent with the actual clinical labels, then ECHO's prediction was considered accurate. The accuracy of ECHO's predictions on all slides was used to evaluate the performance and application value of ECHO.

Five-Classes Expansion of ECHO

Based on the best CNN selected by the above research and the most suitable resolution, we developed a five-class model of ECHO, which can assist the results of binary classification. The

TABLE 1 | Datasets used for training and testing.

Dataset	G1-G2			G3-G4		
	10x	20x	40x	10x	20x	40x
Training	44,987	89,896	281,851	46,903	93,784	252,558
Validation	14,995	29,965	93,950	15,634	31,262	84,186
Test	14,995	29,965	93,950	15,634	31,262	84,186

We divided the tiles under each resolution into training, validation, and test sets in a ratio of 6:2:2.

five classes are as follows: normal organization, G1, G2, G3, and G4. The data preprocessing and training process was the same as before. The effect of this model was evaluated by the confusion matrix and accuracy.

Hardware and Software

Four NVIDIA Tesla K80 graphics cards were used, which contained a total of eight graphics processing units (GPUs). Each model was trained on a single GPU. The construction and training of the model was based on TensorFlow 2.1, and the programming language was Python 3.6.8.

Statistical Analysis

We first got the accuracy and area under the receiver operating characteristic (ROC) curve (AUC) with 95% confidence interval (CI) in test datasets, then assessed the performance of networks. The 95% CI was calculated using the bootstrap method (28). The bootstrap method uses sampling with replacement, a sample size equal to the original sample size, and computes the required statistics. This process was repeated 100 times, and confidence intervals were estimated based on the statistics calculated for these 100 times. In our study, when calculating the accuracy, the original sample refers to whether the class of each whole-slide image (WSI) was judged correctly. When calculating AUC, the raw sample was the ratio of the actual label of each WSI to each class computed by the machine. The accuracy and AUC in test dataset were primary criteria for evaluation. All the statistical analysis was also performed with Python 3.6.8.

RESULT

Model Comparison

TCGA dataset was used to train and compare the performance of different CNNs and resolutions, then the best performing CNN and resolution were selected.

We first determined the epoch to be selected. In general, the accuracy of each model increases as the epoch grows. When the epoch reaches 50-60, the accuracy of the model has increased very little, and the difference was small. Therefore, the maximum value of the epoch was set to 80 and the accuracy of models was compared at each epoch. **Table 2** shows the epoch value of different models at three resolutions, and **Table 3** shows the accuracy and AUC at corresponding epoch values.

TABLE 2 | The epoch value when the three models have the highest test accuracy in the three resolutions.

Model	Resolution		
	10x	20x	40x
EfficientNet	70	70	70
ShuffleNet	72	54	72
Resnet	76	78	64

We choose the epoch value with the highest accuracy as the parameter of the corresponding model. In the next test, we use the corresponding model to evaluate the effect.

Then we evaluated and selected the CNN and the resolution. **Figures 2A–C** shows the ROC of the three models at their best performance at three resolutions. Except for the 10x ResNet, the AUC of other models are all above 0.95, which maintained a high level. The highest among them was the EfficientNet at 20x. **Figure 2D** shows the accuracy of each CNN at different resolutions with 95% CI. The CNN with the highest accuracy was the EfficientNet at 20x, which accuracy reached 0.931 (95% CI: 0.920 to 0.942).

Because EfficientNet has better performance at 20x resolution both in accuracy and AUC, our next research will be based on this model, which is called ECHO.

We also compared the calculation speed of different models. **Figure 3** shows the evaluation time and evaluation results of the three models for random whole-slide images. ResNet had the fastest computing speed. ShuffleNet took about 1.5 times that of ResNet, and EfficientNet took about 2 times that of ResNet. For larger slides, the time difference between the fastest and slowest models could be more than 60s, but the difference was acceptable in clinical practice.

Test on the Whole-Slide Images

The ECHO was used to test on whole-slide image datasets of the TCGA dataset. If the proportion of G3-G4 tiles is more than 50%, the machine will judge the slide as G3-G4. If not, the slide will be classified as G1-G2. We tested a total of 697 slides, the sensitivity reached 98.3% (176/179) and the specificity reached 98.0% (508/518). The total accuracy reached 98.1% (95%CI: 97.1% to 99.1%), and the AUC reached 0.998 (95%CI: 0.995 to 1.000, **Figure 4A**). It proves that the ECHO has very high accuracy on the internal test set. The processing and classification time of a single WSI is about 30-60s (based on the size of the WSI, shown in **Figure 3B**)

Verification on the External Dataset

The TMAs image dataset was used to verify the performance and external use of the ECHO. The TMAs dataset has a total of 337 slides. The slide dyeing method of TMA was different from that of TCGA, so the color characteristics of the image are different. Meanwhile, due to the different sources of patients, the histological structure of the tumor may also be slightly different. Both TCGA and TMA have the tissue which contains sufficient content for pathological diagnosis, and the processing methods are also consistent, so we used TMA to validate the mode to prove that the model has high accuracy when dealing with various types of images. Due to the differences between the TMAs dataset and the TCGA dataset, it was appropriate to use the TMAs dataset to verify the performance and external applicability of ECHO.

The accuracy reached 91.4% (95% CI: 88.4% to 94.4%), and the AUC was 0.992 (95%CI: 0.982 to 1.000). The ROC curve is shown in **Figure 4B**. This proves that the ECHO still has a good effect when faced with test subjects whose sources are quite different and have good applicability.

Visualization of ECHO

When the prediction of each block was finished, the system made a restored slide picture according to the prediction result and the

TABLE 3 | Accuracy and AUC of different CNNs and corresponding resolutions.

Model	Resolution		
	10x	20x	40x
EfficientNet	90.9 (0.97)	93.1 (0.98)	89.3 (0.96)
ShuffleNet	88.9 (0.96)	91.2 (0.96)	90.1 (0.97)
Resnet	76.1 (0.89)	90.8 (0.96)	89.8 (0.96)

*Based on the best epoch value selected in **Table 2**, we tested the accuracy of different models and corresponding resolutions. The table shows the accuracy, with AUC values in parentheses. This result shows that EfficientNet at 20x resolution has the best performance.*

possibility of each block. If the predicted tile result was G1-G2, a black layer was added to the tile. We randomly selected two slides, which belong to G1-G2 and G3-G4. The result was stored through visualization, and the result is shown in **Figure 5**. **Figure 5C** shows the classification results of the five classes ECHO program. The right area shows the visualized heatmap, each tile was added color with transparency. Normal tissues, G1-G4 correspond to colorless, green, blue, yellow, and red, respectively. The detailed stored image can be seen in **Figure 1**, **2**.

Evaluation of Five Classes ECHO

We used the TCGA dataset to test the five classes ECHO. The five classes ECHO accepts a WSI as input and gives the probability of each class. In the test of 447 WSIs, 347 of them were correctly classified. We used the bootstrap method to sample 100 times to calculate the confidence interval, and the final accuracy was 77.63% (95CI: 77.25-78.01). **Figure 6** shows the confusion matrix according to the classification results. The results showed that the classification performance of different classes was different. The classification sensitivity of normal tissues, G1, and G4 were higher, reaching 92.85%, 98.27%, and 100%. The sensitivity of G3 and G2 was poor, 82.88% and 70.48%. In terms of specificity, the classification results of normal tissue and G2 were better, reaching 100% and 94.09%, G4, G3, and G1 are worse, being 77.78%, 71.31%, and 52.29%, respectively. The results show that the five classes ECHO can be used as a reference to complete some auxiliary tasks.

DISCUSSION

The visual inspection of tumor tissue under the light microscope by pathologists is the gold standard for OSCC grading This evaluation is mainly based on the pathologists' clinical pathology knowledge and skills (29). The workload is heavy, and the results are affected by subjectivity. However, the application of DL in the histopathological diagnosis would help the pathologists (12, 30). Recently, there were several studies on OSCC automatic detection. For example, to judge whether OSCC is benign or

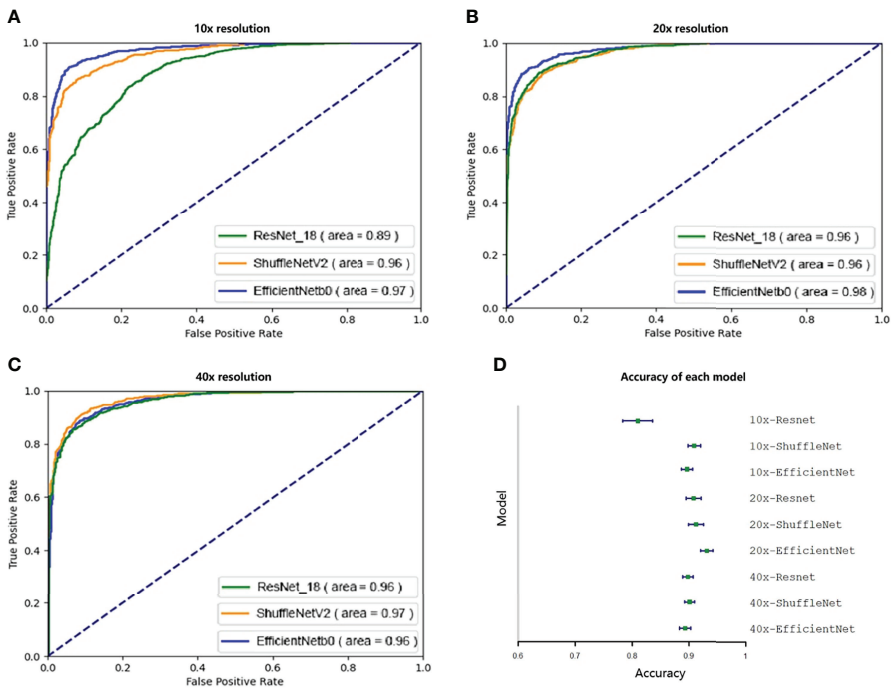


FIGURE 2 | Ninety-five percent confidence interval when testing on a validation set of 10,000 tiles. The test set was resampled and tested one hundred times using the bootstrap method, and the ROC and 95% confidence interval were calculated. **A–C** show the ROC curves of three CNNs tested at 10x, 20x, and 40x resolution. Except for the 10x ResNet, the ROC of other models were all greater than 0.95. **D** shows the accuracy of each model. Except for 10x ResNet, the accuracy of each model was similar, while the accuracy of 20x EfficientNet is slightly higher.

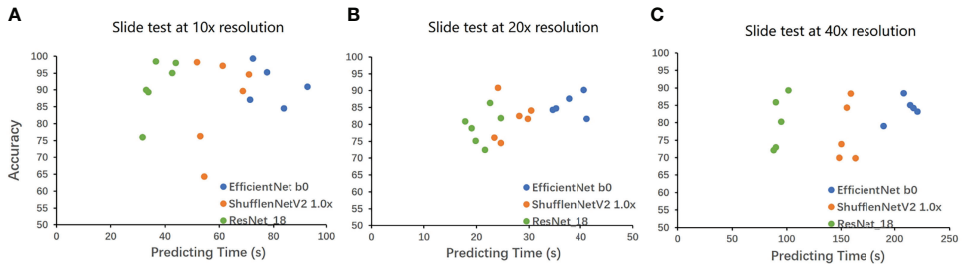


FIGURE 3 | The result on the whole-slide-image. The WSIs was cut into tiles then classified, and the proportion of the correctly classified tiles was used as the accuracy to make Figure 3. The horizontal axis is the classification time (slide cutting time is omitted), and the vertical axis is the proportion of the correct classification. **A** is the result under 10x, **B** is the result under 20x, and **C** is the result under 40x. EfficientNet requires a longer time but has higher accuracy. ResNet has a very powerful speed and a good performance in accuracy. The speed of ShuffleNet is between the two, and the accuracy is not stable.

malignant (14), CNNs were used to classify OSCC epithelial cells (16). However, these studies still have some imperfections, such as the methods used are relatively old, the data resources and evaluation indicators are single, and the accuracy is not high enough. In our study, these problems were basically solved. The ECHO achieved very high accuracy and verified the possibility of external application.

Firstly, we compared the performance of the three CNNs: EfficientNet, ShuffleNet, and ResNet. ShuffleNet is designed for mobile terminals (21), so the model has the smallest amount of parameters and the smallest size, which can be applied to lighter devices. As shown in **Figure 3**, its computing speed is at a medium level, and the highest accuracy reached 91.2% (95%CI: 89.9% to 92.5%, 20x resolution). The calculation speed of EfficientNet is the slowest in our study, but it was still faster than many CNNs (20). The accuracy of EfficientNet is the highest, which reached 93.1% (95%CI: 92.0% to 94.2%, 20x resolution). ResNet was a classic CNN that greatly alleviated the problem of overfitting (22). It has the fastest computing speed and has an accuracy of 90.8% (95CI: 89.5% to 92.1%, 20x resolution).

Secondly, we evaluated the impact of different resolutions on the experimental results. The higher resolution is helpful to improve the model's recognition and extraction of image features, but it may affect the final result due to local overfitting (31). The reason may be that the cut tiles are too small and that the important features are at the edges, so that the key information cannot be extracted. Higher resolution will also greatly increase the workload of model training and the time of slide analysis. Lower resolution can effectively improve the training and recognition speed, but it may cause a potential decrease in accuracy. Because the resolution is too small, the details of the features are not clear, resulting in poor training results. Twenty times resolution also has faster application speed and accuracy, so this resolution was chosen to apply.

In this study, we took two measures to deal with imbalanced datasets. Since the number of slides in the G4 phase was too small, less than one-tenth of that in the G2 phase or G3 phase, we chose to combine G1 and G2 as well-differentiated, and G3 and G4 as poorly-differentiated. After the merger, the imbalance problem was alleviated. In the second step, we processed the dataset by undersampling the majority of class examples.

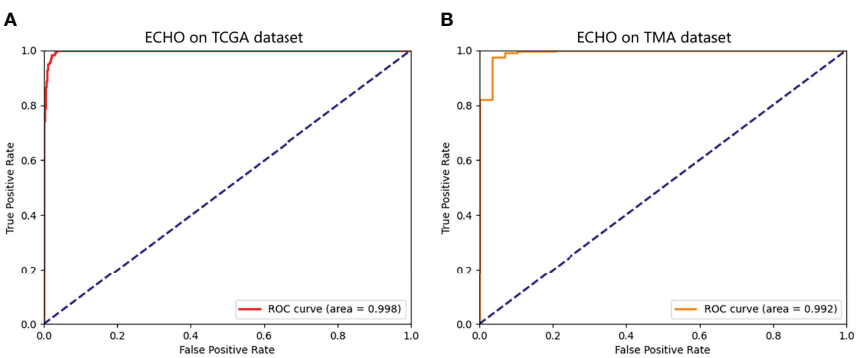


FIGURE 4 | The ROC of the predicted results of ECHO. **A** shows the ROC tested and calculated by ECHO on the test set of TCGA. It can be seen that the area under the curve is as high as 0.998, with an accuracy of 98.1% (684/697). **B** shows the ROC tested and calculated by ECHO on TMA. The test accuracy rate is 0.914 (308/337), and the area under the curve is 0.992.

FIGURE 5 | A slide image made in reverse according to the classification result. Most of the area in **A** is covered by black shadows, so this slide belongs to G1-G2. Picture **B** has almost no area covered by shadows, it belongs to G3-G4. **C** shows the classification results of the five classes ECHO program. The left area shows the input WSI preview, the middle area shows the probabilities of each category, and the right area shows the heatmap, each tile was added a color with transparency. Normal tissues, G1-G4 correspond to colorless, green, blue, yellow, and red, respectively.

To ensure that the information of each WSI can be utilized, we first cut all WSIs into tiles, and then undersampled the imbalanced tile set. The preprocessing measures we took may not be optimal, which leads to a loss of information (32). The undersampling process has produced good results, but in the next research, we will further explore better preprocessing measures (33).

There have been many studies on the machine learning application of OSCC. Mermod et al., 2020, used Random Forest (RF), linear Support Vector Machine (SVM), to judge the metastasis of squamous cell carcinoma of lymph nodes, with an accuracy of 90% (34). Romeo et al., 2020, used Naïve Bayes (NB), Bagging of NB, and K-Nearest Neighbors (KNN) to determine tumor grade with an accuracy of 92.9% (35). These researchers use more traditional machine learning methods, and there was still much room for improvement in accuracy. Ariji et al., 2019, who used deep learning methods, used CNN to evaluate lymph node metastasis, but the accuracy was only 78.1% (36). Jeyaraj & Samuel Nadar, 2019, used CNN to judge benign and malignant tumors with an accuracy of 91.4% (37). Our research is also based on CNN, which has two classification systems and five classification systems. The two-class classification system can accurately determine the tumor differentiation (high or low), and the accuracy has reached an astonishing 98.1%. The five-class classification system can judge the specific differentiation grade of the tumor and can also judge whether the tumor is malignant. The accuracy of judging whether

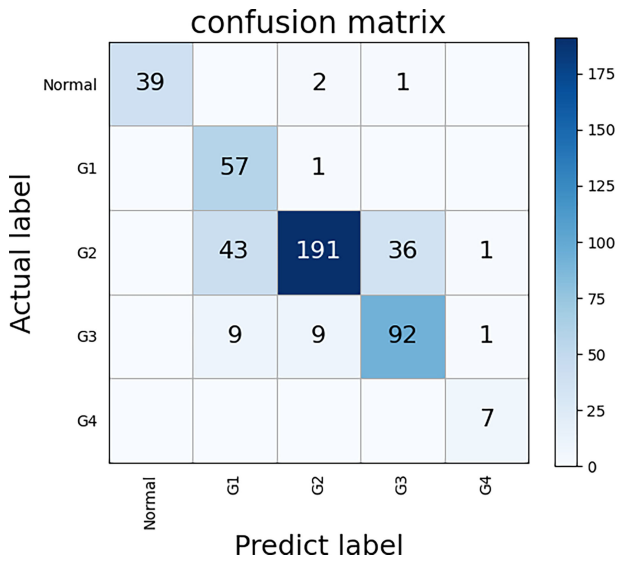


FIGURE 6 | The confusion matrix of five classes ECHO. The correct prediction results are located on a diagonal line from the upper left corner to the lower right corner of the matrix. We used the bootstrap method to sample 100 times to calculate the confidence interval, and the final accuracy was 77.63% (95CI: 77.25-78.01).

it is benign or malignant has reached 92.86% (39/42). Therefore, our study is valuable and far surpasses other current studies in accuracy.

However, our research also needs improvement. Due to the limitation of the number of samples in the dataset, that is, the number of samples of G1, G2, G3, and G4 is too imbalanced, we had to group them to balance the amount of data. In future research, we will obtain more datasets to refine the model and train the system for five classes: normal, G1, G2, G3, and G4. In addition, in the division of G1-G2 and G3-G4, the machine determines whether a slide belongs to G1-G2 or G3-G4 according to the ratio of tiles. When the proportion of G3-G4 tiles is higher than 50%, the machine will classify this slide as 'G3-G4', so 50% is the threshold for machine judgment. It has been reported that when the threshold is 50%, the sensitivity is high and the specificity is low (38). When the threshold is changed, the effect of the model will be different, and this could be further discussed in the future.

CONCLUSION

Oral squamous cell carcinoma is one of the most common head and neck tumors. It is important to determine the grade of tumor differentiation, which has a guiding role in tumor treatment and prognosis prediction. We developed two and five class systems based on CNN. The two classes system can judge whether the tumor is well differentiated or poorly differentiated. The test accuracy on the TCGA dataset reached 98.1% (n=697). The five classes system can judge whether the tissue belongs to normal tissue, G1, G2, G3, or G4. The accuracy reached 77.63%. We've also built visualization programs that can help doctors deal with some controversial slides. The system we developed can effectively reduce the workload of the pathologist and increase the efficiency and speed of the diagnostic process.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov.

ETHICS STATEMENT

This study was approved by the ethics committee of the West China Hospital of Stomatology. The TMAs included 337 available slides of patients recruited from 2004 to 2014 who had received informed consent in this study.

AUTHOR CONTRIBUTIONS

HX, QC, ZX, and JP contributed to conception and design of the study. HX, ZX, XZ, and QC organized the database. ZX, JP, and HX performed the statistical analysis. ZX and JP wrote the first draft of the manuscript, JP, XZ, QC, and HX wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was supported by grants from the National Natural Science Foundation of China (82001059).

ACKNOWLEDGMENTS

The results shown here are based partly on data generated by the TCGA Research Network. The authors would like to thank the oral pathology department of West China School of Stomatology, where all of the slide crafting tasks and diagnostic tasks were completed.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.894978/full#supplementary-material

Supplementary Figure 1 | Another WSI visualization. The differentiation grade of this tissue is G2.

Supplementary Figure 2 | A high-resolution image of the visualization results. G1 tiles are colored green, G2 tiles are colored blue, G3 tiles are colored yellow, and G4 tiles are colored red. Due to the influence of H&E staining, the actual color displayed by the G2 tile is purple.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA-Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Bejnordi BE, Veta M, van Diest PJ, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA-J Am Med Assoc* (2017) 318(22):2199–210. doi: 10.1001/jama.2017.14585
3. El-Naggar AK, Chan JK, Grandis JR, Takata T, Slootweg PJ. *WHO Classification of Head and Neck Tumours*. Lyon: International Agency for Research on Cancer (IARC (2017).
4. Jakobsson P, Eneroth C, Killander D, Moberger G, Mårtensson B. Histologic Classification and Grading of Malignancy in Carcinoma of the Larynx. *Acta Radiologica: Therapy Physics Biol* (1973) 12(1):1–8.
5. Anneroth G, Batsakis J, Luna M. Review of the Literature and a Recommended System of Malignancy Grading in Oral Squamous Cell Carcinomas. *Eur J Oral Sci* (1987) 95(3):229–49. doi: 10.1111/j.1600-0722.1987.tb01836.x
6. Diao S, Hou J, Yu H, Zhao X, Sun Y, Lambo RL, et al. Computer-Aided Pathologic Diagnosis of Nasopharyngeal Carcinoma Based on Deep Learning. *Am J Pathol* (2020) 190(8):1691–700. doi: 10.1016/j.ajpath.2020.04.008
7. Zhang Z, Chen P, McGough M, Xing F, Wang C, Bui M, et al. Pathologist-Level Interpretable Whole-Slide Cancer Diagnosis With Deep Learning. *Nat Mach Intelligence* (2019) 1(5):236–45. doi: 10.1038/s42256-019-0052-1
8. Schmidhuber J. Deep Learning in Neural Networks: An Overview. *Neural Networks* (2015) 61:85–117. doi: 10.1016/j.neunet.2014.09.003
9. Clymer D, Kostadinov S, Catov J, Skvarca L, Pantanowitz L, Cagan J, et al. Decidual Vasculopathy Identification in Whole Slide Images Using Multiresolution Hierarchical Convolutional Neural Networks. *Am J Pathol* (2020) 190(10):2111–22. doi: 10.1016/j.ajpath.2020.06.014
10. Alabi RO, Youssef O, Pirinen M, Elmusrati M, Mäkitie AA, Leivo I, et al. Machine Learning in Oral Squamous Cell Carcinoma: Current Status, Clinical Concerns and Prospects for Future-A Systematic Review. *Artif Intell Med* (2021) 115:102060. doi: 10.1016/j.artmed.2021.102060
11. Courtiol P, Maussion C, Moarii M, Pronier E, Pilcer S, Sefta M, et al. Deep Learning-Based Classification of Mesothelioma Improves Prediction of Patient Outcome. *Nat Med* (2019) 25(10):1519–25. doi: 10.1038/s41591-019-0583-3
12. Ocampo P, Moreira A, Coudray N, Sakellaropoulos T, Narula N, Snuderl M, et al. Classification and Mutation Prediction From Non-Small Cell Lung Cancer Histopathology Images Using Deep Learning. *J Thorac Oncol* (2018) 13(10):S562–2. doi: 10.1016/j.jtho.2018.08.808
13. Sheehan S, Mawe S, Cianciolo RE, Korstanje R, Mahoney JM. Detection and Classification of Novel Renal Histologic Phenotypes Using Deep Neural Networks. *Am J Pathol* (2019) 189(9):1786–96. doi: 10.1016/j.ajpath.2019.05.019
14. Rahman TY, Mahanta LB, Chakraborty C, Das AK, Sarma JD. Textural Pattern Classification for Oral Squamous Cell Carcinoma. *J Microsc-Oxford* (2018) 269(1):85–93. doi: 10.1111/jmi.12611
15. Rahman TY, Mahanta LB, Das AK, Sarma JD. Automated Oral Squamous Cell Carcinoma Identification Using Shape, Texture and Color Features of Whole Image Strips. *Tissue Cell* (2020) 63:101322. doi: 10.1016/j.tice.2019.101322
16. Das N, Hussain E, Mahanta LB. Automated Classification of Cells Into Multiple Classes in Epithelial Tissue of Oral Squamous Cell Carcinoma Using Transfer Learning and Convolutional Neural Network. *Neural Networks* (2020) 128:47–60. doi: 10.1016/j.neunet.2020.05.003
17. Khosravi P, Kazemi E, Imielinski M, Elemento O, Hajirasouliha I. Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images. *EBioMedicine* (2018) 27:317–28. doi: 10.1016/j.ebiom.2017.12.026
18. Wang S, Yang DM, Rong R, Zhan X, Xiao G. Pathology Image Analysis Using Segmentation Deep Learning Algorithms. *Am J Pathol* (2019) 189(9):1686–98.
19. Japkowicz N. *Learning From Imbalanced Data Sets: A Comparison of Various Strategies*. Menlo Park CA: AAAI Press (2000) p. 10–5.
20. Tan M, Le Q. Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning; 2019 Jun 9-15*;California: PMLR (2019). p. 6105–14.
21. Zhang X, Zhou XY, Lin MX, Sun R. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. *Proc Cvpr Ieee* (2018) 2018:6848–56. doi: 10.1109/Cvpr.2018.00716
22. He KM, Zhang XY, Ren SQ, Sun J. Deep Residual Learning for Image Recognition, in: *2016 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr)* Las Vegas: IEEE (2016). pp. 770–8.
23. Network CGA. Comprehensive Genomic Characterization of Head and Neck Squamous Cell Carcinomas. *Nature* (2015) 517(7536):576. doi: 10.1038/nature14129
24. Deo RC. Machine Learning in Medicine. *Circulation* (2015) 132(20):1920–30. doi: 10.1161/Circulationaha.115.001593
25. Lin M, Chen H, Sun X, Qian Q, Li H, Jin R. Neural Architecture Design for Gpu-Efficient Networks. *arXiv [Preprint]* (2020). Available at: https://arxiv.org/abs/2006.14090 (Accessed Aug 11, 2020).
26. Radosavovic I, Kosaraju RP, Girshick R, He K, Dollár P. Designing Network Design Spaces. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2016 Jun 13-19; Seattle: IEEE (2020). p. 10428–36.
27. Poggio T, Kawaguchi K, Liao Q, Miranda B, Rosasco L, Boix X, et al. Theory of Deep Learning Iii: The non-Overfitting Puzzle. *arXiv [Preprint]* (2017). Available at: https://arxiv.org/abs/1801.00173 (Accessed Dec 30, 2017).
28. Ling CX, Huang J, Zhang H. AUC: A Better Measure Than Accuracy in Comparing Learning Algorithms. *Conference of the Canadian society for computational studies of intelligence*; 2003 Jun 11-13; Berlin, Heidelberg: Springer (2003) p. 329–41.
29. Ball CS. The early history of the compound microscope. *Bios* (1966) 2:51–60.
30. Liu Y, Gadepalli K, Norouzi M, Dahl GE, Kohlberger T, Boyko A, et al. Detecting Cancer Metastases on Gigapixel Pathology Images. *arXiv* [Preprint] (2017). Available at: https://arxiv.org/abs/1703.02442 (Accessed Mar 3, 2017).
31. Sabottke CF, Spieler BM. The Effect of Image Resolution on Deep Learning in Radiography. *Radiol: Artif Intelligence* (2020) 2(1):e190015. doi: 10.1148/ryai.2019190015
32. Nalepa J, Kawulok M. Selecting Training Sets for Support Vector Machines: A Review. *Artif Intell Rev* (2019) 52(2):857–900. doi: 10.1007/s10462-017-9611-1
33. Krawczyk B. Learning From Imbalanced Data: Open Challenges and Future Directions. *Prog Artif Intelligence* (2016) 5(4):221–32. doi: 10.1007/s13748-016-0094-0
34. Mermod M, Jourdan EF, Gupta R, Bongiovanni M, Tolstonog G, Simon C, et al. Development and Validation of a Multivariable Prediction Model for the Identification of Occult Lymph Node Metastasis in Oral Squamous Cell Carcinoma. *Head Neck* (2020) 42(8):1811–20. doi: 10.1002/hed.26105
35. Romeo V, Cuocolo R, Ricciardi C, Ugga L, Cocozza S, Verde F, et al. Prediction of Tumor Grade and Nodal Status in Oropharyngeal and Oral Cavity Squamous-Cell Carcinoma Using a Radiomic Approach. *Anticancer Res* (2020) 40(1):271–80. doi: 10.21873/anticanres.13949
36. Ariji Y, Fukuda M, Kise Y, Nozawa M, Yanashita Y, Fujita H, et al. Contrast-Enhanced Computed Tomography Image Assessment of Cervical Lymph Node Metastasis in Patients With Oral Cancer by Using a Deep Learning System of Artificial Intelligence. *Oral Surgery Oral Med Oral Pathol Oral Radiol* (2019) 127(5):458–63. doi: 10.1016/j.oooo.2018.10.002
37. Jeyaraj PR, Samuel Nadar ER. Computer-Assisted Medical Image Classification for Early Diagnosis of Oral Cancer Employing Deep Learning Algorithm. *J Cancer Res Clin Oncol* (2019) 145(4):829–37. doi: 10.1007/s00432-018-02834-7
38. Pham HHN, Futakuchi M, Bychkov A, Furukawa T, Kuroda K, Fukuoka J. Detection of Lung Cancer Lymph Node Metastases From Whole-Slide Histopathologic Images Using a Two-Step Deep Learning Approach. *Am J Pathol* (2019) 189(12):2428–39. doi: 10.1016/j.ajpath.2019.08.014

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.

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 07 July 2022
doi: 10.3389/fonc.2022.915871

Deep Learning Analysis Using ^{18}F-FDG PET/CT to Predict Occult Lymph Node Metastasis in Patients With Clinical N0 Lung Adenocarcinoma

Ming-li Ouyang[1†], Rui-xuan Zheng[1†], Yi-ran Wang[2], Zi-yi Zuo[1], Liu-dan Gu[1], Yu-qian Tian[1], Yu-guo Wei[3], Xiao-ying Huang[1], Kun Tang[4*] and Liang-xing Wang[1*]*

[1] Key Laboratory of Heart and Lung, Division of Pulmonary Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, [2] Department of Medical Engineering, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, [3] Precision Health Institution, General Electric (GE) Healthcare, Hangzhou, China, [4] Department of Nuclear Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

Introduction: The aim of this work was to determine the feasibility of using a deep learning approach to predict occult lymph node metastasis (OLM) based on preoperative FDG-PET/CT images in patients with clinical node-negative (cN0) lung adenocarcinoma.

Materials and Methods: Dataset 1 (for training and internal validation) included 376 consecutive patients with cN0 lung adenocarcinoma from our hospital between May 2012 and May 2021. Dataset 2 (for prospective test) used 58 consecutive patients with cN0 lung adenocarcinoma from June 2021 to February 2022 at the same center. Three deep learning models: PET alone, CT alone, and combined model, were developed for the prediction of OLM. The performance of the models was evaluated on internal validation and prospective test in terms of accuracy, sensitivity, specificity, and areas under the receiver operating characteristic curve (AUCs).

Results: The combined model incorporating PET and CT showed the best performance, achieved an AUC of 0.81 [95% confidence interval (CI): 0.61, 1.00] in the prediction of OLM in internal validation set (n = 60) and an AUC of 0.87 (95% CI: 0.75, 0.99) in the prospective test set (n = 58). The model achieved 87.50% sensitivity, 80.00% specificity, and 81.00% accuracy in the internal validation set and achieved 75.00% sensitivity, 88.46% specificity, and 86.60% accuracy in the prospective test set.

Conclusion: This study presented a deep learning approach to enable the prediction of occult nodal involvement based on the PET/CT images before surgery in cN0 lung adenocarcinoma, which would help clinicians select patients who would be suitable for sublobar resection.

Keywords: positron emission tomography/computed tomography (PET/CT), convolutional neural network, lung adenocarcinoma, sublobar resection, lymph node status

OPEN ACCESS

Edited by:
Wei Wei,
Xi'an University of Technology, China

Reviewed by:
Tsung-Ying Ho,
Chang Gung Memorial Hospital, Taiwan
Vetri Sudar Jayaprakasam,
Memorial Sloan Kettering Cancer Center, United States

****Correspondence:***
Liang-xing Wang
wangliangxing@wzhospital.cn
Kun Tang
kuntang007@163.com
Xiao-ying Huang
huangxiaoying@wzhospital.cn

[†]These authors have contributed equally to this work

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

Received: *08 April 2022*
Accepted: *07 June 2022*
Published: *07 July 2022*

Citation:
Ouyang M-l, Zheng R-x, Wang Y-r, Zuo Z-y, Gu L-d, Tian Y-q, Wei Y-g, Huang X-y, Tang K and Wang L-x (2022) Deep Learning Analysis Using ^{18}F-FDG PET/CT to Predict Occult Lymph Node Metastasis in Patients With Clinical N0 Lung Adenocarcinoma. Front. Oncol. 12:915871. doi: 10.3389/fonc.2022.915871

INTRODUCTION

Lung cancer is one of the most common malignancies and the leading cause of death from cancer worldwide (1). Lung adenocarcinoma (LUAD) is the most common histologic subtype of lung cancer (2). Currently, lobectomy with systemic nodal dissection is the standard treatment for patients with early-stage non–small cell lung cancer (NSCLC) (3), and recently, limited surgery (wedge resection or segmentectomy) has also been performed to preserve healthy lung tissue (4–6). Accurate staging to confirm node-negative (N0) status is required for limited surgery. If N0 status is unreliable, then lobectomy with systemic nodal dissection rather than limited surgery is mandatory. ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) is a valuable imaging modality for evaluation of lymph node (LN) or distant metastasis of lung cancers. Although PET/CT is more sensitive to assess LN status than traditional examinations, occult LN metastasis (OLM) still occurs at a high rate (14%–21%) (7–9). The definition for OLM was that clinical N0 (cN0) staged by PET/CT was pathologically confirmed LN metastasis (LNM) after surgery. Thus, there is a strong need to develop reliable non-invasive methods to identify patients with OLM from cN0 patients staged by PET/CT.

In recent years, radiomics has received increasing attention, and it is a technique for high-throughput extraction of quantitative features from medical images (10, 11). Indeed, many studies have exhibited that quantitative radiomic image features of the primary tumor could be used as non-invasive biomarkers to predict LNM and were good predictive performance (12–14). For OLM of LUAD, Zhong et al. (15) reported that the radiomics signature of the primary tumor based on CT scans had a significant predictive value. Our previous research (16) found that a PET-based radiomics model had achieved success in the prediction of OLM in patients with LUAD. However, traditional radiomic methods are based on four time-consuming and complex steps (tumor segmentation, feature extraction, feature selection, and modeling). Moreover, observer-dependent differences may cause poor repeatability in case of manual segmentation.

Deep learning is a new and especially promising approach that automatically learns powerful feature representations from images, text, or sound and has been shown to sometimes surpass human-level performance in task-specific applications (17–19). Compared with the conventional radiomic methods, the deep learning method simplifies the analysis process and avoids subjective bias because it does not require VOI definition or segmentation. More recently, the deep learning method using convolutional neural network (CNN) has been widely applied to analyze medical images and has been effective in diseases detection and classification (20–22). For classifying LN, Zhao et al. (23) developed a cross-modal deep learning system based on CT images to accurately predict LN metastasis in stage T1 LUADs. Tau et al. (24) reported that using a CNN to analyze PET images can yield a reasonably good prediction of nodal metastasis in patients with NSCLC. In view of the fact that previous studies predicted LN metastasis using deep learning, we hypothesized that deep learning based on PET/CT images might play an important role in predicting OLM.

Hence, the purpose of this study was to evaluate the capability of deep learning analysis based on a two-dimensional (2D) CNN architecture for the prediction of OLM through the use of preoperative FDG-PET/CT images of cN0 LUAD.

MATERIALS AND METHODS

Patients

A total of 434 patients (193 men and 241 women) with cN0 LUAD who had pretreatment FDG PET/CT and underwent surgical resection with the systematic LN dissection from May 2012 to February 2022 were enrolled in this study at The First Affiliated Hospital of Wenzhou Medical University. Among these patients, 343 (79.0%) were pN0 after surgery and pathological examination. In other words, the prevalence of OLM with PET/CT was 21.0% in LUAD, which is basically consistent with previous studies (7–9). The criteria for cN0 on PET/CT was all LNs' short-axis diameter of less than 10 mm without FDG uptake higher than the surrounding background (25). The interval between PET/CT scan and surgery was shorter than 3 weeks in all patients. The exclusion criteria for patients were as follows: (I) history of other malignancy; (II) distant metastasis; (III) multiple lesions; (IV) neoadjuvant chemotherapy/radiotherapy; (V) images with poor quality due to the leakage of ^{18}F-FDG at the injection site, low signal-to-noise ratio, respiratory artifacts, and other movement artifacts. Staging was performed according to the eighth edition of the Union for International Cancer Control TNM classification.

Dataset 1 included 376 consecutive patients from our Hospital between May 2012 and May 2021. Dataset 2 used 58 consecutive patients from June 2021 to February 2022 at the same center. Sixty patients from dataset 1 were randomly allocated to the internal validation dataset, and the remaining 316 patients were assigned to the training set. Dataset 2 was taken as an independent set for the prospective test. The prospective test is a more powerful method for evaluating the model performance than random splitting of a single set or cross-validation because it allows for non-random variation between sets (26). A flowchart of patient selection is shown in **Figure 1**.

This study was approved by the Institutional Review Board of our hospital. Informed consent from the retrospective patients was waived, and written informed consent was provided for patients in prospective test set.

PET/CT Acquisition

An integrated PET/CT scanner (GEMINI TF 64; Philips, The Netherlands) was used for all patients. At least 6-h fasting and serum glucose levels below 110 ml/dl were required before being injected with ^{18}F-FDG (3.7 MBq/kg). Sixty minutes after intravenous injection, the body was scanned in the supine position. A low-dose unenhanced CT scan from skull base to the middle thighs was obtained with the following parameters:

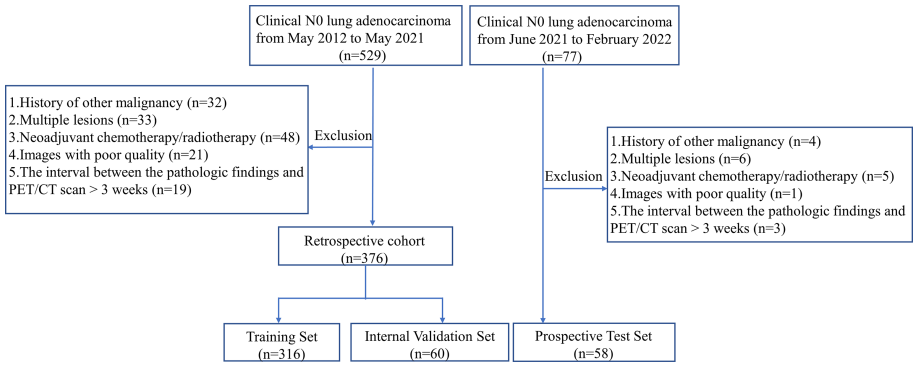


FIGURE 1 | The flowchart of the patient selection.

120 kV, 80 mA, pitch of 0.829, and reconstruction thickness and interval of 5.0 mm. After CT completion, PET images were acquired by using the 3D model with the following parameters: field of view of 576 mm, a matrix of 144 × 144, slice thickness and interval of 5.0 mm, and an emission scan time of each bed position of 1.5 min. PET images were iteratively reconstructed by the ordered subset expectation maximization algorithm, using CT image for attenuation correction.

Image Selection and Processing

FDG uptake at the primary tumor site was identified on PET images with reference to the CT part of PET-CT. Reconstruction in the sagittal and coronal planes was done from the axial images. Slices with the largest tumor area were selected in axial, coronal, and sagittal planes of PET and CT images. To reduce the computational expense and improve model's accuracy, all selected images were cropped to contain only the entire chest as much as possible. Then, the images were converted from the Digital Imaging and Communications in Medicine to Joint Photographic Experts Group format pictures. Subsequently, we resized the images to 299 × 299 pixels and normalized the pixel values to a range of 0 to 1.

There was a higher frequency of OLM negative (OLMN). To overcome the imbalance problem between the two groups (positive or negative), we applied three times oversampling for positive samples and two times oversampling for negative samples to ensure the ratio of the two groups near 1:1. Furthermore, image augmentation, including image rotation and flipping for total of four times, was performed on the training dataset.

CNN Model Architecture

Respective model (PET or CT): The deep CNN model used was the Inception V3 architecture in this study (27). Transfer learning was applied using weights pretrained on the ImageNet dataset. We arranged three channels (299 × 299 × 3 pixels) in the input layer. Three 2D slices (axial, coronal, and sagittal) were used as input to the CNN network rather than 3D volume data because 2D-based analysis enabled us to reduce GPU memory usage and limit the overfitting. The generated features from the Inception V3 were flattened into a 1D feature vector after the average pooling layer. In the end, six fully connected layers and a sigmoid layer were connected to enable the classification of OLMN and OLM positive (OLMP). To avoid overfitting, dropouts were used. The architecture of the CNN is shown in **Figure 2A**.

Combined model (PET + CT): For the construction of complex model, PET and CT were first respectively run to the last full connect layer, then combined them together, and finally connected a sigmoid layer (1 nodes) for classification. Schematic overview of the combined model is shown in **Figure 2B**.

All the above analyses were implemented in the Keras library in Python, using TensorFlow as backend (Python 2.7, Keras 2.6.0, TensorFlow 2.6.0). Adam with a learning rate of 0.000012 and a batch size of 32 was used for parameters optimization. The number of epochs of training was set to 100.

Model Performance

For assessing the performance of prediction models, the receiver operating characteristic (ROC) curves were displayed in the training, internal validation, and prospective test sets, respectively. The performance metrics such as accuracy, sensitivity, specificity, and the area under the curve (AUC) were calculated. Fivefold cross-validation was used to verify the generalization ability.

Statistical Analysis

The statistical analyses were implemented by using IBM SPSS (version 25.0) and Python (version 2.7). Categorical data were analyzed with the chi-square test and the Fisher's exact test. Numerical data were analyzed with the unpaired t-test, Mann–Whitney U-test, ANOVA, and Kruskal–Wallis test. For missing data, mode imputation was used for categorical variables, and mean imputation was used for continuous variables. P-values less than 0.05 indicated a statistically significant difference.

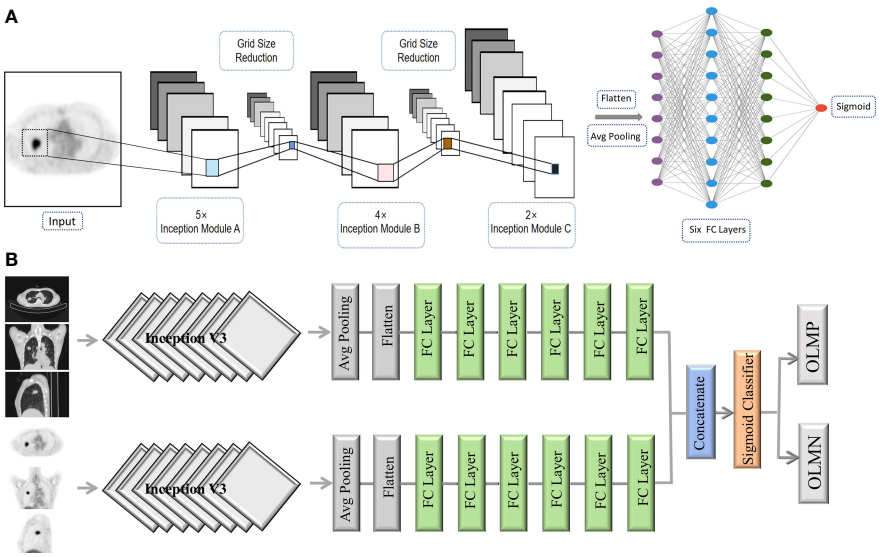


FIGURE 2 | The architecture of the CNN **(A)**. Schematic overview of the combined model (PET + CT) **(B)**. Avg pooling, average pooling; FC layer, fully connected layer; OLMN, occult lymph node metastasis negative; OLMP, occult lymph node metastasis positive.

RESULTS

Baseline Information

The baseline patient characteristics are shown in **Table 1**. The sample sizes of the training, internal validation, and prospective test sets were 316, 60, and 58, respectively. No statistical differences, including age (p = 0.663), gender (p =0.820), smoking history (p = 0.418), tumor location (p = 0.522), radiologic lesion type (p= 0.244), tumor SUVmax (p = 0.261), carcinoembryonic antigen (CEA) (p = 0.250), and predominant subtype (p = 0.088), among the three sets were observed except for pathologic tumor size (p = 0.011) in **Table 1**.

Comparison of Clinicopathologic Data Between OLMN and OLMP Groups

A comparison of clinicopathologic data between OLMN and OLMP groups in the three sets is presented in **Table 2**. OLMP was identified in 91 of all 434 patients (20.9%). The training set of 316 patients included 75 OLMP (23.7%) and 241 OLMN (76.3%). The internal validation set of 60 patients included 8 OLMP (13.3%) and 52 OLMN (86.7%). The prospective test set of 58 patients included 8 OLMP (13.8%) and 50 OLMN (86.2%). Detailed information about the distribution of N stages for OLMP cases of three datasets is shown in **Table 2**. In addition, similar tendencies were observed for pathologic tumor size, CEA, and tumor SUVmax, respectively, in the three sets, although not always statistically significant.

Performance of Deep Learning Models

The deep learning models demonstrated good predictive performance for OLM with the use of the primary lung cancer images of internal validation set, with AUCs of 0.74 [95% confidence interval (CI): 0.58, 0.90) for the PET model, 0.79 (95% CI: 0.58, 1.00) for the CT model, and 0.81 (95% CI: 0.61, 1.00) for the complex model. For prospective test set, the AUCs were 0.73 (95% CI: 0.51, 0.95) for the PET model, 0.79 (95% CI: 0.59, 0.98) for the CT model, and 0.87 (95% CI: 0.75, 0.99) for the complex model (**Figure 3**). The discriminatory ability of the complex model displayed the highest in the validation and test sets.

For internal validation set, the sensitivities of PET, CT, and combined models were 75.00%, 75.00%, and 87.50%, respectively; the specificities of PET, CT, and combined models were 63.46%, 88.46%, and 80.00%, respectively; and the accuracies of PET, CT, and combined models were 65.00%, 86.67%, and 81.00%, respectively (**Table 3**).

For prospective test set, the sensitivities of PET, CT, and combined models were 87.50%, 75.00%, and 75.00%, respectively; the specificities of PET, CT, and combined models were 62.00%, 80.00% and 88.46%, respectively; and the accuracies of PET, CT, and combined models were 65.52%,79.31% and 86.60%, respectively (**Table 3**).

The training curves of PET and CT are provided in **Figure 4**. The validation losses of PET and CT basically reached the minimum at 40~45 epochs, and then losses of training set and validation set estranged after 40 epochs. Therefore, we stopped training at the 40th epoch because no further improvement can be gained in the validation loss. The slowly decrease of validation losses suggests that the models have no overfitting before 45 epochs.

DISCUSSION

Recently, the therapeutic effect of limited surgery in patients with early-stage NSCLC without LNM has been proved to be

TABLE 1 | Baseline characteristics of datasets.

Characteristics	Training Set (n = 316)	Internal Validation Set (n = 60)	Prospective Test Set (n = 58)	P-Value
Age (years) *	62.29 ± 9.73	63.17 ± 9.44	63.36 ± 11.83	0.663
Sex				0.820
Female	178 (56.3)	33 (55.0)	30 (51.7)	
Male	138 (43.7)	27 (45.0)	28 (48.3)	
Smoking history				0.418
Ever smoker	78 (24.7)	16 (26.7)	10 (17.2)	
Never smoker	238 (75.3)	44 (73.3)	48 (82.8)	
Tumor location				0.522
RUL	97 (30.7)	14 (23.4)	19 (32.8)	
RML	20 (6.3)	6 (10.0)	8 (13.8)	
RLL	70 (22.2)	11 (18.3)	10 (17.2)	
LUL	81 (25.6)	18 (30.0)	14 (24.1)	
LLL	48 (15.2)	11 (18.3)	7 (12.1)	
Radiologic lesion type				0.244
Pure solid	288 (91.1)	53 (88.3)	49 (84.5)	
Subsolid	28 (8.9)	7 (11.7)	9 (15.5)	
Tumor SUVmax*	5.62 ± 3.59	4.63 ± 2.43	5.70 ± 4.34	0.261
CEA, ng/ml*	7.76 ± 33.89	4.06 ± 2.73	6.75 ± 11.82	0.25
Pathologic tumor size*	23.31 ± 10.36	19.87 ± 8.83	23.64 ± 9.93	0.011
Predominant subtype				0.088
Acinar	232 (73.4)	41 (68.3)	42 (72.4)	
Papillary	34 (10.8)	6 (10.0)	9 (15.6)	
Lepidic	25 (7.9)	4 (6.7)	0 (0)	
Solid	13 (4.1)	4 (6.7)	6 (10.3)	
Micropapillary	1 (0.3)	1 (1.6)	0 (0)	
Colloid	11 (3.5)	4 (6.7)	1 (1.7)	

RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; CEA, carcinoembryonic antigen.
**Data are means ± standard deviations.*

significant, and limited surgery has more available lung tissue and lower perioperative mortality than standard treatment (28–30). Hence, there is an increasing need for accurately predicting OLM of cN0 LUAD before surgery in a non-invasive way.

Deep learning, which takes raw image pixels and corresponding class labels from image data as inputs and automatically learns representative information, has recently attracted much attention due to its excellent performance in image recognition tasks (17). In this study, we developed three deep learning models based on FDG PET/CT images for preoperative prediction of OLM in patients with cN0 LUAD. Our results presented that the complex model combining ^{18}F-FDG PET and low-dose CT showed better diagnostic performances in distinguishing patients with OLMN and OLMP than either PET or CT alone.

Some studies demonstrated that CNN-based image analysis has been effectively applied in predicting LN status of lung cancer. For example, Zhong et al. (31) showed that a deep learning signature based on CT images could accurately predict occult N2 disease in patients with clinical stage I NSCLC. However, it is already known that PET/CT is more accurate than CT for direct assessment of LN status. Thus, confirming N0 status by CT is not enough. Tau et al. (32) demonstrated that using a CNN to analyze segmented primary tumors with PET in patients with pretreatment NSCLC can yield moderately high accuracy for designation of N category, but the use of segmented tumors as input data for the CNN was time-consuming and might affect the results. Moreover, most recent studies using deep learning (including the two studies discussed above) only performed single-modality analyses because integrating multimodal data is vulnerable to overfitting and poor generalization (22, 33). Wang et al. (34) mixed image patches of both modalities (PET and CT) into the same network, and the result showed that the performance of CNN was not significantly different from the best classical methods and human doctors for the classification of mediastinal LNM in patients with NSCLC. Such mixed setting may affect the final result because two different patches contained different types of diagnostic information. In this study, we processed the PET and CT patches with respective subnetworks and combined the results of the two different subnetworks at the output layers. For the internal validation set, the AUCs of the CNN in predicting nodal metastasis were as follows: ^{18}F-FDG PET alone, 0.74; CT alone, 0.79; and 18F-FDG PET/CT, 0.81. For the prospective test set, the AUCs were as follows: 18F-FDG PET alone, 0.73; CT alone, 0.79; and ^{18}F-FDG PET/CT, 0.87. Our results showed that the combined method, which makes full use of PET functional information and CT anatomic information, showed significantly great diagnostic performances in predicting OLM of LUAD.

A 2D CNN to discriminate between OLMN and OLMP in cN0 LUAD was successfully trained, validated, and tested in this study. Previous studies proposed that 3D CNN–based CT image analysis was used for classification in patients with lung cancer (23, 35). However, the increased complexity comes at a high computational cost. Another factor to consider is whether adding

TABLE 2 | Comparison of clinical features between OLMN and OLMP groups in the three sets.

Characteristics	Training Set			Internal Validation Set			Prospective Test Set		
	(OLMN = 241; OLMP = 75)			(OLMN = 52; OLMP = 8)			(OLMN = 50; OLMP = 8)		
	OLMN	OLMP	P	OLMN	OLMP	P	OLMN	OLMP	P
Age (years) *	63.03 ± 9.46	59.91 ± 10.24	0.015	63.29 ± 9.44	62.38 ± 10.01	0.801	63.22 ± 11.70	64.25 ± 13.48	0.822
Sex			0.125			0.939			0.299
Female	130 (53.9)	48 (64.0)		28 (53.8)	5 (62.5)		24 (48.0)	6 (75.0)	
Male	111 (46.1)	27 (36.0)		24 (46.2)	3 (37.5)		26 (52.0)	2 (25.0)	
Smoking history			0.004			1			0.375
Ever smoker	69 (28.6)	9 (12.0)		14 (26.9)	2 (25.0)		10 (20.0)	0 (0)	
Never smoker	172 (71.4)	66 (88.0)		38 (73.1)	6 (75.0)		40 (80.0)	8 (100)	
Tumor location			0.650			0.736			0.597
RUL	76 (31.5)	21 (28.0)		13 (25.0)	1 (12.5)		16 (32.0)	3 (37.5)	
RML	14 (5.8)	6 (8.0)		6 (11.5)	0 (0)		8 (16.0)	0 (0)	
RLL	52 (21.6)	18 (24.0)		10 (19.2)	1 (12.5)		8 (16.0)	2 (25.0)	
LUL	65 (27.0)	16 (21.3)		14 (26.9)	4 (50.0)		11 (22.0)	3 (37.5)	
LLL	34 (14.1)	14 (18.7)		9 (17.4)	2 (25.0)		7 (14.0)	0 (0)	
Radiologic lesion type			0.031			0.608			0.436
Pure solid	215 (89.2)	73 (97.3)		45 (86.5)	8 (100)		41 (82.0)	8 (100)	
Subsolid	26 (10.8)	2 (2.7)		7 (13.5)	0 (0)		9 (18.0)	0 (0)	
Tumor SUVmax*	4.96 ± 3.24	7.74 ± 3.85	< 0.001	4.45 ± 2.40	5.82 ± 2.42	0.064	5.22 ± 4.36	8.64 ± 2.95	0.002
CEA, ng/mL*	5.62 ± 9.15	15.24 ± 67.42	0.029	3.11 ± 2.19	4.5 ± 2.46	0.046	4.64 ± 2.88	19.18 ± 29.52	0.311
Pathologic tumor size*	22.18 ± 9.62	26.93 ± 11.78	< 0.001	19.23 ± 8.82	24.00 ± 8.25	0.056	21.24 ± 7.19	38.63 ± 11.94	< 0.001
Predominant subtype			0.318			0.399			0.629
Acinar	171 (71.0)	61 (81.3)		36 (69.2)	5 (62.5)		37 (74.0)	5 (62.5)	
Papillary	26 (10.8)	8 (10.7)		5 (9.6)	1 (12.5)		7 (14.0)	2 (25.0)	
Lepidic	23 (9.5)	2 (2.6)		4 (7.7)	0 (0)		0 (0)	0 (0)	
Solid	10 (4.1)	3 (4.0)		2 (3.9)	2 (25.0)		5 (10.0)	1 (12.5)	
Micropapillary	1 (0.4)	0 (0)		1 (1.9)	0 (0)		0 (0)	0 (0)	
Colloid	10 (4.2)	1 (1.4)		4 (7.7)	0 (0)		1 (2.0)	0 (0)	
pN (8th ed.)									
N1a (single N1)		31 (41.3)			3 (37.5)			3 (37.5)	
N1b (multiple N1)		6 (8.0)			0 (0)			0 (0)	
N2a (single N2)		17 (22.7)			2 (25.0)			0 (0)	
N2b (multiple N2)		21 (28.0)			3 (37.5)			5 (62.5)	

OLMN, occult lymph node metastasis negative; OLMP, occult lymph node metastasis positive; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; CEA, carcinoembryonic antigen.
**Data are means ± standard deviations.*

these interslice features would improve classification performance. Lee et al. (36) reported that a 2D CNN slice-based approach had better performance than 3D-CNN case-based approach for detecting intrapelvic tumor recurrence and metastases. The study of Vries et al. (21) also showed that the sagittal 2D CNN already performed with very high accuracy for discriminating between Aß-negative and -positive PET scans in patients with subjective cognitive decline. Therefore, we

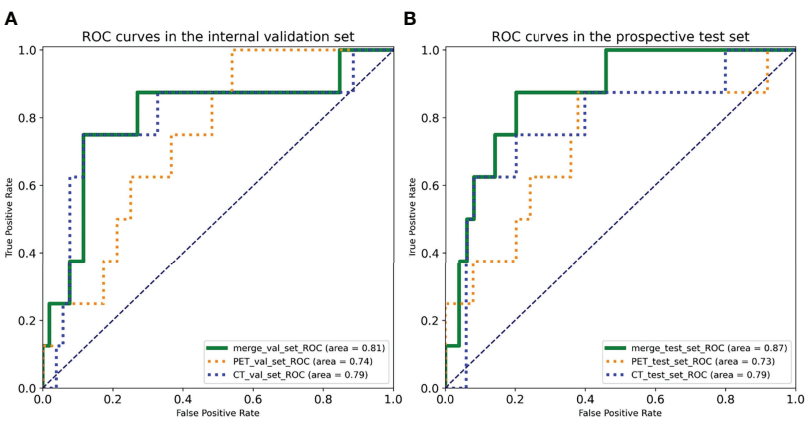


FIGURE 3 | Receiver operating characteristic (ROC) curves of three deep learning models in the **(A)** internal validation set and the **(B)** prospective test set.

TABLE 3 | Performance of the three deep learning models.

	PET			CT			Combined		
	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
Internal Validation Set	75.00%	63.46%	65.00%	75.00%	88.46%	86.67%	87.50%	80.00%	81.00%
Prospective Test Set	87.50%	62.00%	65.52%	75.00%	80.00%	79.31%	75.00%	88.46%	86.60%

hypothesized that patients without a very large number of cases may be more applicable to 2D CNN architectures.

For clinical features, statistical analysis showed a significant difference in pathologic tumor size, CEA, and tumor SUVmax in the training set, which is consistent with our previous findings (16, 37). However, these clinical features were not all statistically significant in our validation and test sets, which may imply that the clinical utility of these features is limited.

There are several limitations to our current study. First, this was a single-center study with a relatively small sample size. Further improvement with multicenter and large-sample studies must be achieved before clinical use. Second, patients with multiple lesions were excluded because it is difficult to determine which lesion would cause OLM and should be input in the model. Therefore, predicting OLM of multifocal lung cancer needs to be further verified. Third, although statistical analysis of clinical data was performed, we did not integrate these clinical features into the deep learning model. Therefore, clinical parameters as another modality combined DL model should be studied in the future. Fourth, we did not use PET/CT fusion images because PET scan is difficult to rigidly match with CT scan in spatial location due to cardiac and respiratory motion artifacts. Last, limitation also obviously includes the opaque black box nature of the deep learning technology.

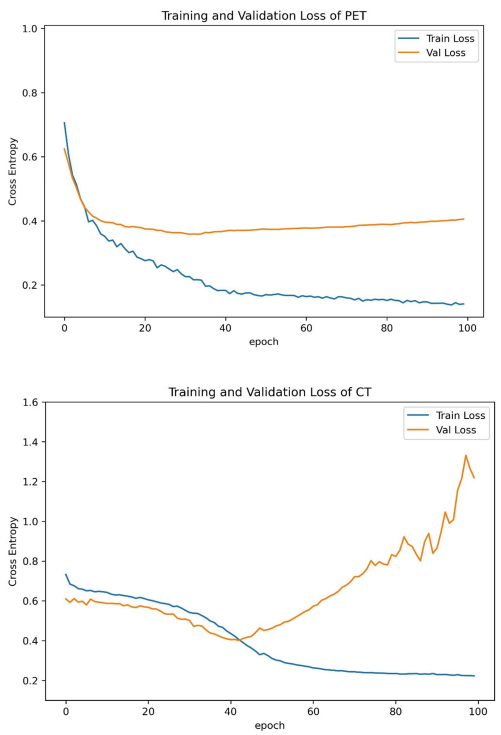


FIGURE 4 | Training curves of PET and CT models. We stopped training at the 40th epoch because no further improvement can be gained in the validation loss.

CONCLUSIONS

We constructed a deep learning model that can successfully incorporate PET and CT images into a 2D CNN architecture to accurately predict OLM in patients with cN0 LUAD. Moreover, the deep learning model demonstrated a good predictive performance. This model may help to determine the patients who are eligible for limited resection.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The medical images are not publicly available due to the ethical considerations. The analysis code used in this study can be obtained by the corresponding author upon reasonable request. Requests to access these datasets should be directed to M-LO, 1427738937@qq.com.

ETHICS STATEMENT

Informed consent from the retrospective patients was waived, and written informed consent was provided for patients in prospective test set.

AUTHOR CONTRIBUTIONS

R-XZ, Y-RW, L-DG, Y-QT, and M-LO collected the clinical information and the imaging data. R-XZ, Z-YZ, Y-GW, and M-LO were responsible for writing code and data analysis.

FUNDING

This study was supported by Zhejiang Public Welfare Technology Application Research Project, China (LGF21H010009) and Wenzhou Science and Technology Program (no. Y20210222).

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68 (6):394–424. doi: 10.3322/caac.21492
2. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-Small-Cell Lung Cancers: A Heterogeneous Set of Diseases. *Nat Rev Cancer* (2014) 14 (8):535–46. doi: 10.1038/nrc3775
3. Ettinger DS, Wood DE, Aggarwal C, Aisner DL, Akerley W, Bauman JR, et al. NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 1.2020. *J Natl Compr Canc Netw* (2019) 17(12):1464–72. doi: 10.6004/jnccn.2019.0059
4. Yerokun BA, Yang CJ, Gulack BC, Li X, Mulvihill MS, Gu L, et al. A National Analysis of Wedge Resection Versus Stereotactic Body Radiation Therapy for Stage IA non-Small Cell Lung Cancer. *J Thorac Cardiovasc Surg* (2017) 154 (2):675–86.e4. doi: 10.1016/j.jtcvs.2017.02.065
5. Zhang Z, Feng H, Zhao H, Hu J, Liu L, Liu Y, et al. Sublobar Resection is Associated With Better Perioperative Outcomes in Elderly Patients With Clinical Stage I non-Small Cell Lung Cancer: A Multicenter Retrospective Cohort Study. *J Thorac Dis* (2019) 11(5):1838–48. doi: 10.21037/jtd.2019.05.20
6. Grills IS, Mangona VS, Welsh R, Chmielewski G, McInerney E, Martin S, et al. Outcomes After Stereotactic Lung Radiotherapy or Wedge Resection for Stage I non-Small-Cell Lung Cancer. *J Clin Oncol* (2010) 28(6):928–35. doi: 10.1200/JCO.2009.25.0928
7. Gomez-Caro A, Garcia S, Reguart N, Arguis P, Sanchez M, Gimferrer JM, et al. Incidence of Occult Mediastinal Node Involvement in Cn0 non-Small-Cell Lung Cancer Patients After Negative Uptake of Positron Emission Tomography/Computer Tomography Scan. *Eur J Cardiothorac Surg* (2010) 37(5):1168–74. doi: 10.1016/j.ejcts.2009.12.013
8. Kirmani BH, Rintoul RC, Win T, Magee C, Magee L, Choong C, et al. Stage Migration: Results of Lymph Node Dissection in the Era of Modern Imaging and Invasive Staging for Lung Cancer. *Eur J Cardiothorac Surg* (2013) 43 (1):104–9. doi: 10.1093/ejcts/ezs184
9. Veeramachaneni NK, Battafarano RJ, Meyers BF, Zoole JB, Patterson GA. Risk Factors for Occult Nodal Metastasis in Clinical T1N0 Lung Cancer: A Negative Impact on Survival. *Eur J Cardiothorac Surg* (2008) 33(3):466–9. doi: 10.1016/j.ejcts.2007.12.015
10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
11. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006
12. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol* (2016) 34(18):2157–64. doi: 10.1200/JCO.2015.65.9128
13. Wu S, Zheng J, Li Y, Yu H, Shi S, Xie W, et al. A Radiomics Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer. *Clin Cancer Res* (2017) 23(22):6904–11. doi: 10.1158/1078-0432.CCR-17-1510
14. Li Q, He XQ, Fan X, Zhu CN, Lv JW, Luo TY. Development and Validation of a Combined Model for Preoperative Prediction of Lymph Node Metastasis in Peripheral Lung Adenocarcinoma. *Front Oncol* (2021) 11:675877. doi: 10.3389/fonc.2021.675877
15. Zhong Y, Yuan M, Zhang T, Zhang YD, Li H, Yu TF. Radiomics Approach to Prediction of Occult Mediastinal Lymph Node Metastasis of Lung Adenocarcinoma. *AJR Am J Roentgenol* (2018) 211(1):109–13. doi: 10.2214/AJR.17.19074
16. Wang L, Li T, Hong J, Zhang M, Ouyang M, Zheng X, et al. (18)F-FDG PET-Based Radiomics Model for Predicting Occult Lymph Node Metastasis in Clinical N0 Solid Lung Adenocarcinoma. *Quant Imaging Med Surg* (2021) 11 (1):215–25. doi: 10.21037/qims-20-337
17. LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature* (2015) 521(7553):436–44. doi: 10.1038/nature14539
18. Fayek HM, Lech M, Cavedon L. Evaluating Deep Learning Architectures for Speech Emotion Recognition. *Neural Netw* (2017) 92:60–8. doi: 10.1016/j.neunet.2017.02.013
19. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial Intelligence in Radiology. *Nat Rev Cancer* (2018) 18(8):500–10. doi: 10.1038/s41568-018-0016-5
20. Sibille L, Seifert R, Avramovic N, Vehren T, Spottiswoode B, Zuehlsdorff S, et al. (18)F-FDG PET/CT Uptake Classification in Lymphoma and Lung Cancer by Using Deep Convolutional Neural Networks. *Radiology* (2020) 294 (2):445–52. doi: 10.1148/radiol.2019191114
21. de Vries BM, Golla SSV, Ebenau J, Verfaillie SCJ, Timmers T, Heeman F, et al. Classification of Negative and Positive (18)F-Florbetapir Brain PET Studies in Subjective Cognitive Decline Patients Using a Convolutional Neural Network. *Eur J Nucl Med Mol Imaging* (2021) 48(3):721–8. doi: 10.1007/s00259-020-05006-3
22. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, et al. A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using (18)F-FDG PET of the Brain. *Radiology* (2019) 290(2):456–64. doi: 10.1148/radiol.2018180958
23. Zhao X, Wang X, Xia W, Li Q, Zhou L, Li Q, et al. A Cross-Modal 3D Deep Learning for Accurate Lymph Node Metastasis Prediction in Clinical Stage T1 Lung Adenocarcinoma. *Lung Cancer* (2020) 145:10–7. doi: 10.1016/j.lungcan.2020.04.014
24. Tau N, Stundzia A, Yasufuku K, Hussey D, Metser U. Convolutional Neural Networks in Predicting Nodal and Distant Metastatic Potential of Newly Diagnosed Non-Small Cell Lung Cancer on FDG PET Images. *AJR Am J Roentgenol* (2020) 215(1):192–7. doi: 10.2214/AJR.19.22346
25. Schmidt-Hansen M, Baldwin DR, Hasler E, Zamora J, Abraira V, Roque IFM. PET-CT for Assessing Mediastinal Lymph Node Involvement in Patients With Suspected Resectable non-Small Cell Lung Cancer. *Cochrane Database Syst Rev* (2014) 11):CD009519. doi: 10.1002/14651858.CD009519.pub2
26. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* (2015) 162(1):W1–73. doi: 10.7326/M14-0698
27. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. *Rethinking the Inception Architecture for Computer Vision*. IEEE (2016) Conference on Computer Vision and Pattern Recognition p. 2818–26.
28. Okada M. Radical Sublobar Resection for Small-Diameter Lung Cancers. *Thorac Surg Clin* (2013) 23(3):301–11. doi: 10.1016/j.thorsurg.2013.04.003
29. Cao J, Yuan P, Wang Y, Xu J, Yuan X, Wang Z, et al. Survival Rates After Lobectomy, Segmentectomy, and Wedge Resection for Non-Small Cell Lung Cancer. *Ann Thorac Surg* (2018) 105(5):1483–91. doi: 10.1016/j.athoracsur.2018.01.032
30. Altorki NK, Yip R, Hanaoka T, Bauer T, Aye R, Kohman L, et al. Sublobar Resection is Equivalent to Lobectomy for Clinical Stage 1A Lung Cancer in Solid Nodules. *J Thorac Cardiovasc Surg* (2014) 147(2):754–62. doi: 10.1016/j.jtcvs.2013.09.065
31. Zhong Y, She Y, Deng J, Chen S, Wang T, Yang M, et al. Deep Learning for Prediction of N2 Metastasis and Survival for Clinical Stage I Non-Small Cell Lung Cancer. *Radiology* (2022) 302(1):200–11. doi: 10.1148/radiol.2021210902
32. Kocak B, Ates E, Durmaz ES, Ulusan MB, Kilickesmez O. Influence of Segmentation Margin on Machine Learning-Based High-Dimensional Quantitative CT Texture Analysis: A Reproducibility Study on Renal Clear Cell Carcinomas. *Eur Radiol* (2019) 29(9):4765–75. doi: 10.1007/s00330-019-6003-8
33. Zhang Q, Liao Y, Wang X, Zhang T, Feng J, Deng J, et al. A Deep Learning Framework for (18)F-FDG PET Imaging Diagnosis in Pediatric Patients With Temporal Lobe Epilepsy. *Eur J Nucl Med Mol Imaging* (2021) 48(8):2476–85. doi: 10.1007/s00259-020-05108-y
34. Wang H, Zhou Z, Li Y, Chen Z, Lu P, Wang W, et al. Comparison of Machine Learning Methods for Classifying Mediastinal Lymph Node Metastasis of non-Small Cell Lung Cancer From (18)F-FDG PET/CT Images. *EJNMMI Res* (2017) 7(1):11. doi: 10.1186/s13550-017-0260-9
35. Zhao W, Yang J, Sun Y, Li C, Wu W, Jin L, et al. 3d Deep Learning From CT Scans Predicts Tumor Invasiveness of Subcentimeter Pulmonary Adenocarcinomas. *Cancer Res* (2018) 78(24):6881–9. doi: 10.1158/0008-5472.CAN-18-0696
36. Lee JJ, Yang H, Franc BL, Iagaru A, Davidzon GA. Deep Learning Detection of Prostate Cancer Recurrence With (18)F-FACBC (Fluciclovine, Axumin(R)) Positron Emission Tomography. *Eur J Nucl Med Mol Imaging* (2020) 47 (13):2992–7. doi: 10.1007/s00259-020-04912-w
37. Ouyang ML, Tang K, Xu MM, Lin J, Li TC, Zheng XW. Prediction of Occult Lymph Node Metastasis Using Tumor-To-Blood Standardized Uptake Ratio

and Metabolic Parameters in Clinical N0 Lung Adenocarcinoma. *Clin Nucl Med* (2018) 43(10):715–20. doi: 10.1097/RLU.0000000000002229

Conflict of Interest: Author Y-GW was employed by General Electric (GE) Healthcare.

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

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

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 13 July 2022
doi: 10.3389/fonc.2022.868186

Imaging-Based Deep Graph Neural Networks for Survival Analysis in Early Stage Lung Cancer Using CT: A Multicenter Study

Jie Lian[1], Yonghao Long[2], Fan Huang[1], Kei Shing Ng[1], Faith M. Y. Lee[3], David C. L. Lam[4], Benjamin X. L. Fang[5], Qi Dou[2] and Varut Vardhanabhuti[1]**

[1] Department of Diagnostic Radiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong SAR, China, [2] Department of Computer Science, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China, [3] Faculty of Medicine, University College London, London, United Kingdom, [4] Department of Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong SAR, China, [5] Department of Radiology, Queen Mary Hospital, Hong Kong, Hong Kong SAR, China

OPEN ACCESS

Edited by:
Shahid Mumtaz, Instituto de Telecomunicações, Portugal

Reviewed by:
Pallavi Tiwari, Case Western Reserve University, United States
Chenbin Liu, Chinese Academy of Medical Sciences and Peking Union Medical College, China

***Correspondence:**
Varut Vardhanabhuti
varv@hku.hk
Qi Dou
qidou@cuhk.edu.hk

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

Received: *02 February 2022*
Accepted: *16 June 2022*
Published: *13 July 2022*

Citation:
Lian J, Long Y, Huang F, Ng KS, Lee FMY, Lam DCL, Fang BXL, Dou Q and Vardhanabhuti V (2022) Imaging-Based Deep Graph Neural Networks for Survival Analysis in Early Stage Lung Cancer using CT: A Multicenter Study. Front. Oncol. 12:868186. doi: 10.3389/fonc.2022.868186

Background: Lung cancer is the leading cause of cancer-related mortality, and accurate prediction of patient survival can aid treatment planning and potentially improve outcomes. In this study, we proposed an automated system capable of lung segmentation and survival prediction using graph convolution neural network (GCN) with CT data in non-small cell lung cancer (NSCLC) patients.

Methods: In this retrospective study, we segmented 10 parts of the lung CT images and built individual lung graphs as inputs to train a GCN model to predict 5-year overall survival. A Cox proportional-hazard model, a set of machine learning (ML) models, a convolutional neural network based on tumor (Tumor-CNN), and the current TNM staging system were used as comparison.

Findings: A total of 1,705 patients (main cohort) and 125 patients (external validation cohort) with lung cancer (stages I and II) were included. The GCN model was significantly predictive of 5-year overall survival with an AUC of 0.732 ($p < 0.0001$). The model stratified patients into low- and high-risk groups, which were associated with overall survival (HR = 5.41; 95% CI:, 2.32–10.14; $p < 0.0001$). On external validation dataset, our GCN model achieved the AUC score of 0.678 (95% CI: 0.564–0.792; $p < 0.0001$).

Interpretation: The proposed GCN model outperformed all ML, Tumor-CNN, and TNM staging models. This study demonstrated the value of utilizing medical imaging graph structure data, resulting in a robust and effective model for the prediction of survival in early-stage lung cancer.

Keywords: lung cancer, graph convolutional networks, cox proportional-hazards, survival prediction, lung graph model

INTRODUCTION

Lung cancer is the leading cause of cancer-related mortality around the world, accounting for more than 1.80 million deaths in 2020 (1). It is commonly accepted that early detection and treatment improve patients' outcomes (2). Although medical imaging technologies such as computed tomography (CT) scan have made significant advances in recent years, accurate diagnosis, particularly of early lung cancer on CT images, and corresponding individual survival prediction remains a challenge. In recent years, using machine learning and deep learning approaches have recently become a promising tool for helping radiologists and physicians improve detection and prognostication (3, 4).

For example, Jin et al. (5) used the convolution neural network (CNN) as a classifier in their computer-aided diagnosis method to detect lung pulmonary nodules on CT images, achieving an accuracy of 84.6% and sensitivity of 82.5% on the Lung Image Database Consortium image collection (LIDC-IDRI). Sangamithraa et al. (6) applied a K-mean learning algorithm for clustering-based segmentation and a back propagation network for classification to achieve an accuracy of 90.7% on their own dataset. Besides, She et al. (7) applied deep learning models with radiomic features as input and achieved a C-index of 0.7 for survival prediction after surgery. While the approaches described above achieved a good level of prediction performance for nodule detection and prognosis, their models have the following limitations. First, the majority of studies used small patient numbers, which resulted in the respective models only performing well on specific datasets, thus limiting generalizability. Second, most of the previous research used strict criteria for their input images; for example, some pre-trained models performed well only on contrast-enhanced CT, although there was a considerable amount of non-contrast CT being used in practice. Additionally, a substantial number of current machine learning models with radiomic features required expert radiologists to manually segment tumors (8–11), which is time consuming, and the relevant findings heavily relied on radiologists' experience. Moreover, the majority of the models was constructed using pixels that focused exclusively on the tumor, without reference to surrounding structures or patient-specific clinical data, despite the fact that they may also contain disease-related information. In clinical practice, clinicians use that additional information to make treatment decisions and risk stratify patients for more accurate treatment and prognosis (12). In essence, these additional features are analogous to "domain knowledge," which has been underutilized in prior research.

Graph convolutional neural network (GCN) (13) is an emerging technique used to tackle data with graph structures, owing to its effectiveness to model relationships across different factors. In graph, nodes are regarded as different entities, while edges present the relationship between each pair of nodes. This approach is unique in that it is able to elegantly incorporate connections from various features. In recent years, graph presentation has been widely used, for instance, social network analysis, language translation, and point cloud, also in the medical field such as vascular segmentation (14) and airway segmentation (15) due to the fact that some organs and systems within the human body are inherently based on graph or network structures (e.g., vascular structures such as retinal vessels) (16, 17). Lungs also inherently have graph structures (18) if we regard every lung lobe as nodes connected by the airway which can be regarded as edges. In theory, the relationship between different parts of the lungs can be modeled and GCN can be applied on lung CT images to tackle clinical problems.

In this study, we developed a graph representation to summarize information of stage I and II lung cancer patients and to forecast their 5-year overall survival rates using CT and clinical data. This study demonstrated the utility of applying medical domain knowledge to create graph structure data and making predictions with state-of-the-art graph convolutional neural network models, which provided a robust and effective model for early stage lung cancer survival prediction.

MATERIALS AND METHODS

Data Description

The Institutional Review Board of Shanghai Pulmonary Hospital has approved this retrospective study protocol and waived the requirement for informed consent for all included patients. The main cohort of the study included consecutive patients who underwent surgery for early stage non-small cell lung cancer (NSCLC) from January 2011 to December 2013. The inclusion criteria were as follows: (I) pathologically confirmed stage (I) and (II) NSCLC, (II) availability of preoperative thin-section CT image data, and (III) complete follow-up of survival data. Patients receiving neoadjuvant therapy were excluded. An external validation set of 125 patients who met our criteria were also retrieved from the NSCLC Radiogenomics (19) dataset (please refer to original reference for related data information). We only used the one single CT image when patient was diagnosed as NSCLC. Both contract and non-contrast CT were included.

Scanning Parameters

The CT scans were performed using Somatom Definition AS+ (Siemens Medical Systems, Germany) and iCT256 (Philips Medical Systems, Netherlands). Detailed scanning parameters can be found in Supplementary Material I. Intravenous contrast was given according to institutional clinical practice. Relevant clinical data were manually extracted from medical records. The follow-up data were acquired from outpatient records and telephone interviews. Overall survival (OS) was defined as the time interval between the date of surgery and the date of mortality or the last follow-up. Recurrence-free survival (RFS) was measured from the time of surgery to the date of recurrence or death or last follow-up (more details can be found at Supplementary Material II).

Lung CT images Segmentation

Lung CT segmentation is a necessary first step in analyzing the pulmonary structures, and it has been regarded as a necessary prerequisite for accurate CT image analysis tasks (20). Before segmentation, every CT data were preprocessed with slice

thickness of 1 mm and matrix of 512×512 mm, following normalization. Several image segmentation approaches were adopted in this project to ensure accurate preparation for the graph modeling and analysis. The 3D airways were segmented using an adaptation of the region-growing method (21), where we randomly picked a seed point from non-background region in the CT image, and neighbor pixels were examined until the borders. The generated airway segments was then skeletonized with a skeleton algorithm (22) to obtain the main structure of the airways. We then applied a searching algorithm to find the four most important points, namely, the root point, the center point, the left point, and the right point (see Supplementary Material III), and segmented a bounding box of 64×64×64 from the original CT to represent the main properties of the corresponding area of the tissue around the airway. Furthermore, for each patient, a public pre-trained UNet (23) model called lung mask (24) was adapted to segment the five lung lobes. In the last step, tumor image was cropped with the bounding box from CT by using the corresponding annotation information provided by radiologists. For each patient, this resulted in images for 10 separate lung structures, namely, five lung lobes, four airway landmarks, and one tumor segment (**Figure 1**).

Graph Building and Graph Convolutional Neural Network Architecture

The very first step in this study is to build meaningful structure of the lung graphs, particularly defining the vertices and their connections. To use the natural structure of the lung, we considered the four airway landmarks and five lung lobe segments as nodes in each graph, and all nodes were connected in their natural ways. To emphasize the significance of the tumor, we added a tumor node to each patient's lung graph, and the tumor node was connected to their corresponding lobes in which the tumor was located. For example, if the tumor was detected on the left upper lobe, the tumor node will be connected to the left upper lobe node. Each CT were modeled as a 10-node graph for further analysis.

For each patient node, a feature vector should be defined to represent the corresponding properties. In this study, we used the pre-trained MedicalNet (25) to get the relevant image features, followed by an average pooling layer to reduce the dimension space to one dimension (1D). The MedicalNet is a collection of ResNet (26) models that have been pre-trained on a variety of large medical datasets and have demonstrated exceptional performance on medical deep learning tasks such as organ segmentation and nodule detection. To keep the feature vectors simpler and more representative, a linear ridge transform method was used to lower the dimension of each node's feature vector from 1,024 to 96 as the final feature vectors on patients' lung graphs (**Figure 2**).

The goal of GCN is to learn the graph or node embedding using the node's neighborhood information with a neural network. Recently, an inductive framework called GraphSage (27), which allows updating node features by sampling and aggregating information from the neighboring nodes, achieved promising performance among various graph neural network topologies on networks. This network was deemed highly suitable for our study, as our lung CT graph was designed to emphasize the interaction within different parts of a patient's lung structure. Therefore, we designed a survival prediction graph neural network predictor composed of SageConv blocks, a mean-readout layer, and a fully

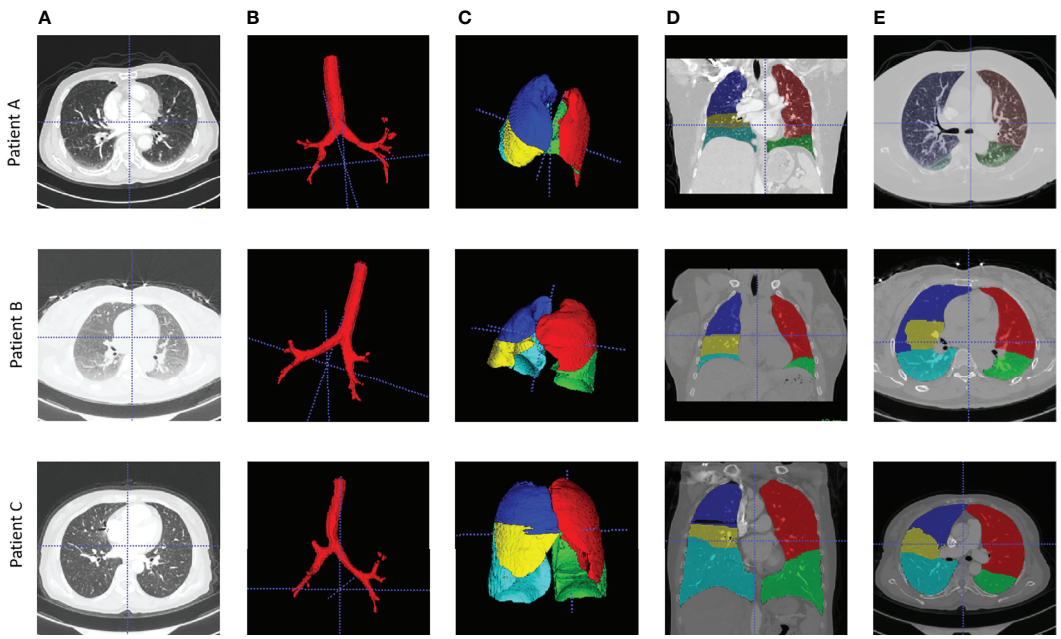


FIGURE 1 | Examples of airway and lung lobes segmentation. **(A)** Patient raw CT scan; **(B)** Airway segments produced by region-growing algorithm; **(C)** lung lobes segments, 3D; **(D)** Lung lobes segments, x-axial 2D; **(E)** Lung lobes segments, z-axial 2D.

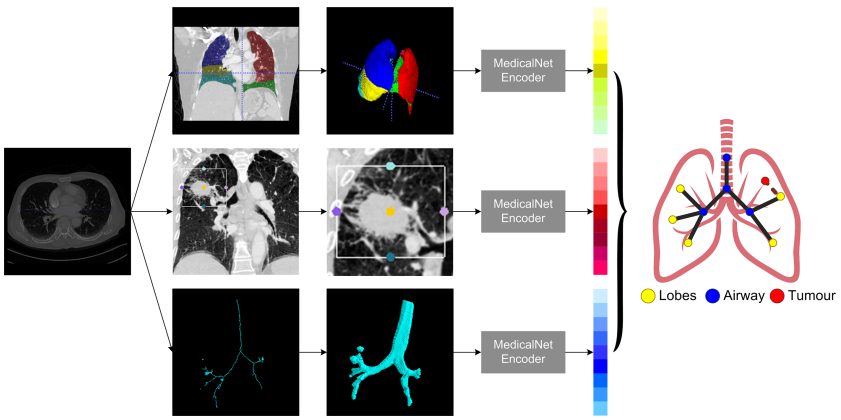


FIGURE 2 | The pipeline of building patients' lung graph building.

connected layer. This model will output a survival label for each patient graph. In detail, the SageConv block consists of a GraphSage Convolution layer with a long short-term memory (LSTM) aggregator, a ReLu activation layer, a dropout layer, and a layer normalization function, which are all efficiently extracting the diagnosis knowledge from the patient graph. The entire model was trained on two GPU nodes in parallel, with a total training epoch of 100. We set a reduced learning rate method to find the optimal training with an initialization value of 0.01 and a minimal value of 0.00001 in order to train the model effectively. In addition, to avoid overfitting when training the model, a weight decay function with value of 0.00005 was added. In order to get the best-performed graph structure, we tested the number of layers of SageConv blocks from 1 to 4, and only the best-performed model was reported.

Experiment Design and Statistical Analysis

To demonstrate the performance of the GCN model on lung cancer survival prediction, a set of experiments were implemented on our dataset. The whole patient cohort was randomly split as training, validation, and testing sets with a ratio of 75% (1278), 12.5% (213), and 12.5% (214) stratified for survival, keeping the survival rate almost equal when splitting the dataset, and there was no significant difference in age and sex among each subset (**Table 1**). We evaluated the performance of the lung graph model by using the area under the receiver operating characteristics (AUC) score, sensitivity, specificity, and precision scores. In order to put emphasis on the model and not to miss the true positive cases, we also added F_2 score (28) as one of the metrics. All relevant results can be found in **Supplementary Table S1**. Wilcoxon rank sums tests were performed to compare performance with baseline model.

In order to see the performance of this graph presentation method with both current clinical assessment and novel deep learning methods, we selected the standard clinical model (TNM staging), commonly used clinical Cox proportional-hazard model, traditional machine learning methods, along with a state-of-the-art deep learning model to make comparison:

1) TNM staging model: using T, N, and M information to make prediction (baseline model I);

2) a Cox proportional-hazard model: using the clinical features (patient sex, age, tumors size, tumors staging, and histology information) as input (baseline model II);

3) a set of machine learning (ML) models: using 103 tumors radiomic features as input (baseline model III), with only the best performer used as the baseline model to be compared;

(4) Tumor-CNN: using individual's tumor segments as input for a ResNet-50 deep neural network.

All models were trained and tested on the same dataset to predict an individual patient's 5-year overall survival, and the best results were reported in comparison to GCN model. We further implemented the survival analysis with Kaplan–Meier estimates for low- and high-risk patients based on the scores predicted by the best three performing models on the testing set, along with a log-rank test. Hazard ratio of our GCN biomarker was calculated by a Cox proportional-hazard model. Finally, a subanalysis was implemented to evaluate the GCN model's performance for predicting overall survival and relapse-free survival on stage I and II patients dataset separately.

All experiments were performed using Python 3.7. The statistics analysis was implemented with the package of Pandas (version 1.3.0) and statistics (version 3.4). Radiomic features were calculated with the PyRadimics package (version 3.0.1). The machine learning models were implemented with the library of Scikit-Learn (version 0.24). Both the Cox regression and the Kaplan–Meier curve were calculated by using the Lifelines package (version 0.26.03). The whole GCN structure was implemented using Deep Graph Library (version 0.6.1) and PyTorch (version 1.8.0).

RESULTS

Patient Information Statistics

A total of 1,705 NSCLC patients were included in the main cohort. There were 1,010 men (59.2%) and 695 women (40.8%) with a median age of 61 years (range: 55–66 years). The median

TABLE 1 | Feature distribution in the total patient cohorts, training and validation cohorts, and the test cohorts.

		Patients Characteristics (*n* = 1,705)	TRAIN and VAL (n = 1,492)	Test (n = 213)	EXTERNAL (n= 125)
Feature	**Content**	**Mean, SD, 95% CI/Count and percentage (%)**			
Age	Age	60.6, 8.8, (CI: 60.2- 61.0)	60.6, 8.7, (CI: 60.1- 61.0)	60.7, 9.5, (CI: 59.4- 62.0)	69.0, 8.90, (CI: 67.4- 70.5)
Sex	Female No. (%);	695 (33.3);	602 (33.3);	93 (33.3);	33 (26.4);
	Male No. (%)	1010 (66.7)	890 (66.7)	120 (66.7)	92 (73.6)
Resection	Sublobar Resection No. (%);	146 (8.6);	123 (8.2);	23 (10.8);	/
	Lobectomy No. (%);	1472 (86.3);	1,292 (86.6);	180 (84.5);	
	Bilobectomy No. (%);	66 (3.9);	59 (3.95);	7 (3.3);	
	Pneumonectomy No. (%)	21 (1.2)	18 (1.2)	3 (1.4)	
Histology	Adenocarcinoma No. (%);	1,235 (72.4);	1,072 (71.4);	163 (76.5);	97 (77.6);
	Squamous Cell Carcinoma No. (%);	391 (22.9);	351 (23.5);	40 (18.8);	26 (20.8);
	Others No. (%)	79 (4.6)	69 (4.6)	10 (4.7)	2 (1.6)
Tumor Size	Tumor Size	2.66, 1.37, (CI: 2.60- 2.73)	2.68, 1.38, (CI: 2.61- 2.75)	2.55, 1.25, (CI: 2.38-2.71)	/
pTNM stage	Stage I No. (%);	1,398 (82.0);	1,219 (81.7);	179 (84.0);	63 (50.4);
	Stage II No. (%)	306 (18.0)	273 (18.3)	34 (16.0)	62 (49.6)
RFS Status	RFS No. (survival %)	1,243 (72.9)	1,089 (73.0)	154 (72.3)	93 (74.4)
RFS Month	RFS Month	57.6, 24,4, (CI: 56.4- 58.7)	57.5, 24.5, (CI: 56.2- 58.7)	58.4, 23.4, (CI: 55.2- 61.5)	/
OS Status	OS No. (survival %)	1,333 (78.2)	1,166 (78.2)	167 (78.4)	79 (63.2)
OS Month	OS Month	62.5, 19.8, (CI: 61.6- 63.5)	62.4, 19.9, (CI: 61.4- 63.4)	63.4, 18.4, (CI: 60.9- 65.9)	/

follow-up time is 70.9 months. Of these, 145 patients (8.5%) received sub-total lobectomy, 1,472 patients (86.3%) underwent lobectomy, 66 patients (3.9%) received bi-lobectomy, 21 patients (1.2%) underwent pneumonectomy, and one patient received sub-total lobectomy of one lobe plus total lobectomy of another lobe. Tumors were most commonly located in the upper lobe [419 left upper lobe (LUL), 24.6%, and 565 right upper lobe (RUL), 33.1%]. A total of 1,235 tumors (72.4%) were diagnosed as adenocarcinoma, and 391 tumors (22.9%) were squamous cell carcinoma. The distribution of pathological stages was as follows: stage IA in 791 patients (46.4%), stage IB in 607 patients (35.6%), stage IIA in 133 patients (7.8%), and stage IIB in 174 patients (10.2%). In the whole main cohort, the 3-year OS and RFS were 98.4% and 81.1%, respectively, and the 5-year OS and RFS were 78.2% and 74.2%, respectively.

There were 33 (26.4%) female and 92 (73.6%) male patients in the external validation dataset, with a median age of 69 (range, 43–87 years). Tumors in the upper lobe were also the most common [41 at right upper lobe (RUL), 32.8%, and 32 at left upper lobe (LUL), 25.6%]. There were 97 patients with adenocarcinoma and 26 with squamous cell carcinoma among them. The pathological stages were distributed as follows: stage IA in 40 patients (32.0%), stage IB in 23 patients (18.4%), stage IIA in 45 patients (36.0%), and stage IIB in 17 patients (13.6%). The RFS was 74.4%, while the 5-year OS was 63.2%. **Table 1** provides the rest of the patient's detailed information.

Model Evaluation

As shown in **Table 2**, the Cox modeling and ML radiomic feature baseline models showed poor performance on the testing set. The best performing ML radiomic model was from the decision tree (DT) model, while other ML models such as SVM, linear classification, K-means, LASSO, and KNN methods had worse performance than the DT predictor. The Tumor-CNN model had a significantly improved performance (AUC=0.614; 95% CI: 0.519–0.710; $p < 0.05$) compared with the two baseline models, although the TNM method performed better (AUC=0.633; 95% CI: 0.539–0.728; $p < 0.005$). The GCN model achieved the highest AUC score of 0.732 (95% CI: 0.643–0.821; $p < 0.0001$) among all models in survival prediction for early-stage lung cancer. On external validation dataset, our GCN model achieved the AUC score of 0.678 (95% CI: 0.564–0.792; $p < 0.0001$).

For survival analysis, both GCN the cancer staging system and Tumor-CNN shared a similar trend and, based on Kaplan–Meir analysis, were able to demonstrate significant separation of high- and low-risk groups (**Figure 3**), while the p-value of the log rank sums test suggested that GCN has a stronger separation ability compared with the others. Comparable results were found in the prediction of 5-year survival outcomes with the hazard ratios, respectively, for GCN (HR = 5.41; 95% CI: 2.32–10.14; p=0.000014), and TNM (HR = 3.85; 95% CI: 1.91–7.02; p=0.00015).

For the stage I dataset (n=179) analysis, as per **Figure 4**, our GCN model achieved a clear separation of low- and high-risk groups in 5-year overall survival prediction ($p < 0.0001$) and relapse-free survival prediction ($p < 0.0001$), with AUC of 0.728 (CI: 0.618–0.839) and 0.660 (CI: 0.555–0.757) separately. Referencing stage II (n=55), the model showed slightly weaker performance of separation for 5-year overall survival (AUC = 0.647, CI: 0.461–0.834, $p = 0.132$) comparing with stage I dataset, while better performance for relapse-free survival prediction (AUC = 0.702, CI: 0.532–0.877, $p < 0.01$) was achieved.

TABLE 2 | Performance for each model based on AUC scores and the Wilcoxon rank-sum tests.

ML models	AUC scores (95% CI)	p-values
CPH Model	0.549 (0.454–0.645)	.45
DT-radiomics	0.572 (0.476–0.668)	.33
Tumor-CNN	0.614 (0.519–0.710)	.02
TNM	0.633 (0.539–0.728)	.002
GCN	0.732 (0.643–0.821)	< 0.0001

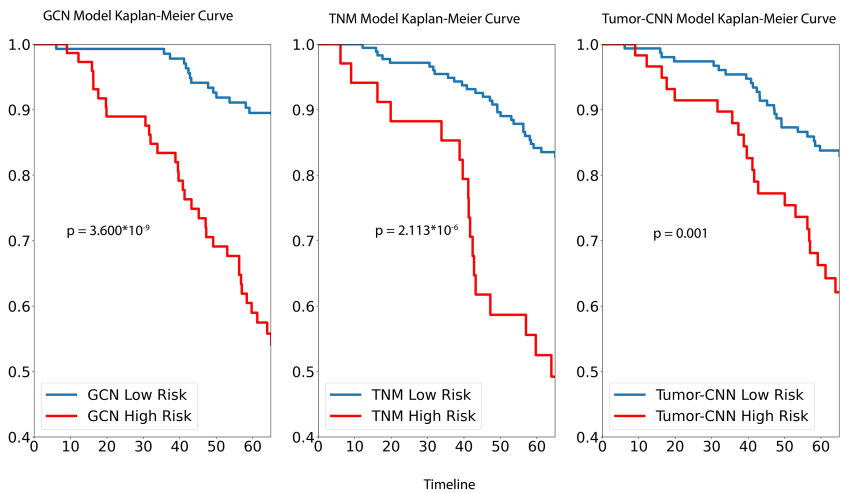


FIGURE 3 | Performance of GCN, TNM and Tumor-CNN models on testing dataset.

DISCUSSION

Prediction of survival of early-stage lung cancer patients remains a challenging task. In this paper, we proposed a graph-based method to represent a patient's lung CT images and applied the state-of-the-art graph convolutional neural network to improve 5-year survival predictions for individual patients. In previous studies, especially for some small size cohorts, the radiomic feature methods (our baseline models) were commonly used. The results in this study showed that when applied to a large patient cohort in which CTs were collected from multiple data sources, this radiomic feature method demonstrated poor performance, which may be due to the heterogeneity in image acquisition, reconstruction methods, or effects of post-processing.

Deep learning approaches have demonstrated impressive performance in recent years in medical fields such as automatic segmentation and diagnostic task such as lung nodule detection. Due to the fact that deep learning models are generally robust and can be applied to a wide variety of scenarios once properly trained with enough data, it has been previously applied to the task of survival prediction. In this project, we applied a ResNet-50 deep neural network, which took tumor segments (Tumor-CNN model) as input resulting in an AUC score of 0.6144. When analyzing the Tumor-CNN model's performance from the medical perspective, we demonstrated that tumors contained the majority of prognostic information, yet adjacent non-tumor regions and their interactions with each other may have an effect on an individual patient's survival. This hypothesis was based on our intuition that tumors spread from the primary sites *via* lymphatic drainage, hematogenous (*via* the vascular supply) or directly to the surrounding lungs (29). We therefore reasoned that such regional information can potentially be mapped *via* a graph representation method to represent the entire lung as input with an emphasis on the tumors as an additional node on an individual patient's basis. Moreover, the best performance achieved by our GCN model demonstrated that using a relational data representation method can help improve the performance when compared to traditional deep learning models. To this end, our model demonstrated best accuracy in identifying high-risk patients, particularly on stage I patient group, demonstrating that features generated by GCN can find the survival-relevant information from early-stage patients' CT image. The RFS Kaplan–Meier analysis revealed that the GCN approach also contained information that related

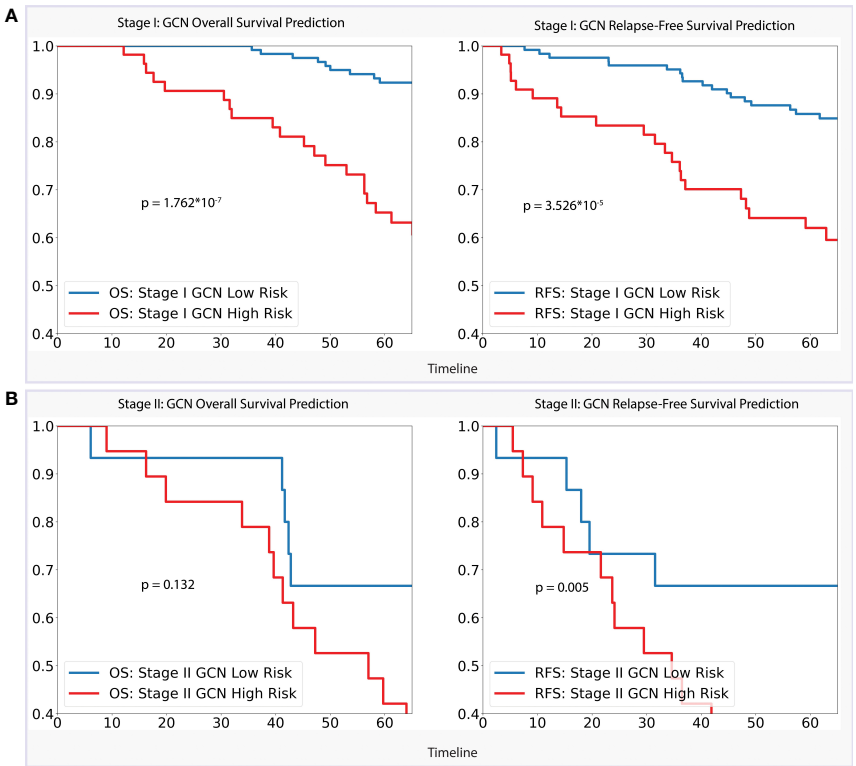


FIGURE 4 | **(A)**. Stage I Analysis: Performances of GCN models on OS prediction and RFS prediction separately; **(B)**. Stage II Analysis: Performances of GCN models on OS prediction and RFS prediction separately.

to disease relapse, and combining that information from both of the above two aspects to analyze individual's survival result likely contributes to improve performance.

On reviewing the whole process of our graph survival predictor formulation, all the steps were fully automated and could be easily applied to prospective patients in the future. Unlike radiomic approach, there was no need to specifically segment the tumors with our proposed method. By including regional information in graph structures likely contributes to improved prediction performance.

The results from our study have the following strengths. First, our dataset is large and has incorporated images from one large volume center with a standardized acquisition method, including contrast and non-contrast CT scans. Our model was found to be more generalizable as a result of training based on this large dataset with reasonable performance on external validation set. Second, our model's whole procedure was fully automated. For example, segmenting the lung and airway took only a few seconds to obtain accurate results, which would allow ease of clinical translation. Finally, we conducted a series of experiments comparing our graph model to traditional model, widely used radiomic approaches and the most cutting-edge deep learning models, which supported our conclusion that the GCN models can outperform other conventional methods. We acknowledge, however, that due to differences in input features between these different models, comparison of performance may not be a fair one.

There are a few limitations in our study. First, while we achieved the best performance with the graph neural network, we did not investigate the model's ability to discover new features, but it was apparent from our results that graph models have greater potential for future development due to their input of relational graph structures. Second, we used only CT images as input in this experiment because we have yet to develop a method for incorporating imaging data with demographic data such as age and gender information, which may improve the model's performance. Some future work is being planned to improve the performance of our models. More anatomically relevant information could be incorporated into the graphs. For example, one could consider edge weight based on the location of the tumors for individual patients and create some other lung graph structures to better represent patients' survival information. Furthermore, we intend to combine whole-slide imaging data from lung patients with CT data to better represent disease information in the future.

In this study, we presented a graph presentation model for describing CT data from early stage lung cancer patients and predicting their 5-year overall survival. Numerous experiments were conducted to compare our GCN model to traditional clinical model based on TNM staging, commonly used

radiomic feature approaches, and state-of-the-art deep learning methods. We demonstrated that our graph methods performed significantly better compared with other existing models.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Tongji University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conception and design: JL, QD, and VV. Administrative support: QD and VV. Provision of study materials or patients: JL, FL, BF, and DL. Collection and assembly of data: JL and YL. Data analysis and interpretation: JL, YL, FH, and KSN. Manuscript writing: all authors. Final approval of manuscript: all authors. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

ACKNOWLEDGMENTS

We would like to thank Yunlang She, Jiajun Deng, and Chang Chen from Tongji University School of Medicine for their contribution to this project. To aid reproducibility of research, our codes are published on the Github repository: https://github.com/SereneLian/Lung_Graph.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.868186/full#supplementary-material

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Mirsadraee S, Oswal D, Alizadeh Y, Caulo A, van Beek EJr. The 7th Lung Cancer TNM Classification and Staging System: Review of the Changes and Implications. *World J Radiol* (2012) 4(4):128–34. doi: 10.4329/wjr.v4.i4.128
3. Xu Y, Hosny A, Zeleznik R, Parmar C, Coroller T, Franco I, et al. Deep Learning Predicts Lung Cancer Treatment Response From Serial Medical Imaging. *Clin Cancer Res* (2019) 25(11):3266–75. doi: 10.1158/1078-0432.CCR-18-2495
4. Wang S, Liu Z, Chen X, Zhu Y, Zhou H, Tang Z, et al. (2018) Unsupervised Deep Learning Features for Lung Cancer Overall Survival Analysis. 2018., in: *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018/07. IEEE.
5. Jin X-Y, Zhang Y-C, Jin Q-L. Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network. in: *2016 9th International Symposium on Computational Intelligence and Design (ISCID)*, 2016/12. IEEE.
6. Sangamithraa PB, Govindaraju S. (2016) 2016. Lung Tumour Detection and Classification Using EK-Mean Clustering, in: *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2016/03. IEEE (2019).
7. She Y, Jin Z, Wu J, Deng J, Zhang L, Su H, et al. Development and Validation of a Deep Learning Model for Non-Small Cell Lung Cancer Survival. *JAMA Network Open* (2020) 3(6):e205842–e. doi: 10.1001/jamanetworkopen.2020.5842
8. Altintas Z, Tothill I. Biomarkers and Biosensors for the Early Diagnosis of Lung Cancer. *Sensors Actuators B: Chem* (2013) 188:988–98. doi: 10.1016/j.snb.2013.07.078
9. Buizza G, Toma-Dasu I, Lazzeroni M, Paganelli C, Riboldi M, Chang Y, et al. Early Tumor Response Prediction for Lung Cancer Patients Using Novel Longitudinal Pattern Features From Sequential PET/CT Image Scans. *Physica Medica* (2018) 54:21–9. doi: 10.1016/j.ejmp.2018.09.003
10. Wang T, Deng J, She Y, Zhang L, Wang B, Ren Y, et al. Radiomics Signature Predicts the Recurrence-Free Survival in Stage I Non-Small Cell Lung Cancer. *Ann Thorac Surg* (2020) 109(6):1741–9. doi: 10.1016/j.athoracsur.2020.01.010
11. Wolf M, Holle R, Hans K, Drings P, Havemann K. Analysis of Prognostic Factors in 766 Patients With Small Cell Lung Cancer (SCLC): The Role of Sex as a Predictor for Survival. *Br J cancer* (1991) 63(6):986–92. doi: 10.1038/bjc.1991.215
12. Liao Y, Wang X, Zhong P, Yin G, Fan X, Huang C. A Nomogram for the Prediction of Overall Survival in Patients With Stage II and III Non-Small Cell Lung Cancer Using a Population-Based Study. *Oncol lett* (2019) 18(6):5905–16. doi: 10.3892/ol.2019.10977
13. Kipf TN, Welling M. Semi-Supervised Classification With Graph Convolutional Networks. *arXiv preprint arXiv* (2016), 160902907. doi: 10.48550/arXiv.1609.02907
14. Chapman BE, Berty HP, Schulthies SL. Automated Generation of Directed Graphs From Vascular Segmentations. *J Biomed Inform* (2015) 56:395–405. doi: 10.1016/j.jbi.2015.07.002
15. Qin Y, Chen M, Zheng H, Gu Y, Shen M, Yang J, et al. AirwayNet: A Voxel-Connectivity Aware Approach for Accurate Airway Segmentation Using Convolutional Neural Networks. Medical Image Computing and Computer Assisted Intervention – MICCAI (2019): *22nd International Conference*, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI; Shenzhen, China: Springer-Verlag; . p. 212–20.
16. Pan H, Gu J, Han Q, Feng X, Xie X, Li P. Medical Image Clustering Algorithm Based on Graph Model. In: K Li, Z Xiao, Y Wang, J Du, K Li*Parallel Computational Fluid Dynamics. ParCFD 2013. Communications in Computer and Information Science* (2014) 405:54–65. Springer, Berlin: Heidelberg. doi: 10.1007/978-3-642-53962-6_5
17. Dicente Cid Y, Jiménez-del-Toro O, Platon A, Müller H, Poletti P-A. From Local to Global: A Holistic Lung Graph Model. In:A Frangi, J Schnabel, C Davatzikos, C Alberola-López, G Fichtinger (eds) *Med Image Computing Comput Assisted Intervention – MICCAI 2018:* (2018) 11071:786–93. Springer: Cham. doi: 10.1007/978-3-030-00934-2_87
18. Dicente Cid Y, Müller H, Platon A, Janssens JP, Lador F, Poletti PA, et al. A Lung Graph–Model for Pulmonary Hypertension and Pulmonary Embolism Detection on DECT Images. *Medical Computer Vision and Bayesian and Graphical Models for Biomedical Imaging. BAMBI MCV 2016* Lecture Notes in Computer Science (2016) 10081:58–68. doi: 10.1007/978-3-319-61188-4_6
19. Bakr S, Gevaert O, Echegaray S, Ayers K, Zhou M, Shafiq M, et al. A Radiogenomic Dataset of non-Small Cell Lung Cancer. *Sci Data* (2018) 5 (1):1–9. doi: 10.1038/sdata.2018.202

20. Armato S, McLennan G, McNitt-Gray M, Meyer C, Reeves A, Bidaut L, et al. WE-B-201b-02: The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Public Database of CT Scans for Lung Nodule Analysis. *Med Phys* (2010) 37(6Part6):3416–7. doi: 10.1118/1.3469350
21. Lo P, van Ginneken B, Reinhardt JM, Yavarna T, de Jong PA, Irving B, et al. Extraction of Airways From CT (Exact'09). *IEEE Trans Med Imaging* (2012) 31(11):2093–107. doi: 10.1109/TMI.2012.2209674
22. Xie W, Thompson RP, Perucchio R. A Topology-Preserving Parallel 3D Thinning Algorithm for Extracting the Curve Skeleton. *Pattern Recognit* (2003) 36(7):1529–44. doi: 10.1016/S0031-3203(02)00348-5
23. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect Notes Comput Sci: Springer Int Publ* (2015) p:234–41. doi: 10.1007/978-3-319-24574-4_28
24. Hofmanninger J, Prayer F, Pan J, Röhrich S, Prosch H, Langs G. Automatic Lung Segmentation in Routine Imaging Is Primarily a Data Diversity Problem, Not a Methodology Problem. *Eur Radiol exp* (2020) 4(1):50–. doi: 10.1186/s41747-020-00173-2
25. Chen S, Ma K, Zheng Y. *Med3D: Transfer Learning for 3D Medical Image Analysis2019 April 01* (2019). Available at: https://ui.adsabs.harvard.edu/abs/2019arXiv190400625C.
26. Khanna A, Londhe ND, Gupta S, Semwal A. A Deep Residual U-Net Convolutional Neural Network for Automated Lung Segmentation in Computed Tomography Images. *Biocybern Biomed Eng* (2020) 40(3):1314–27. doi: 10.1016/j.bbe.2020.07.007
27. Hamilton WL, Ying R, Leskovec J. (2017)., in: Proceedings of the 31st International Conference on Neural Information Processing Systems (2017) Long Beach, California, USA: Curran Associates Inc; p. 1025–35.
28. Goutte C, Gaussier E. Inductive Representation Learning on Large Graphs. In: DE Losada, JM Fernández-Luna (eds) *Advances in Information Retrieval. ECIR 2005. Lecture Notes in Computer Science* 3408. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-540-31865-1_25
29. Lee GM, Stowell JT, Pope K, Carter BW, Walker CM. Lymphatic Pathways of the Thorax: Predictable Patterns of Spread. *AJR Am J Roentgenol* (2021) 216 (3):649–58. doi: 10.2214/AJR.20.23523

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 | Frontiers in Oncology

TYPE Original Research
PUBLISHED 22 July 2022
DOI 10.3389/fonc.2022.956705

OPEN ACCESS

EDITED BY
Shahid Mumtaz,
Instituto de Telecomunicações, Portugal

REVIEWED BY
Maxim A. Dulebenets,
Florida Agricultural and Mechanical University, United States
Marcin Wozniak,
Silesian University of Technology, Poland

*CORRESPONDENCE
Shanshan Kong
kongss@ncst.edu.cn

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

RECEIVED 30 May 2022
ACCEPTED 28 June 2022
PUBLISHED 22 July 2022

CITATION
Liu J, Zhou Z, Kong S and Ma Z (2022) Application of random forest based on semi-automatic parameter adjustment for optimization of anti-breast cancer drugs. *Front. Oncol.* 12:956705. doi: 10.3389/fonc.2022.956705

Application of random forest based on semi-automatic parameter adjustment for optimization of anti-breast cancer drugs

Jiajia Liu[1,2,3], Zhihui Zhou[1,3,4], Shanshan Kong[1,2,3,4,5]* and Zezhong Ma[1,2,5]

[1]College of Science, North China University of Science and Technology, Tangshan, China, [2]Hebei Engineering Research Center for the Intelligentization of Iron Ore Optimization and Ironmaking Raw Materials Preparation Processes, North China University of Science and Technology, Tangshan, China, [3]The Key Laboratory of Engineering Computing in Tangshan City, North China University of Science and Technology, Tangshan, China, [4]Hebei Key Laboratory of Data Science and Application, North China University of Science and Technology, Tangshan, China, [5]Tangshan Intelligent Industry and Image Processing Technology Innovation Center, North China University of Science and Technology, Tangshan, China

The optimization of drug properties in the process of cancer drug development is very important to save research and development time and cost. In order to make the anti-breast cancer drug candidates with good biological activity, this paper collected 1974 compounds, firstly, the top 20 molecular descriptors that have the most influence on biological activity were screened by using XGBoost-based data feature selection; secondly, on this basis, take pIC50 values as feature data and use a variety of machine learning algorithms to compare, soas to select a most suitable algorithm to predict the IC50 and pIC50 values. It is preliminarily found that the effects of Random Forest, XGBoost and Gradient-enhanced algorithms are good and have little difference, and the Support vector machine is the worst. Then, using the Semi-automatic parameter adjustment method to adjust the parameters of Random Forest, XGBoost and Gradient-enhanced algorithms to find the optimal parameters. It is found that the Random Forest algorithm has high accuracy and excellent anti over fitting, and the algorithm is stable. Its prediction accuracy is 0.745. Finally, the accuracy of the results is verified by training the model with the preliminarily selected data, which provides an innovative solution for the optimization of the properties of anti- breast cancer drugs, and can provide better support for the early research and development of anti-breast cancer drugs.

KEYWORDS

anti-breast cancer, parameter optimization, random forest, xgboost, bioactivity

1 Introduction

At the present, stroke, ischemic heart disease and other cardiovascular diseases, as well as malignant tumors represented by breast cancer have become the main cause of premature death in our population, seriously threatening human health. Global incidence rate and mortality associated with breast cancer have been increasing (1), and breast cancer has officially replaced lung cancer as the number one cancer worldwide (2), and its incidence rate among women's cancers worldwide is as high as 24.2%, becoming one of the most common cancers in women (3–5), which seriously affects women's health (6). Estrogen receptor is a hormone receptor and an effective nonstandard RNA binding protein. It is a biomarker of breast cancer and affects the choice of endocrine therapy for breast cancer. It has a very important role in the process of breast development. It is considered as an important target for the treatment of breast cancer and plays an important role in the treatment of breast cancer. Therefore, compounds that can antagonize the activity of ERα may be drug candidates for the treatment of breast cancer.

Active compounds are compounds that can have an effect on disease sites in the human body, or compounds with more pronounced pharmacological effects and clear structures, which are widely used in research fields such as cancer, stem cells and immunity. Molecular descriptors refer to the measurement of the properties of molecules in a certain aspect, which can be either the physical and chemical properties of molecules or the numerical indicators derived from various algorithms according to the molecular structure. The target is a kind of biological macromolecule which has pharmacodynamic function and can be acted by drugs. The biological activity of a compound refers to its ability to bind, inhibit or activate the target. The higher the biological activity, the stronger the ability of the compound. Quantitative Structure-Activity Relationship aims to establish the quantitative relationship between the physiological activities or some properties of a series of compounds and their physical and chemical property parameters or structural parameters through reasonable mathematical statistical methods.

Currently, in the research and development of cancer drugs, the process of screening and developing new drugs through experiments is very slow and requires a lot of manpower and material resources, how to effectively and quickly select drugs to treat breast cancer and improve the therapeutic effect has become an important topic in the research of cancer drugs. In order to save time and cost, the method of establishing compound activity prediction models are usually used to screen potentially active compounds. The specific method is as follows: for a target related to disease, collect a series of compounds acting on the target and their biological activity data, and then take a series of analytical structural descriptors as independent variables and the biological activity value of compound as dependent variables to build a Quantitative Structure-Activity Relationship (QSAR) model of the compound. Then using the model to predict new molecules with better bioactivity, or to guide the structural optimization of existing active compounds.

This paper presents a four-part study on the optimization of anti-breast cancer drug properties. (1) In section 1, this paper analyzes the research that has been completed in related fields, the application of artificial intelligence algorithm and describes the research content of this paper. (2) In section 2, the theoretical basis of the used algorithm is described. (3) In section 3, the data set, preprocessing of the data as well as the analysis of the results are presented, including the XGBoost-based data feature selection results and prediction results of Quantitative Structure-Activity Relationship (QSAR) model. (4) In section 4, a summary of the work done throughout the text is presented.

2 Related work

B Zhao (7) considered the independence, coupling and correlation of bioactivity descriptors to screen the most potentially valuable bioactivity descriptors, and then used an optimized back propagation neural network pair to make predictions, and used a gradient boosting algorithm to verify the pharmacokinetics and safety of the screened bioactivity descriptors, and the results showed that the bioactivity descriptors screened by this method not only fit the non-linear relationship of activity well, but also accurately predicted their pharmacokinetic characteristics and safety. The results showed that the screened bioactivity descriptors could not only fit the nonlinear relationship of activity, but also accurately predict the pharmacokinetic characteristics and safety, with an average accuracy of 89.92 ~ 94.80%. S Leya, P N Kumar (8) established a deep learning based cancer drug screening model to predict the activity in the GDB13 data set after confirming the importance of synergy between an effective mimetic drug or compound and its target, and achieved good results in identifying anti-cancer drugs with improved performance metrics. B Xu (9) constructed a QSAR model based on three traditional neural network models (BP neural network, Elman neural network and wavelet neural network) and a neural network model improved by optimization algorithm (SSA-BP neural network), and the results showed that the BP neural network can predict the biological activity of compounds more accurately, and the optimized model can further improve the predictive performance of the BP neural network, which can help to better screen efficient compound molecules and guide the structural optimization of existing active compounds and the development of quality breast cancer drugs. X Liu, W Zhang,

W Zheng, et al. (10) considered that the low content of traditional drug screening platforms limits the process of drug evaluation, so it proposed a micropatterned co-culture-based high content (μCHC) platform to study neuronal cancer cell interactions and drug screening, and finally obtained a high efficiency and fidelity of clinical cancer treatment by screening drug candidates or drug combinations through the μCHC system. Y Zhu, T Brettin, Y A Evrard, et al. (11) extended the classical transfer learning framework by integration and demonstrated its general utility with a gradient advancement model and two deep neural networks for three representative prediction algorithms, and finally tested the integrated transfer learning framework on an *in vitro* drug screening benchmark dataset, and the results showed that the established framework extensively improved the prediction algorithms in prediction applications prediction performance. Z Xiong, D Wang, X Liu, et al. (12) used a new graph neural network structure (Attentive FP), which uses a graph attention mechanism to achieve learning from relevant drugs in a data set, and experimental results showed that Attentive FP achieved state-of-the-art prediction performance in various datasets. P Wongyikul, N Thongyot, P Tantrakoolcharoen, et al (13) developed an had screening protocol using Gradient Boosting Classifier machine learning model and screening parameters to identify HAD prescription error events from drug prescriptions. The experimental results show that machine learning plays an important role in screening and reducing HAD prescription errors and has potential benefits. D FernándezLlaneza, S Ulander, D Gogishvili, et al. (14) proposed a Siamese recurrent neural network model (SiameseCHEM) based on bidirectional long-term and short-term memory structure with self attention mechanism, which can automatically learn the discriminant features from the SMILES representation of small molecules. Then it is trained with random SMILES strings, which proves that it is robust to binary or classification tasks of biological activity. M Kumari, N Subbarao (15) proposed a new deep learning based approach to implement virtual screening with convolutional neural network architecture as a way to predict the inhibitory activity of 3CLpro against unknown compounds during SARS-CoV virtual screening. Experimental results show that their proposed convolutional neural network model can prove useful for the development of novel target-specific anti-SARS-CoV compounds. A Abdo, M Pupin (16) proposed a turbine prediction model using nearest neighbor structure to improve prediction accuracy in order to study how to use learning data to enhance prediction model. The experimental results show that Turbo prediction can improve the prediction quality of the traditional prediction model. For heterogeneous data sets, it can predict with minimal computational cost without additional efforts of users. A Gupta, H Zhou (17) accelerated the screening of drugs by opening a machine-learning driven large-scale virtual screening pipeline in order to handle the growing library of drug-like compounds and to separate true positives from false positives. K Carpenter, A Pilozzi, X Huang (18) created a virtual screener for protein kinase inhibitors and achieved prediction of IC50 values for target compounds by transforming and feeding the data as input into two majority-invariant recurrent neural networks (RNN).

With the progress of science and technology, the development of artificial intelligence technology is changing with each passing day. Its application fields are very wide, and it can be effectively applied to all fields of production and life. Of course, the application advantages of artificial intelligence are very obvious. More and more enterprises are committed to the R&D and application of artificial intelligence. With the deepening of research, the application rate and popularity of artificial intelligence technology are also gradually increasing. For example, artificial intelligence can be applied to online learning. By applying artificial intelligence technology and educational psychology theory, personalized online learning resource recommendation schemes can be designed to improve students' learning outcomes (19). Artificial intelligence can also be applied to multi-objective optimization. For example, in order to improve the existing technology of proton exchange membrane fuel cell (PEMFC), multi-objective optimization based on artificial intelligence can be adopted to facilitate the design and application of PEMFC (20). To address the issue of geographic emergency evacuation of vulnerable population groups, multi-objective planning can be used to improve the safety of evacuees during the natural disaster preparation phase and to ensure timely evacuation from areas expected to be affected by major natural disasters (21). Artificial intelligence can also be applied to the field of transportation. With highly interconnected road networks placing higher demands on road safety and efficiency, intelligent transportation systems have received widespread attention. Artificial intelligence technology can provide various support for road routing and traffic congestion management, and can effectively support intelligent transportation systems (22). Artificial intelligence can also be applied to several fields in the medical field, such as neural disease prediction and modeling, bioinformatics, surgery, physical rehabilitation, medical robot and hospital clinical data management (23). The most basic is the grass-roots medical institutions, which are the first line of defense for the health of grass-roots residents. Its informatization construction is an important means to realize the modernization of medical services. Artificial intelligence technology can promote the informatization of grass-roots medical institutions, so as to optimize the process of medical treatment, improve the service capacity of high-quality medical resources and reduce costs (24). Artificial intelligence technology can be applied to the financial field. With the continuous expansion of the scale, quantity and scope of international trade and the increase of trade complexity and

uncertainty, artificial intelligence technology can predict and select international trade and play an important role in the healthy development of international trade (25). The Financial Stability Board (FSB) also released the development of artificial intelligence and machine learning in the financial service market and their impact on financial stability. Artificial intelligence and machine learning can certainly strengthen financial supervision (26). Artificial intelligence techniques can also be applied in industry, for example, the surface roughness induced by grinding operations can affect the corrosion resistance, wear resistance, and contact stiffness of ground parts, which can be predicted using artificial intelligence algorithms, helping to provide real-time feedback control of grinding parameters for the purpose of reducing production costs (27). Artificial intelligence techniques can also protect the network from data transmission, for example, P Rani, Kavita, S Verma, et al. (28) proposed a new update routing protocol combining the advantages of artificial bee colony, artificial neural network and support vector machine techniques as a way to protect the network from black hole attacks. Artificial intelligence techniques can also be applied in the field of scheduling, for example, to solve the scheduling problem of CDT trucks, M Dulebenets (29) proposed a new adaptive multiplicative modal algorithm, which can assist in the correct planning of CDT jobs. Artificial intelligence techniques can also improve algorithms; for example, to address the problem of Gaussian noise impeding the unbiased aggregation capability of GNN models, W Dong, M Wozniak, J Wu, et al., (30) proposed a method that uses principal component analysis to retain the aggregated true signal from adjacent features and simultaneously removes filtered Gaussian noise to achieve a more advantageous denoising capability.

The main contributions of this paper are as follows: XGBoost is used for data feature selection so as to select the 20 molecular descriptors with the most significant impact, and then the 20 molecular descriptors screened are used as input variables and the pIC50 value as output variables from the perspective of the compound molecular descriptors, and four machine learning algorithms, namely Gradient-enhanced regression, XGBoost regression, Support vector machine, and Random Forest regression are used for comparison. The results of Random Forest regression, XGBoost regression and Gradient-enhanced regression are preliminarily screened out to be good. Then the Semi-automatic parameter adjustment method is used to adjust the parameters of the three algorithms, and subsequently the most appropriate algorithms is selected to determine the core algorithm of the prediction model as a way to predict the IC50 and pIC50 values. The highest accuracy rate of 74.5% is finally obtained for Random Forest regression, and the Random Forest algorithm is considered to be the core algorithm.

3 Theoretical foundation

3.1 XGBoost-based data feature selection

Feature selection refers to the selection of some effective features from the original features to reduce the dimensionality of the data set (31). XGBoost is an integrated learning model that can fit the residuals of the previous tree by generating a new tree in successive iterations and its accuracy increases with the number of iterations (32), which can be effectively used for classification and regression (33).

XGBoost is an improvement of the Gradient boosting algorithm by using Newton's method when solving the extrema of the loss function, Taylor expansion of the loss function to the second order, and additionally a regularization term is added to the loss function. The objective function at training time consists of two parts, the first part is the Gradient boosting algorithm loss and the second part is the regularization term. The loss function is defined as:

$$L(\varnothing) = \sum_{i=1}^{n} \left(y_i^{'}, y_i \right) + \sum_{k} \Omega(f_k)$$

Where n is the number of training function samples, l is the loss for a single sample, which is assumed to be a convex function, $y_i^{'}$ is the predicted value of the model for the training samples, and y_i is the true label value of the training samples.

The regularization term defines the complexity of the model:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \| w \|^2$$

Where λ and λ are manually set parameters, w is a vector formed by the values of all leaf nodes of the Decision tree, and T is the number of leaf nodes.

3.2 Random forest

Random Forest is a supervised learning algorithm which tends to find the best grouping features recursively (34). The "forest" it builds is an integration of Decision tree, which is mostly trained using Bagging methods. The Bagging method uses randomly selected training data with playback and then constructs a classifier, and finally combines the learned models to increase the overall effect.

The growth of the tree in the Random Forest algorithm introduces additional randomness to the model. Unlike Decision tree where each node is partitioned into the best features that minimize the error, in a Random Forest we randomly select features to construct the best partition. Thus, when you are in a Random Forest, consider only the random subset used to segment the nodes, or even make the tree more random by using a random threshold on each feature instead of searching

for the best threshold as in a normal Decision tree. This process yields a wide range of diversity and usually leads to better models, and the Random Forest algorithm proceeds as follows:

The input is the sample set D={ $(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)$ } and the number of weak classifier iterations T.

The output is the final strong classifier f(x).

1) For t=1,2,⋯,T: divided into two steps. a) The training set is randomly sampled for the tth time, and a total of m times are taken to obtain the sampling set D_t containing m samples. b) Train the first t Decision tree model $G_t(x)$ with the sample set D_t. When training the nodes of the Decision tree model, select a part of the sample features among all the sample features on the nodes, and choose an optimal feature among these randomly selected part of the sample features to do the left and right subtree partitioning of the Decision tree.

2) In case of classification algorithm prediction, the category or one of the categories with the most votes cast by T weak learners is the final category. In case of regression algorithms, the value obtained by arithmetic averaging of the regression results obtained by T weak learners is the final model output.

3.3 Gradient-enhanced regression tree

Gradient-enhanced regression tree(GBR) is a nonparametric machine learning method based on propulsion strategies and Decision trees (35), whose basic idea is to use regression trees as weak learners and replace a single strong learner with a superposition of multiple weak learners. We train multiple layers of weak classifiers for the same training set, and each layer uses the training set to train a weak classification model, from which we obtain the prediction results. We then determine the weights that should be reassigned to each sample based on whether the samples in the training set are correctly classified and the accuracy of the overall classification, and train a classifier for the next layer with the new data set after the modified weights. This training is continued until there are few misclassified samples, and finally the classifiers of each layer with weight assignments are fused together so that the final decision classifier is composed down.

3.4 Support vector machine

Support vector machine (SVM) is a supervised machine learning that can deal with classification and regression problems (36), the basic idea is to find the optimal classification hyperplane that completely separates the two classes of samples in the original space in the linearly divisible case and to use kernel methods in the nonlinear case to solve problems that are nonlinear in low-dimensional space as linearly integrable problems in high-dimensional space (37). Delineating the hyperplane can be defined as a linear equation:

$$w^T x + b = 0$$

Where w={ $w_1,\cdots,w_d$ } is a normal vector that determines the direction of the hyperplane, d is the number of eigenvalues, x is the sample to be trained, and b is the displacement term that determines the distance between the hyperplane and the origin.

Suppose $P(x_1,\cdots,x_n)$ is a point in the training sample, where x_i denotes the ith feature variable of that sample. Then the formula for the distance from the point to the hyperplane is:

$$d = \frac{|w_1 * x_1 + w_2 * x_2 + \cdots + w_n * x_n + b|}{\sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}} = \frac{|W^T * X + b|}{\| W \|}$$

Where ‖W‖ is the parametrization of the hyperplane and the constant b is the intercept in the linear equation.

In the case that the hyperplane is determined, the full support vector can be found and then the hyperplane interval can be calculated. The next step is to determine w and b so that the interval is maximum. This is an optimization problem whose objective function can be written as:

$$\arg\max\left\{\min\left(y\left(w^T + b\right)\right) * \frac{1}{\| W \|}\right\}$$

Where y denotes the label of the training sample point and its value is -1 or 1, and $y(w^T+b)$ denotes the distance. If the training sample points are in the positive direction of the hyperplane, then $y(w^T+b)$ is a positive number, and the opposite is a negative number. This is an optimization problem with constraints and can usually be solved by the Lagrange multiplier method.

$$L(w, b, a) = \frac{1}{2} * \| w \| - \sum_{i=1}^{n} a_i (y_i (w * x + b) - 1)$$

This optimization algorithm gives us a^*, and then we can solve for w and b according to a^*. The purpose of the classification is to find the hyperplane, i.e., the “decision plane”.

3.5 Semi-automatic parameter adjustment

Semi-automatic parameter adjustment is a parameter adjustment method combining manual parameter adjustment and grid search. For different algorithms, it has different sequence of parameter adjustment steps. Taking XGBoost parameter adjustment as an example, the process is as follows:

1) First grid search n_ estimators parameter, other parameters take fixed values;

2) Take the optimization result in (1) and add it to the parameter setting, and grid search min_child_ weight and max_ depth two parameters;

3) Take the optimization result in (1)(2) and add it to the parameter setting, and grid search gamma parameter;

4) Take the optimization result in (1)~(3) and add it to the parameter setting, and grid search subsample and colsample_bytree two parameters;

5) Take the optimization result in (1)~(4) and add it to the parameter setting, and grid search reg_alpha and reg_lambda two parameters;

6) Take the optimization result in (1)~(5) and add it to the parameter setting, and grid search learning_rate parameter.

4. Experiment

4.1 Data import

The data set collected in this paper contains biological activity values IC50 and pIC50 of compounds ERα, information on 729 molecular descriptors, interpretation of molecular descriptor meanings.

4.2 Data pre-processing

In this data set, IC50 is the biological activity value of the compound against ERα, which is an experimental measurement, where a smaller value represents greater biological activity and more effective in inhibiting ERα activity. The pIC50 is obtained by converting the IC50 value (i.e., the negative logarithm of the IC50 value), which usually has a positive correlation with biological activity, i.e., a higher pIC50 value indicates higher biological activity. In practical QSAR modeling, pIC50 is generally used to represent the bioactivity value. Variable selection is first performed for 729 molecular descriptors of 1974 compounds, and the top 20 molecular descriptors (i.e., variables) with the most significant effect on biological activity are selected by using a XGBoost-based data feature selection method to rank the variables according to their importance on biological activity.

4.3 XGBoost-based data feature selection

The data of 729 molecular descriptors of 1974 compounds are initially analyzed. We find that the data are of a certain scale and the influence factors obtained by adopting simple correlation analysis are not representative, so we adopt a more rigorous XGBoost-based data feature selection to calculate all 729 feature weights, as shown in Figure 1.

As seen in Figure 1, the feature weights of the 729 molecular descriptors varies greatly overall, with the maximum weight exceeding 0.12 and the minimum weight close to 0. It is thus clear that the degree of influence of different molecular descriptors on biological activity varies greatly. Therefore, all the calculated molecular descriptor feature weights are output in descending order, and the top 20 molecular descriptors are intercepted as the most significant variables affecting biological activity, as shown in Figure 2.

The top 20 feature weights of 729 molecular descriptors of 1974 compounds are summarized according to Figure 2, which are ranked and summarized to finally obtain the top 20 molecular descriptors affecting biological activity, as shown in Table 1.

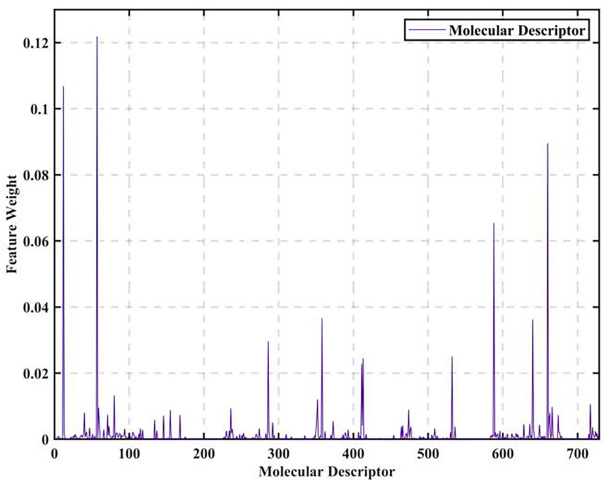


FIGURE 1
Molecular descriptor feature weights.

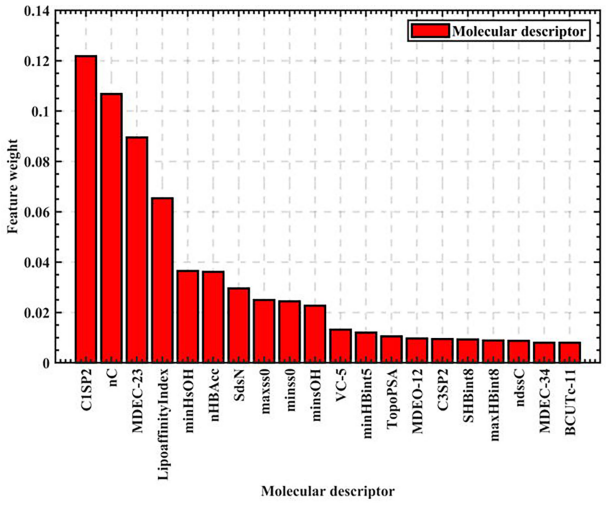


FIGURE 2
Top 20 ranking chart of feature weights.

4.4 Quantitative structure-activity relationship model

Based on the 20 molecular descriptors screened previously, we build the model according to the known data types by the four algorithms that have been selected. A total of 1974 sets of data exist in the data set, so we randomly select 50 compounds as the test set for IC50 values and corresponding pIC50 values prediction, and the remaining 1924 sets of data as the prediction set.

Firstly, we eliminate the selected 50 compounds, and the remaining 1924 sets of compound data with the filtered 20 molecular descriptors as input and pIC50 values as output, use the cross validation method to segment the data with 0.35 as the sample ratio, so as to obtain the training set and test set, and then use the training set to train the Gradient-enhanced regression, XGBoost regression, Support vector machine and Random Forest regression models respectively, and the training process is shown in Figure 3.

As seen in Figure 3, the training results of Support vector machine are the worst, and the training results of Gradient-enhanced regression, XGBoost regression, and Random Forest regression models are better and do not differ much from each other. Therefore, we perform Semi-automatic parameter adjustment combining manual parameter adjustment and grid parameter adjustment on the three algorithm models of Gradient-enhanced regression, XGBoost regression and Random Forest regression to find the core algorithm.

After that, we optimize the parameters of the three algorithms, and consider obtaining the model that is closest to the actual accuracy through the optimal parameters. Among them, in the Random Forest algorithm model, we optimize the number of trees, the maximum depth of trees, the maximum number of features, and the minimum number of samples

TABLE 1 Range of molecular descriptor values found by the optimization search model.

Molecular descriptors	Weighting value	Molecular descriptors	Weighting value
C1SP2	0.121828	VC-5	0.013187
nC	0.106756	minHBint5	0.011962
MDEC-23	0.089470	TopoPSA	0.010527
LipoaffinityIndex	0.065367	MDEO-12	0.009671
minHsOH	0.036505	C3SP2	0.009461
nHBAcc	0.036146	SHBint8	0.009285
SdsN	0.029560	maxHBint8	0.008846
maxss0	0.024947	ndssC	0.008720
minss0	0.024412	MDEC-34	0.007981
minsOH	0.022663	BCUTc-11	0.007961

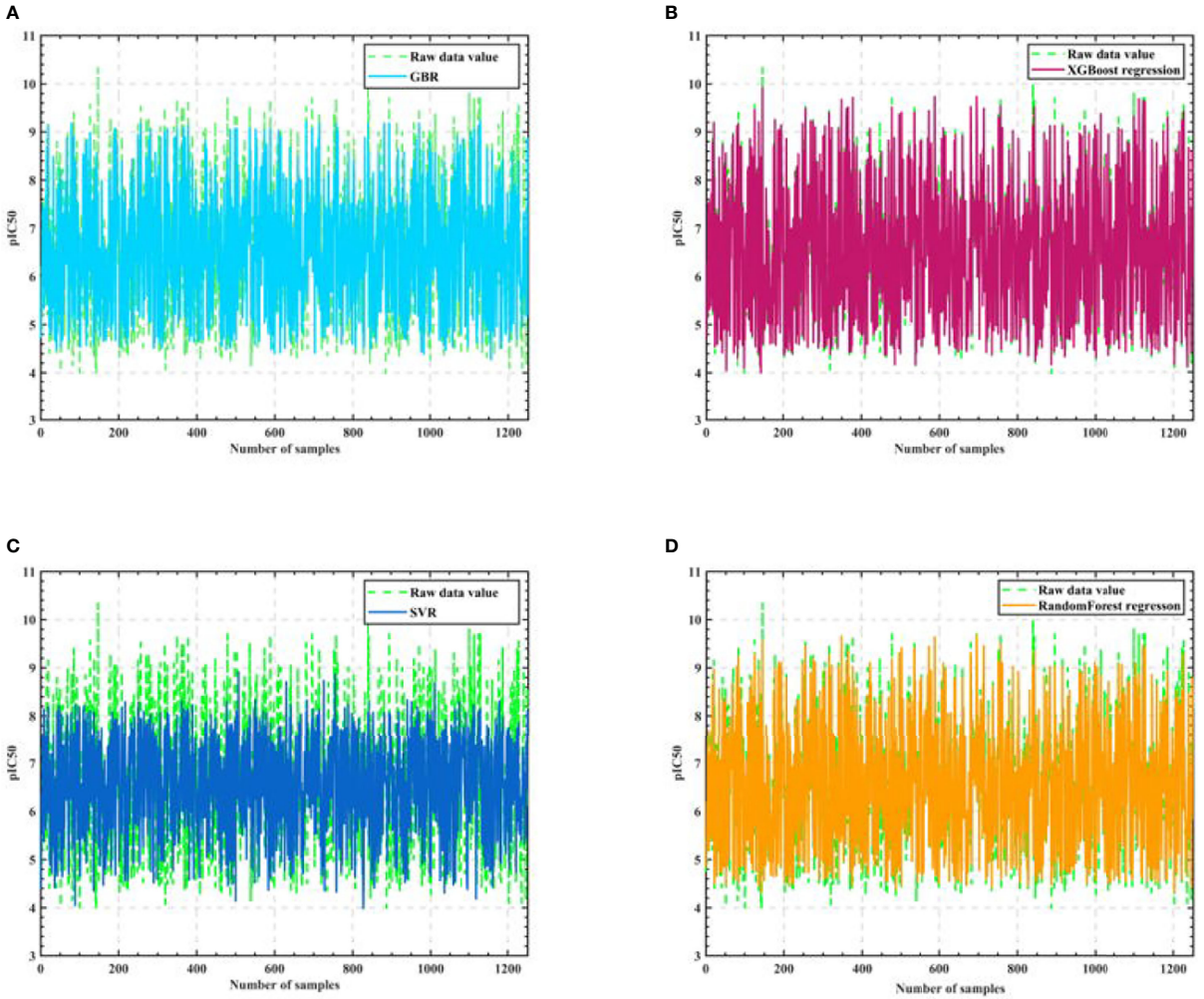


FIGURE 3
Algorithm model training process. **(A)** GBR **(B)** XGBoost regression **(C)** SVR **(D)** Random Forest regression.

10.3389/fonc.2022.956705

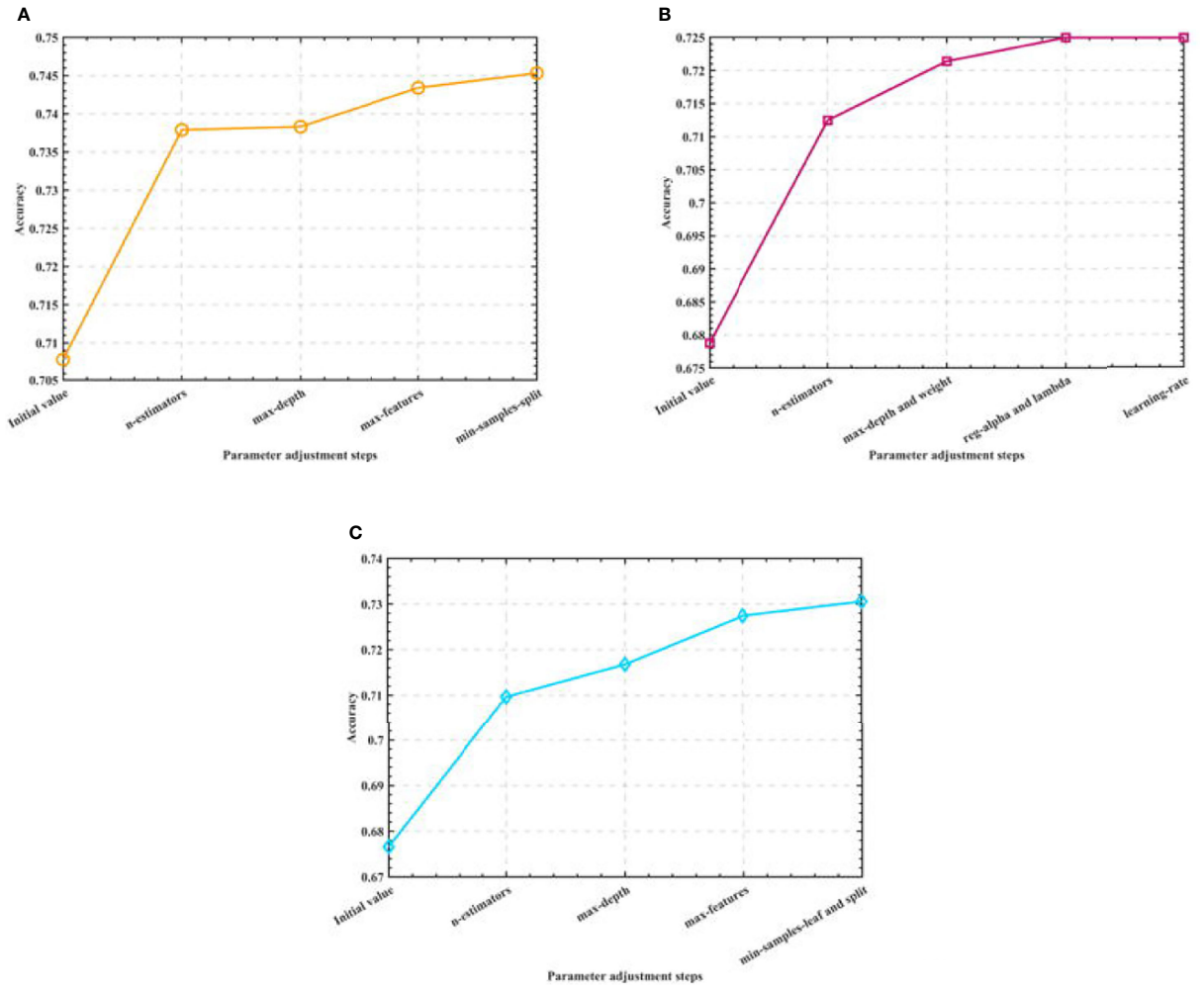


FIGURE 4
Semi-automatic parameter adjustment process. **(A)** Optimization process of Random Forest **(B)** Optimization process of XGBoost **(C)** Optimization process of GBR.

allowed to split nodes. In the XGBoost algorithm model, we optimize the number of learners, the depth of the tree, the minimum weight of the subset, L1 regularization, L2 regularization and the learning rate. In the Gradient-enhanced algorithm model, we optimize the parameters of the maximum number of weak learners, the maximum depth of learners, the maximum number of features of learners, the minimum number of samples required by leaf nodes and the minimum number of samples divided into internal nodes. The process is shown in Figure 4.

By adjusting the parameters of the three algorithm models, we finally determine the optimal parameter combination of the Random Forest algorithm as: n_estimators=500, max_depth=90, max_features=0.1, min_samples_split=2; the optimal parameter combination of the XGBoost algorithm as: n_estimators=20, max_depth=7, min_child_weight=5, reg_alpha=1, reg_lambda=0.1, learning_rate=0.3; the optimal parameter combination of the Gradient-enhanced algorithm as: n_estimators=300, max_depth=4, max_features=0.3, min_samples_leaf=8,min_samples_split=9; and use the test set to test the algorithms before and after parameter adjustment. The test accuracy of the three algorithms is shown in Table 2.

In order to determine the core algorithm more accurately, we use the training data set and the test data set to train the above three parameter adjusted models, combined with a variety of regression model training error analysis methods to determine the algorithm with the best training effect. Among them, the regression model training error analysis methods we selected Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), absolute coefficient (R-Square), and Explained Variance score (EV).

The model evaluation error analysis table, model testing error analysis table, and cross-validation results are shown in Tables 3, 4.

After adjusting the error analysis of the model, it can be concluded that the Random Forest has the best accuracy and excellent anti over fitting, and the stability of the algorithm is high.

Then, in order to verify that Random Forest is the best algorithm, we use three training and parameter adjusted algorithms: Random Forest, Gradient-enhanced and XGBoost to predict the IC50 value and the corresponding pIC50 value of the selected 50 groups of compound data, and the experimental results are shown in Figure 5.

We experimentally compare the three regression algorithm models after tuning the parameters on 50 sets of test set data as in Figure 6.

We finally give the accuracy rates, as in Table 5.

As can be seen from the table, Random Forest regression has the highest accuracy rate of 76.685%, so it can be considered reasonable for the Random Forest algorithm to be the core algorithm.

5 Conclusion

With the development of computer technology, in the early stage of anti-breast cancer drug research and development, using computer models to predict the biological activity of compounds is conducive to reducing the failure rate of drug research and development and saving a lot of research and development time and cost for research and development institutions. Therefore, in order to solve the problem of bioactivity prediction during the early development of anti-breast cancer candidate drugs, this paper has carried out relevant work and obtained the following conclusions.

(1) In this paper, we investigate compounds capable of antagonizing ERα activity by facilitating XGBoost-based data feature selection thereby screening the top 20 molecular descriptors with the most significant impact on biological activity.

(2) Then, from the perspective of molecular descriptors, with 20 molecular descriptors selected based on XGBoost feature as input and pIC50 value as output, multiple regression prediction models of Random Forest, Gradient-enhanced, XGBoost and SVM are constructed to predict ER biological activity. According to the degree of fitting between the predicted value and the real value, the Random Forest, Gradient-enhanced and XGBoost are preliminarily selected with good results. In order to select the best algorithm from the three algorithms, the Semi-automatic parameter adjustment method is used to adjust the parameters of the three algorithms. The Random Forest has the highest accuracy, the best accuracy and excellent anti over fitting, and the algorithm has high stability.

(3) Finally, by training the initial randomly selected data, it is verified that the Random Forest with Semi-automatic parameter adjustment has the best effect. It can be seen that using the Semi-automatic parameter adjusted Random Forest model to predict the bioactivity of compounds against breast cancer drugs can provide a good reference, and can play a certain role in promoting the optimization of drug properties in the process of cancer drug development.

TABLE 2 Comparison of accuracy before and after parameter adjustment.

Algorithm	Random Forest regression	XGBoost regression	GBR
Accuracy before parameter adjustment	0.707829	0.678772	0.676599
Accuracy after parameter adjustment	0.745329	0.724967	0.730603

TABLE 3 Model evaluation error analysis.

Evaluation Metrics	MSE	MAE	EV	R^2
Random Forest	0.077646	0.203693	0.961473	0.961529
GBR	0.135665	0.268423	0.932684	0.932684
XGBoost	0.152522	0.290378	0.924355	0.924319

TABLE 4 Model test error analysis.

Evaluation Metrics	MAE	RMSE	MAPE	EV
Random Forest	0.544467	0.135665	0.089044	0.715558
GBR	0.557161	0.135665	0.090808	0.701300
XGBoost	0.567658	0.135665	0.091912	0.687493

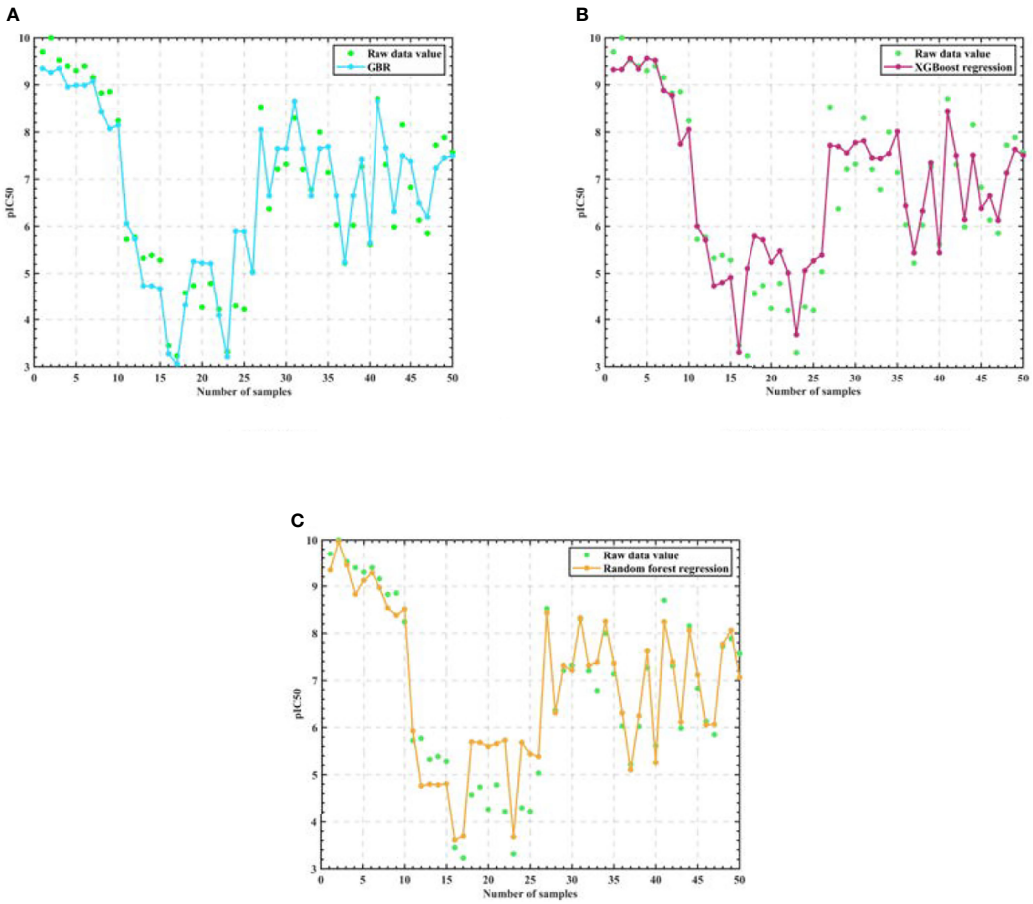


FIGURE 5
Algorithm model regression test results. **(A)** GBR **(B)** XGBoost regression **(C)** Random Forest regression

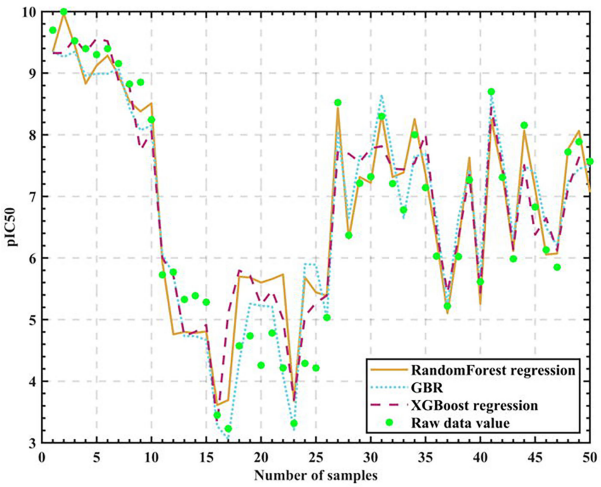


FIGURE 6
Algorithm model test results.

TABLE 5 Final prediction accuracy.

Algorithm	Random Forest regression	GBR	XGBoost regression
Accuracy	0.766850	0.758679	0.756774

The model proposed in this paper can provide better support for the early development of anti- breast cancer drugs, and the model can also be extended to other areas of prediction. However, the test accuracy of the model did not reach a particularly high level, so the subsequent optimization of drug properties for the problem will provide more solutions using artificial intelligence technology.

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

The first author did the experiment of this paper, and analyzed and compared the experimental results. The second author wrote this article, the third author edited the language of this article, and the fourth author and corresponding author put forward the idea of this article. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by: North China University of Science and Technology, Project Name : Research on Trusted Verification Technology of Cloud Outsourcing Computing, Project Number:0088/28415599.

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

References

1. Zhang Y, Li H, Zhang J, Zhao C, Lu S, Qiao J, et al. The combinatory effects of natural products and chemotherapy drugs and their mechanisms in breast cancer treatment. *Phytochem Rev* (2020) 19(5):1179–97. doi: 10.1007/s11101-019-09628-w

2. Ferlay J, Colombet M, Soerjomataram I, Parkin D, Pineros M, Znaor A, et al. Cancer statistics for the year 2020: An overview. *Int J Cancer* (2021) 149(4):778–89. doi: 10.1002/ijc.33588

3. Jain V, Kumar H, Anod HV, Chand P, Gupta NV, Dey S, et al. A review of nanotechnology-based approaches for breast cancer and triple-negative breast cancer. *J Controlled Release* (2020) 326:628–47. doi: 10.1016/j.jconrel.2020.07.003

4. Cui C, Ding X, Wang D, Chen L, Xiao F, Xu T, et al. Drug repurposing against breast cancer by integrating drug-exposure expression profiles and drug–drug links based on graph neural network. *Bioinformatics* (2021) 37(18):2930–7. doi: 10.1093/bioinformatics/btab191

5. Aggarwal S, Verma SS, Aggarwal S, Gupta SC. Drug repurposing for breast cancer therapy: Old weapon for new battle. *Semin Cancer Biol Acad Press* (2021) 68:8–20. doi: 10.1016/j.semcancer.2019.09.012

6. Chan H, Shan H, Dahoun T, Dahoun T, Vogel H, Yuan S. Advancing drug discovery *via* artificial intelligence. *Trends Pharmacol Sci* (2019) 40(8):592–604. doi: 10.1016/j.tips.2019.06.004

7. Zhao B. Anti-breast cancer drug screening based on neural networks and QSAR model. *Med Rep Case Stud* (2022) 7(1):1–5.

8. Leya S, Kumar PN. *Virtual screening of anticancer drugs using deep learning*. Springer International Publishing (2020) p. 1293–8. doi: 10.1007/978-3-030-41862-5_131

9. Xu B. Optimal modeling of anti-breast cancer candidate drugs based on SSA-BP. *Helsinki Finland* (2021) 15. doi: 10.25236/ICBCME.2021.032

10. Liu X, Zhang W, Zheng W, Jiang X. Micropatterned coculture platform for screening nerve-related anticancer drugs. *ACS nano* (2021) 15(1):637–49. doi: 10.1021/ACSNANO.0C06416

11. Zhu Y, Brettin T, Evrard YA, Partin A, Xia F, Shukla M. Ensemble transfer learning for the prediction of anti-cancer drug response. *Sci Rep* (2020) 10 (1):18040–0. doi: 10.1038/s41598-020-74921-0

12. Xiong Z, Wang D, Liu X, Zhong F, Wan X, Li X. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J med Chem* (2019) 63(16):8749–60. doi: 10.1021/acs.jmedchem.9b00959

13. Wongyikul P, Thongyot N, Tantrakoolcharoen P, Seephueng P, Khumrin P. High alert drugs screening using gradient boosting classifier. *Sci Rep* (2021) 11 (1):20132–2. doi: 10.1038/S41598-021-99505-4

14. FernándezLlaneza D, Ulander S, Gogishvili D, Nittinger E, Zhao H, Tyrchan C. Siamese Recurrent neural network with a self-attention mechanism for bioactivity prediction. *ACS omega* (2021) 6(16):11086–94. doi: 10.1021/ACSOMEGA.1C01266

15. Kumari M, Subbarao N. Deep learning model for virtual screening of novel 3C-like protease enzyme inhibitors against SARS coronavirus diseases. *Comput Biol Med* (2021) 132:104317. doi: 10.1016/J.COMPBIOMED.2021.104317

16. Abdo A, Pupin M. Turbo prediction: a new approach for bioactivity prediction. *J computer-aided Mol design* (2022) 36(1):1–9. doi: 10.1007/S10822-021-00440-3

17. Gupta A, Zhou H. A machine learning-enabled pipeline for Large-scale virtual drug screening. *J Chem Inf Modeling* (2021) 61(9):4236–44. doi: 10.1021/ACS.JCIM.1C00710

18. Carpenter K, Pilozzi A, Huang X. A pilot study of multi-input recurrent neural networks for drug-kinase binding prediction. *Molecules* (2020) 25(15):3372. doi: 10.3390/molecules25153372

19. Xin W, Sun S, Wu D, Zhou L. Personalized online learning resource recommendation based on artificial intelligence and educational psychology. *Front Psychol* (2021) 12:767837. doi: 10.3389/FPSYG.2021.767837

20. Feng Z, Huang J, Jin S, Wang G, Chen Y. Artificial intelligence-based multi-objective optimisation for proton exchange membrane fuel cell: A literature review. *J Power Sources* (2022) 520. doi: 10.1016/J.JPOWSOUR.2021.230808

21. Dulebenets M, Pasha J, Kavoosi M, Abioye FO, Ozguven EE, Moses R. Multiobjective optimization model for emergency evacuation planning in geographical locations with vulnerable population groups. *J Manage Eng* (2019) 36(2):1–17. doi: 10.1061/(ASCE)ME.1943-5479.0000730

22. Boukerche A, Tao Y, Sun P. Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems. *Comput Networks* (2020) 182. doi: 10.1016/j.comnet.2020.107484

23. Mishra S, Abbas M, Jindal K, Narayan J, Dwivedy SK. *Artificial intelligence-based technological advancements in clinical healthcare applications: A systematic review*. Springer Singapore (2022) p. 207–27. doi: 10.1007/978-981-16-9455-4_11

24. Zhang H, Xu J. Design and implementation of information application platform for primary medical institutions based on artificial intelligence. *Int Conf Comput Graphics Artif Intell Data Process* (2022) 12168:1216826–1216826-11. doi: 10.1117/12.2631112

25. Guo S. *Application of artificial intelligence technology in international trade finance*. Springer International Publishing (2020) p. 155–62. doi: 10.1007/978-3-030-62743-0_22

26. Li C. *The application of artificial intelligence and machine learning in financial stability*. Springer International Publishing (2020) p. 214–9. doi: 10.1007/978-3-030-62743-0_30

27. Pan Y, Zhou P, Yan Y, Anupam A, Wang Y, Guo D. New insights into the methods for predicting ground surface roughness in the age of digitalisation. *Precis Eng* (2021) 67:393–418. doi: 10.1016/j.precisioneng.2020.11.001

28. Rani P, Kavita, Verma S, Kaur N, Wozniak M, Shafi J. Robust and secure data transmission using artificial intelligence techniques in ad-hoc networks. *Sensors* (2021) 22(1):251–1. doi: 10.3390/S22010251

29. Dulebenets M. An adaptive polyploid memetic algorithm for scheduling trucks at a cross-docking terminal. *Inf Sci* (2021) 565:390–421. doi: 10.1016/J.INS.2021.02.039

30. Dong W, Wozniak M, Wu J, Li W, Bai Z. *De-noising aggregation of graph neural networks by using principal component analysis*. IEEE Transactions on Industrial Informatics (2022). doi: 10.1109/TII.2022.3156658

31. Zhou T, Lu H, Wang W, Yong X. GA-SVM based feature selection and parameter optimization in hospitalization expense modeling. *Appl Soft Computing* (2018) 75:323–32. doi: 10.1016/j.asoc.2018.11.001

32. Yang Y, Zhang X, Yang L. Data-driven power system small-signal stability assessment and correction control model based on XGBoost. *Energy Rep* (2022) 8 (S5):710–7. doi: 10.1016/J.EGYR.2022.02.249

33. Sheng X, Huo W, Zhang C, Zhang X, Han Y. A paper quality and comment consistency detection model based on feature dimensionality reduction. *Alexandria Eng J* (2022) 61(12):10395–405. doi: 10.1016/J.AEJ.2022.03.074

34. Verdaasdonk MJA, Carvalho RMDE. From predictions to recommendations: Tackling bottlenecks and overstaying in the emergency room through a sequence of random forests. *Healthcare Analytics* (2022) 2(2). doi: 10.1016/J.HEALTH.2022.100040

35. Liao S, Liu Z, Liu B, Cheng C, Jin X, Zhao Z. Multistep-ahead daily inflow forecasting using the ERA-interim reanalysis data set based on gradient-boosting regression trees. *Hydrol Earth System Sci* (2020) 24(5):2343–63. doi: 10.5194/hess-24-2343-2020

36. Pannakkong W, Harncharnchai T, Buddhakulsomsiri J. Forecasting daily electricity consumption in Thailand using regression, artificial neural network, support vector machine, and hybrid models. *Energies* (2022) 15(9):3105–5. doi: 10.3390/EN15093105

37. Cai W, Qu Z. HRM risk early warning based on a hybrid solution of decision tree and support vector machine. *Wirel Commun Mobile Computing* (2022) 2022. doi: 10.1155/2022/8396348

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 01 August 2022
DOI 10.3389/fonc.2022.919088

OPEN ACCESS

EDITED BY
Shahid Mumtaz,
Instituto de Telecomunicações,
Portugal

REVIEWED BY
Zahoor Ahmad,
Counties Manukau District Health
Board, New Zealand
Tae Keun Yoo,
B&VIIT Eye center/Refractive surgery
&AI Center, South Korea
Johannes Zenk,
Augsburg University Hospital, Germany

*CORRESPONDENCE
Zitong Lin
linzitong_710@163.com
Ying Chen
yingchen@nju.edu.cn
Xudong Yang
yangxd66@163.com

†These authors share first authorship

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

RECEIVED 13 April 2022
ACCEPTED 28 June 2022
PUBLISHED 01 August 2022

CITATION
Hu Z, Wang B, Pan X, Cao D, Gao A,
Yang X, Chen Y and Lin Z (2022) Using
deep learning to distinguish malignant
from benign parotid tumors on plain
computed tomography images.
Front. Oncol. 12:919088.
doi: 10.3389/fonc.2022.919088

Using deep learning to distinguish malignant from benign parotid tumors on plain computed tomography images

Ziyang Hu[1†], Baixin Wang[2†], Xiao Pan[1], Dantong Cao[1], Antian Gao[1], Xudong Yang[3*], Ying Chen[2*] and Zitong Lin[1*]

[1]Department of Dentomaxillofacial Radiology, Nanjing Stomatological Hospital, Medical School of Nanjing University, Nanjing, China, [2]School of Electronic Science and Engineering, Nanjing University, Nanjing, China, [3]Department of Oral and Maxillofacial Surgery, Nanjing Stomatological Hospital, Medical School of Nanjing University, Nanjing, China

Objectives: Evaluating the diagnostic efficiency of deep-learning models to distinguish malignant from benign parotid tumors on plain computed tomography (CT) images.

Materials and methods: The CT images of 283 patients with parotid tumors were enrolled and analyzed retrospectively. Of them, 150 were benign and 133 were malignant according to pathology results. A total of 917 regions of interest of parotid tumors were cropped (456 benign and 461 malignant). Three deep-learning networks (ResNet50, VGG16_bn, and DenseNet169) were used for diagnosis (approximately 3:1 for training and testing). The diagnostic efficiencies (accuracy, sensitivity, specificity, and area under the curve [AUC]) of three networks were calculated and compared based on the 917 images. To simulate the process of human diagnosis, a voting model was developed at the end of the networks and the 283 tumors were classified as benign or malignant. Meanwhile, 917 tumor images were classified by two radiologists (A and B) and original CT images were classified by radiologist B. The diagnostic efficiencies of the three deep-learning network models (after voting) and the two radiologists were calculated.

Results: For the 917 CT images, ResNet50 presented high accuracy and sensitivity for diagnosing malignant parotid tumors; the accuracy, sensitivity, specificity, and AUC were 90.8%, 91.3%, 90.4%, and 0.96, respectively. For the 283 tumors, the accuracy, sensitivity, and specificity of ResNet50 (after voting) were 92.3%, 93.5% and 91.2%, respectively.

Conclusion: ResNet50 presented high sensitivity in distinguishing malignant from benign parotid tumors on plain CT images; this made it a promising auxiliary diagnostic method to screen malignant parotid tumors.

KEYWORDS

deep learning, convolutional neural network, residual neural network, parotid tumor, computed tomography

Introduction

Parotid tumor is the most common type of salivary gland tumor. The acinar, ductal, and myoepithelial cells that comprise parotid tissues can give rise to a variety of benign and malignant neoplasms. Pre-operative recognition of malignancy in parotid tumors is useful in that it may alert the surgeon to more stringent attention to the operative margin and hence better tumor clearance (1).

From the clinical aspect, although there are some clues of malignancy-rapid growth, skin fixation, ulceration, facial nerve palsy, pain, or cervical node metastasis, but only 30% malignant parotid tumors present with these features (1, 2). Fine-needle biopsy is helpful in differentiating malignancy; however, it is invasive and more dependent on technical skill and experience to obtain adequate specimens, and the few tissues obtained always could not represent the whole tumor (3–5). Computed tomography (CT), as a commonly used imaging technique, is useful to identify the location and size of parotid tumors. However, benign and malignant parotid tumors always have similar CT features; the sensitivity of CT in identifying malignant tumors is unsatisfactory (6–8).

Recently, deep learning methods, especially convolutional neural network (CNN), have demonstrated effectiveness in image recognition tasks. CNN-based tumor segmentation and classification have been widely used in breast cancer (9), lung cancer (10), liver tumor (11, 12), and nasopharyngeal carcinoma (13). For parotid tumors, Xia et al. and Chang et al. had utilized neural network to differentiating benign and malignant parotid tumors on magnetic resonance imaging (MRI) (14, 15). To date, there were no CNN models based on plain CT images to differentiate benign and malignant parotid tumors. Because of the fatty nature of parotid gland (16), the plain CT images usually could visualize the tumors in parotid gland well and provides abundant texture information of parotid tumors (17). So, in this study, we explored using CNN to diagnose parotid tumors on plain CT images.

Residual neural network (ResNet) is a CNN network proposed in 2015. The framework reformulates the layers as learning residual functions with reference to the layer inputs to obtain deeper networks with higher accuracy (18). The ResNet model could employ the entire image and is capable of retaining image information more completely than many CNN networks. It exhibits high diagnostic efficiency for liver fibrosis staging and lung nodule segmentation (19–22). In this study, the applicability of using ResNet to classify benign and malignant parotid tumor on plain CT images was investigated, and the diagnostic efficiency of it was compared with other two networks and oral radiologists.

Abbreviations: CT, Computed tomography, CNN, convolutional neural network, ResNet, residual neural network, SGD, random gradient descent, CAM, class activation map, PPV, positive predictive value, NPV, negative predictive value, ROC, receiver operating characteristic, AUC, area under the curve.

Methods and materials

Data acquisition

An oral radiologist collected the CT images of patients with parotid tumors in our hospital from 2008 to 2020. The inclusion criteria were as follows: (1) primary parotid tumor; (2) definite pathological diagnosis was available after surgery. (3) The CT images were of good quality, without motion artifacts and foreign body artifacts. The approval from the Ethics Committee of our University was obtained prior to performing this study (NJSH-2022NL-069).

The plain CT images of 283 patients (113 males and 170 females; mean age, 50.5 ± 15.6 years; range, 18–73 years) with parotid tumors were included. Of them, 150 were benign (55 males and 95 females; mean age, 51.7 ± 16.3 years) and 133 were malignant (58 males and 75 females; mean age, 50.3 ± 15.2 years). No statistical difference of age and gender was found between benign and malignant tumor group. The pathological classification of the 283 tumors was showed in Table 1.

All patients were performed CT examination before surgery; the parameters of CT were as follows: tube potential: 130 kVp, tube current 56 mA, slice thickness: 3 mm, matrix: 512 × 512, window width: 200 Hounsfield units (Hu), window level: 40 Hu.

Image processing

Two radiologists manually selected the region of interest; axial CT images with lesions were randomly selected and then the regions of interest were obtained by square cropping the CT images (Figure 1). For each patient, three or five axial CT images including tumors were selected and cropped. It was confirmed by another radiologist, and if there was any doubt about the area of

TABLE 1 The pathological classification of the 283 tumors included.

Benign		
	Pleomorphic adenoma	76
	Warthin tumor	46
	Basal cell adenoma	20
	Other	8
Malignant		
	Adenocarcinoma	
	Mucoepidermoid carcinoma	32
	Pleomorphic adenocarcinoma	19
	Acinar cell carcinoma	15
	Adenoid cystic carcinoma	11
	Other	13
	Lymphoma	15
	Squamous cell carcinoma	11
	Other	17

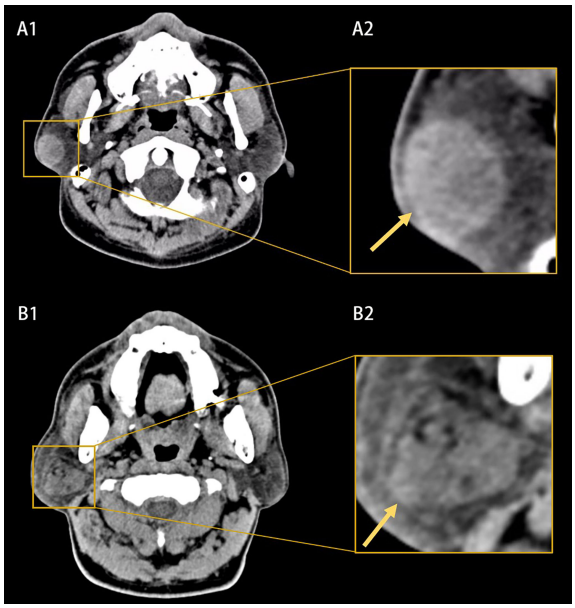


FIGURE 1
Example of computed tomography (CT) images **(A1, B1)** and the extracted region of interests **(A2, B2)**. **(A1, A2)** Showed a benign parotid tumor with a well-defined and smooth border and homogeneous appearance (yellow arrow). **(B1, B2)** Showed a malignant parotid tumor with a poor-defined border and heterogeneous appearance (yellow arrow).

interest, the two radiologists would work together to re-crop the CT image. Neither of them knew the patients' pathological diagnosis. As the resolution of images cropped was not equal, the resolution of image was adjusted to a uniform size of 317 × 317 pixels.

A total of 917-cropped CT images were finally obtained (Figure 2). The subjects in dataset were divided into two subcategories: training and testing. The training set (approximately 75% of the database [687 images for 213 patients] was used to train variant versions of the model with different initialization conditions and hyper parameters. Once the models have been trained, their performance was evaluated using test set (approximately 25% of the database [230 images for 70 patients]). When building the CNN model, a series of methods were performed on the input images in order to reduce over-fitting of the model. These data argument methods included random horizontal and vertical flipping, random image rotation within 90°.

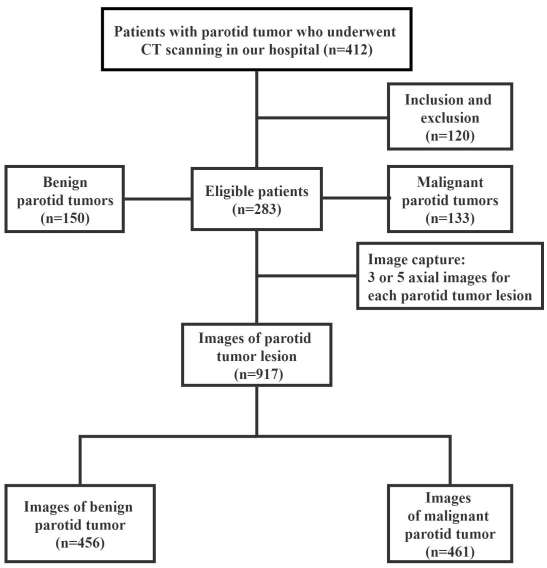


FIGURE 2
A flowchart of plain computed tomography (CT) image inclusion and exclusion.

Network structure and voting

The CNN models were implemented on hardware with following specification: Intel processor i7, 64 GB RAM with NVIDIA Tesla V100 GPU, 1 TB hard disk for implementing.

ResNet 50-layer structure was shown in Figure 3 with pre-trained model on the ImageNet database. The input data were grayscale image with a resolution of 317 × 317. The input data were gradually processed by ResNet50 through five blocks. In the first block, the image was converted into a 159 × 159 × 64 tensor. Between the second and fifth stages, a residual block structure

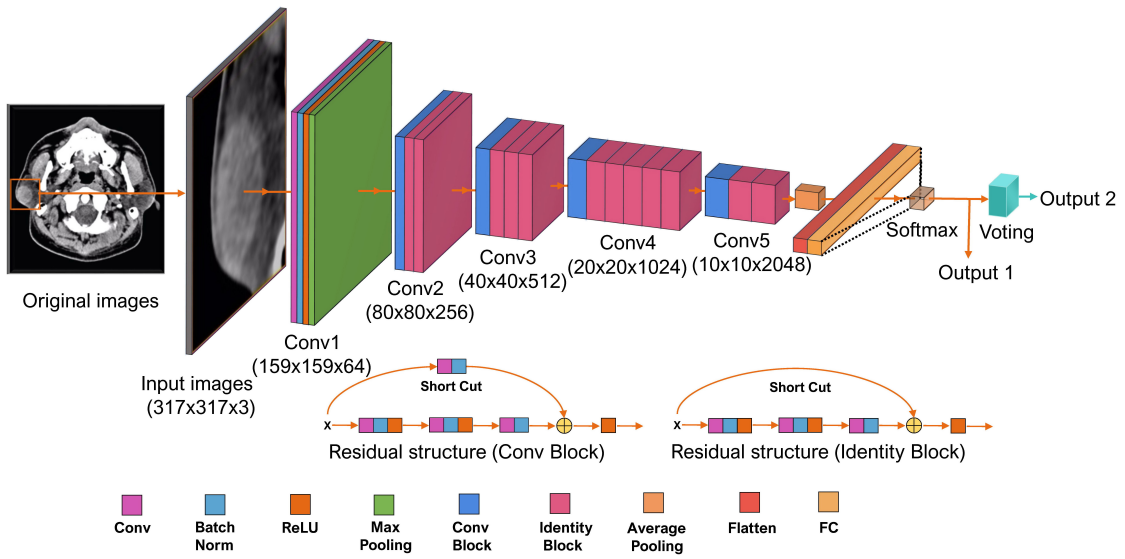


FIGURE 3
The structure of Residual neural network (ResNet) 50-layer model. The input data are the cropped plain computed tomography (CT) images with a resolution of 317 × 317. It propagated by ResNet through five convolutional phases. Through the five convolutional phases, the data were then processed with three fully connected layers. This ResNet model was structured to output two values; the bigger one indicating the classification of label 0 or label 1 for each CT images is output1. The results of output1 of each CT images are used to perform the final voting. The most labels of output1 for a parotid lesion are output2. Batch Norm is the batch normalization; Conv is the Convolutional layer; FC is the fully connected layer; ReLU is the rectified linear unit function.

was introduced to overcome the problems of vanishing and exploding gradients. After five blocks, the input was converted into a 10 × 10 × 2,048 tensor. Both the height and width were greatly reduced, and the number of dimensions was increased from three to 2,048, which indicated that the extracted information was much more than the original RGB pixel information. According to the 2,048 features extracted by ResNet50, the tensor was fattened into 2,048 vector elements. The model loss function was the cross-entropy loss function and the Random gradient descent (SGD) model optimization method was used. The initial learning rate was 5e−3. The batch size of the model training was 16; the final model selected for the test group was the model with the smallest loss function value for the test group. Fivefold cross-validation was used to establish the ResNet model. The proportion of patients corresponding to benign and malignant parotid tumor was equal for the training and test groups. The final result was the average of the fivefold cross-validation for the test group.

In order to simulate the process of human diagnosis and take the spatial information into account, a voting model was added at the end of Resnet. The input was classified as 0 or 1 (0 and 1 represent benign and malignant parotid tumor, respectively). For each parotid lesion (three or five CT images), the most classification was counted as the final result of the parotid lesion. Generated activation maps by class activation map (CAM) on test dataset were applied to evaluate the region of interest for further clinical review.

Other two neural networks (VGG16_bn and DenseNet169) were used to classify the benign and malignant parotid tumors; the voting model was also added at the end of these networks and the diagnostic efficiency of them was compared with ResNet50. These two networks were also models pre-trained using the ImageNet database (23–25).

Manual classification of parotid tumor on plain computed tomography images

After development of the CNN models were complete, the 283 parotid tumors were classified into benign or malignant ones by two radiologists (A with 3 and B with 12 years of experience, respectively), using the same CT images. These two observers did not take part in the model training process and were blinded to lesion selection. The observers were also unaware of patient names, laboratory results, other imaging findings, or final diagnosis. After 3 months, observer B re-classified the 283 tumors into malignant or benign; this time, all the original CT images without cropping of the 283 patients were used. The following characteristics were used for classification: tumor location, number of tumors (single or multiple), the size of the tumor (the size based on the selected CT images), tumor shape (regular, e.g., round or oval, irregular, e.g., polycyclic, lobular), tumor density (uniform, uneven), and tumor margins (well defined, poorly defined).

Statistics

The diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the three CNN networks was calculated on the 917 CT images. The receiver operating characteristic (ROC) curves and the area under the curve (AUC) of the three networks were constructed and calculated. The diagnostic accuracy, sensitivity, specificity, PPV, and NPV; the three deep-learning network models (after voting); and the two radiologists were also calculated for the 283 tumors. The diagnostic accuracy, sensitivity, and specificity of onefold were compared, and the statistical significance was calculated between VGG19_BN, DenseNet169, radiologist A, radiologist B, and radiologist B (second time) with ResNet50 (after voting) using McNemar's test. The statistical analyses were conducted using SPSS 23.0 software (IBM SPSS Statistics Base Integrated Edition 23, Armonk, NY, USA).

Results

Diagnostic performance of three convolutional neural network models and radiologists

The classification performance of three networks was shown in Table 2. The accuracy of ResNet50, VGG16_bn, and DenseNet169 was 90.8%, 90.0%, and 87.3%, respectively (Figure 4). The ROC curves of the three networks were shown in Figure 5. The AUC of Resnet50, VGG16_bn, and DenseNet169 to differentiate malignant from benign tumors was 0.96, 0.96, and 0.95, respectively.

The attention heatmap was generated by CAM and then the heatmap was super-imposed on the original CT image, so that the location of parotid tumor and the region highlighted by ResNet could be compared. As showed in Figure 6, the attention heatmap highlighted important sub-regions for further clinical review. This showed that the abnormal characteristics of malignant parotid tumors had been learned by Resnet and used as the basis for its classification of benign and malignant tumors.

Diagnostic performance of different convolutional neural network models after voting

The accuracy, sensitivity and specificity, PPV, and NPV of the three networks after voting and the two radiologists were shown in Table 3. Statistical significance between VGG19_BN, DenseNet169, radiologists, and ResNet50 of onefold was shown in Table 4.

Discussion

The pre-operative diagnosis of benign and malignant tumors of the parotid gland is of great clinical significance and can have an important impact on surgical planning. Because of the important function of facial nerve, to preserve facial nerve function is a general and important principle in parotid tumors treatments. For benign lesions, local excision or partial parotidectomy is sufficient and every attempt to preserve facial nerve function should be made during surgery; however, for malignant tumors, a total parotidectomy with sacrifice of any part of the nerves overtly involved in tumor is desirable (26). For surgeons, pre-operatively recognition, the malignancy of parotid tumors is urgently hoped to be resolved, since this is helpful for more adequate pre-operative preparation, more appropriate operation (balance between preserving facial nerve function and avoiding recurrence).

Fine-needle aspiration cytology is used for the pre-operative diagnosis, and high specificity was showed by Piccioni et al.'s study and Dhanani et al.'s study (27, 28). However, due to the difficulty of sampling and the heterogeneity of the tumor, the sensitivity of recognition malignancy was not quite satisfactory (sensitivity: 73%–97%, specificity: 83%–97.9%) (5, 27, 28). In a meta-analysis by Schmidt et al., the sensitivity and specificity were 79% and 96% for malignancy (4). The relatively low sensitivity was due to few tissues obtained for diagnosis, so some malignant tumors would be misdiagnosed (false negative). Ultrasound-guided core needle biopsy could obtain an adequate tissue sample for histological evaluation, which allows classification of malignant and benign tumors and tumor grading. The sensitivity and specificity of it was much higher than fine-needle aspiration (29). However, compared with imaging technique, these two pre-operative techniques are invasive and has a risk of infection (15). A summary of the diagnosis results of fine-needle aspiration cytology and core needle biopsy was showed in Table 5.

In clinic, ultrasound, CT, and MRI are widely used in parotid tumors diagnosis (Table 5). The CT technique has high spatial resolution and rapid acquisition (34). CT images are useful in defining the anatomic localization, the extent, the density, the border, and delineation of tumors, and they are useful for detecting metastases and lymph nodes. However, it was not

TABLE 2 The diagnostic performance of the three convolutional neural network (CNN) models.

	Accuracy	Sensitivity	Specificity	PPV	NPV
ResNet50	90.8%	91.3%	90.4%	90.5%	91.2%
VGG19_BN	90.0%	85.2%	94.7%	94.2%	86.4%
DenseNet169	87.3%	86.1%	88.6%	88.4%	86.3%

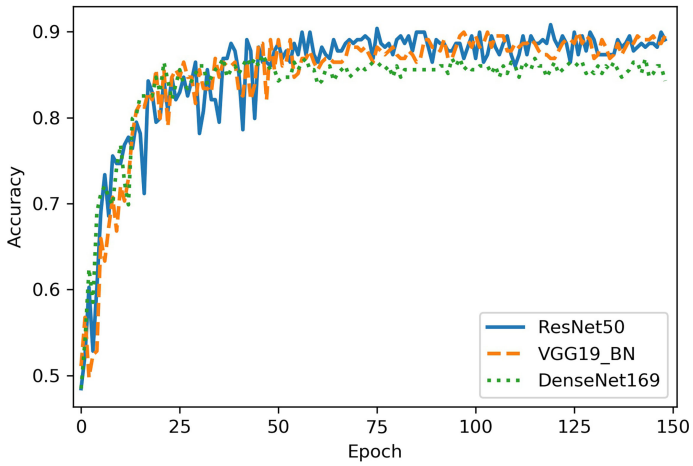


FIGURE 4
The diagnostic accuracy of Resnet50, VGG16_bn, and DenseNet169 in test set. The horizontal axis represents the training epochs, and vertical axis represents diagnostic accuracy. The best accuracy of ResNet50 is higher than VGG16_bn and DenseNet169.

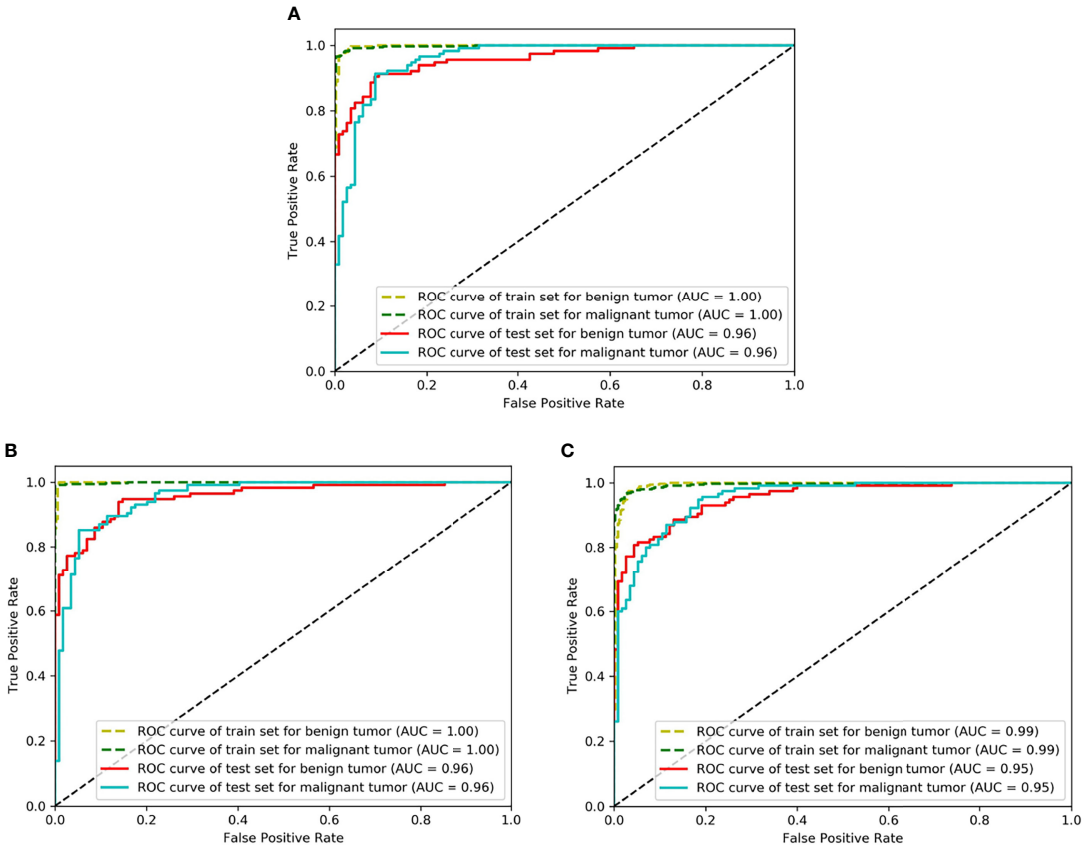


FIGURE 5
The receiver operating characteristic (ROC) curves of Resnet50, VGG16_bn, and DenseNet169. The horizontal axis represents the false positive rate and vertical axis represents the true positive rate. **(A)** Is the ROC curve of Resnet50 (AUC = 0.96 for malignant tumor), **(B)** Is the ROC curve of VGG16_bn (AUC = 0.96 for malignant tumor), and **(C)** Is the ROC curve of DenseNet169 (AUC = 0.95 for malignant tumor).

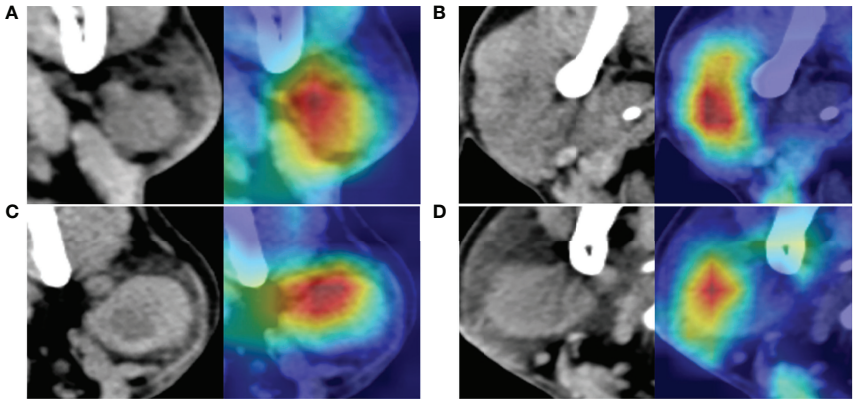


FIGURE 6
The images at the left side are the original computed tomography (CT) images; the images at the right side are the heatmaps drawn by class activation map (CAM). The red color shows where the network is focused to differentiate benign and malignant parotid tumor. **(A)** Is a benign tumor with homogeneous density and well-defined margin, network mainly focused on the texture and margin of tumor. **(B)** Is a malignant tumor with heterogeneous density and poor-defined margin, network mainly focused on the texture of tumor. **(C)** Is a benign tumor with intra-tumoral cystic component and well-defined margin, whereas network mainly focused on the upper part of the tumor and the margin of the tumor but not the intra-tumoral cystic component. **(D)** Is a malignant tumor with homogeneous density and relative well-defined margin, network mainly focused on texture of the left part of tumor but not the margin of tumor. All these tumors are correctly recognized by the neural network.

TABLE 3 The diagnostic performance of three convolutional neural network (CNN) models after voting and two radiologists.

	Accuracy	Sensitivity	Specificity	PPV	NPV
ResNet50	92.3%	93.5%	91.2%	90.6%	93.9%
VGG19_BN	92.3%	87.1%	97.1%	96.4%	89.2%
Densenet169	90.8%	90.3%	91.2%	90.3%	91.2%
Radiologist A	58.6%	32.9%	83.2%	65.3%	53.7%
Radiologist B	68.5%	49.7%	86.6%	78.0%	64.2%
Radiologist B (second time)	77.1%	74.2%	79.7%	76.6%	77.6%

Radiologist B (second time): The classification using all the original computed tomography (CT) images without cropping.

reliable in differentiating benign and malignant parotid tumors. Generally speaking, benign lesions reveal a well-defined and smooth border and have a homogeneous appearance (35–37). However, malignant parotid tumors could also display as a homogeneous density mass with well-defined border (6, 38). Recently, CNNs present high efficiency in image processing and classification tasks in many medical fields. Because of textures of parotid tumors differ depending on the underlying histopathological composition, neural network with pixel level of receptive field could extract more detailed image feature than human (39, 40). This supply a non-invasive pre-surgery malignancy identification of parotid tumors based on various images. By providing accurate, consistent, and instant results for the same input image, it could also increase the accuracy of diagnosis and reduce rote manual tasks, helping to simplify clinical workflow integration for radiologist.

In this study, the sensitivity and specificity of ResNet50 in distinguishing malignant from benign tumors were 91.3% and 90.4%, and the sensitivity and specificity reached 93.5% and 91.2% after voting. VGG16_bn presented a sensitivity of 87.1%, and Densenet169 presented a sensitivity of 90.3% after voting. All the three CNN networks presented high sensitivity, and the ResNet50 presented the relatively higher sensitivity. Meanwhile, the two radiologists had a sensitivity of 32.9% and 49.7%, and a specificity of 83.2% and 86.6%. And radiologist B had a sensitivity of 74.2% and a specificity of 79.7% using the whole original CT images without cropping. There were significant differences between radiologists with ResNet50 for diagnostic accuracy and sensitivity ($P < 0.05$). Because radiologist B (second time) had a diagnosis using all the original CT images without cropping, the diagnosis of accuracy and sensitivity increased. Our manual classification results were also similar with previous

10.3389/fonc.2022.919088

TABLE 4 Statistical significance between VGG19_BN, DenseNet169, radiologists, and ResNet50 of onefold.

	Accuracy (*P*-value)	Sensitivity (*P*-value)	Specificity (*P*-value)
VGG19_BN vs. ResNet50	1.00	0.51	1.00
DenseNet169 vs. ResNet50	1.00	1.00	1.00
Radiologist A vs. ResNet50	0.00	0.00	0.06
Radiologist B vs. ResNet50	0.00	0.00	0.07
Radiologist B (second time) vs. ResNet50	0.04	0.03	0.04

study (6–8). The inconsistency of manual classification also reflects the instability of manual diagnosis, and manual classification is more dependent on experience of radiologist. Because approximately 80% of parotid tumors are benign (26), this priori experience will make the radiologist more inclined to diagnose parotid tumors as benign. Considering the relatively low sensitivity of fine-needle aspiration cytology and manual classification, we think that the high sensitivity of our ResNet50 could be an important auxiliary diagnosis. For the CNN highly suspected malignant tumors, if the diagnosis of fine needle aspiration cytology is benign, maybe re-sampling, re-evaluation, or consultation of experienced cytologist is needed.

Recently, Chang et al. and Xia et al. also utilized neural network to differentiate benign and malignant parotid tumors on MRI images. In Chang et al.'s study, U-Net model based on MRI images of 85 parotid tumors (60 benign tumors and 25 malignant tumors) was used, and the diagnostic accuracy, sensitivity, and specificity were 71%, 33%, and 87% for malignant tumors. In Xia et al.'s study, a modified ResNet model was developed based on MRI images of 233 parotid tumors (153 benign tumors and 80 adenocarcinoma), and the accuracy, sensitivity, and specificity were 88.2%, 94.6%, and 81.7% for differentiating benign from malignant parotid lesions. And studies using CNN networks based on portal phase CT images (contrast-enhanced CT) and ultrasound images were also published recently (32, 33). In our study, 283 parotid tumors (150 benign tumors and 133 malignant tumors) were used for training and testing. More malignant parotid tumors were included in our database, and a high sensitivity of differentiating malignant from benign parotid lesions was presented. Compared with these CNN studies (Table 5), our study had a relatively large sample size, more balanced benign and malignant parotid tumors and relatively high sensitivity of differentiating malignant from benign parotid tumors. We speculate that maybe more malignant parotid samples trained are the reason for high sensitivity of recognition malignant ones in this study.

In this study, in order to simulate the process of human diagnosis, a voting model was built at the end of the three deep-learning network models, and the accuracy, sensitivity, and specificity of the three CNN models were calculated for the 283 tumors. After voting, the three CNN models all showed higher diagnostic efficiency than the models without voting. The pathology diagnosis was a microscopic diagnosis, so the tumors will be diagnosed as malignant if there are malignant cells. However, imaging reflects the macroscopic morphology, so imaging diagnosis is a probabilistic diagnosis and it needs comprehensive analysis. Zhao et al. recently reported a hybrid algorithm, a Bayesian network branch performing probabilistic causal relationship reasoning and a graph convolutional network branch performing more generic relational modeling and reasoning using a feature representation (41). And their hybrid algorithm achieves a much more robust performance than pure neural network architecture.

TABLE 5 The diagnostic sensitivity and specificity of different diagnostic methods.

Diagnostic method	Sensitivity	Specificity
Fine-needle biopsy (4)	79%	96%
Core-needle biopsy (29)	98%	94%
Conventional MRI (30)	76%	91%
Plain CT (6–8)	10-50%	85-95%
Ultrasound (elastography) (31)	67%	64%
CT enhanced scan (DL) (32)	96.7% (first group) 76.7% (second group)	98.9% (first group) 78.8% (second group)
Ultrasound (DL) (33)	77%	81%
MRI (DL) (15)	33%-81.7%	87%-94.6%

DL, deep learning.

In this study, we re-analyzed the mis-diagnosed parotid tumors of CNNs. And we found that the accuracy for lymphoma diagnosis was 73.3% (11/15), for ResNet50 (after voting), this was much lower than the whole database. The lymphomas usually appear as homogeneous and sharply demarcated nodes, just like benign tumors; this maybe the reasons of misdiagnosis. Moreover, most mis-diagnosed malignant parotid tumors were small ones; they were homogeneous and well-defined, and similar with benign tumors. So, lymphoma and small tumors are more likely be mis-diagnosed even for CNNs. Furthermore, using the attention heatmap, we can infer which part of the input image is focused on by the neural network. For some tumors, the highlighted areas were on the margins, and for others, the highlighted areas were intra-tumoral, which means that the neural network focused on

the texture of images of these tumors (Figure 6). Interpretability is increased through the attention heatmap generated. It may provide new ways of thinking in the diagnosis of parotid tumor.

This study still has several limitations. First, the tumor data included need to be further expanded to get a stable result for neural network. Although most of the parotid tumor is included in this study, some types of tumor are still limited for clinical application. Second, there is no auto-segment or automated detection (R-CNN or Yolo) built in the neural networks. A neural network with auto-segmentation or automated detection need to be explored in further study. And networks using a voxel-based domain and the whole CT images without cropping, combining the radiological with clinical findings are also needed to be explored in the future. Third, this study was based on a single center; an external validation study is needed to validate its diagnostic performance and generalizability. Prospective and multi-institutional datasets are also needed in future studies.

Conclusion

ResNet50 presented high sensitivity in distinguishing malignant from benign parotid tumors on plain CT images, and this made it a promising auxiliary diagnostic method to screen malignant parotid tumors.

Data availability statement

The data used to support the findings of this study were supplied by Zitong Lin under license and so cannot be made readily available. Requests for access to these data should be made to linzitong_710@163.com.

Ethics statement

The approval from the Ethics Committee of our University was obtained prior to performing this study (NJSH-2022NL-069). The data are anonymous, and the requirement for written informed consent was therefore waived.

Author contributions

Z.Y. Hu contributed to writing of this manuscript, neural network modeling, data analysis and visualization. B.X. Wang contributed to neural network modeling and optimization, data analysis and visualization. X. Pan contributed to collection of materials and data curation. D.T. Cao contributed to data analysis. A.T. Gao contributed to collection of materials. X.D. Yang contributed to study design and interpretation of the results. Y. Chen contributed to study design and neural network modeling and optimization. Z.T. Lin contributed to conceptualization, CT image consultation, review and editing of this manuscript, and funding acquisition. All authors gave final approval and agree to be accountable for all aspects of the work.

Funding

This work was supported by the General project of Jiangsu Commission of Health (M2021077), the Jiangsu Province Medical Association Roentgen Imaging Research and Special Project Funds (SYH-3201150-0007), the Medical Science and Technology Development Foundation (YKK19090), and the Nanjing Clinical Research Center for Oral Diseases (no. 2019060009).

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

References

1. Ettl T, Schwarz-Furlan S, Gosau M, Reichert TE. Salivary gland carcinomas. *Oral Maxillofac Surg* (2012) 16:267–83. doi: 10.1007/s10006-012-0350-9

2. Wong DS. Signs and symptoms of malignant parotid tumours: an objective assessment. *J R Coll Surg Edinb* (2001) 46:91–5.

3. Ozawa N, Okamura T, Koyama K, Nakayama K, Kawabe J, Shiomi S, et al. Retrospective review: usefulness of a number of imaging modalities including CT, MRI, technetium-99m pertechnetate scintigraphy, gallium-67 scintigraphy and f-18-FDG PET in the differentiation of benign from malignant parotid masses. *Radiat Med* (2006) 24:41–9. doi: 10.1007/BF02489988

4. Schmidt RL, Hall BJ, Wilson AR, Layfield LJ. A systematic review and meta-analysis of the diagnostic accuracy of fine-needle aspiration cytology for parotid gland lesions. *Am J Clin Pathol* (2011) 136:45–59. doi: 10.1309/AJCPOIE0CZNAT6SQ

5. Kristjan GJ, Aida A, Jahan A. The accuracy of fine-needle aspiration cytology for diagnosis of parotid gland masses: a clinicopathological study of

114 patients. *J Appl Oral ence* (2016) 24:561–7. doi: 10.1590/1678-775720160214

6. Berg HM, Jacobs JB, Kaufman D, Reede DL. Correlation of fine needle aspiration biopsy and CT scanning of parotid masses. *Laryngoscope* (1986) 96:1357–62. doi: 10.1288/00005537-198612000-00008

7. Whyte AM, Byrne JV. A comparison of computed tomography and ultrasound in the assessment of parotid masses. *Clin Radiol* (1987) 38:339–43. doi: 10.1016/S0009-9260(87)80203-9

8. Urquhart A, Hutchins LG, Berg RL. Preoperative computed tomography scans for parotid tumor evaluation. *Laryngoscope* (2001) 111:1984–8. doi: 10.1097/00005537-200111000-00022

9. Qi X, Hu J, Zhang L, Bai S, Yi Z. Automated segmentation of the clinical target volume in the planning CT for breast cancer using deep neural networks. *IEEE Trans Cybern* (2020) 52(5):3446–3456. doi: 10.1109/TCYB.2020.3012186

10. Hamidian S, Sahiner B, Petrick N, Pezeshk A. 3D convolutional neural network for automatic detection of lung nodules in chest CT. *Proc SPIE Int Soc Opt Eng* (2017), 10134:1013409. doi: 10.1117/12.2255795

11. Yasaka K, Akai H, Abe O, Kiryu S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: A preliminary study. *Radiology* (2018) 286:887–96. doi: 10.1148/radiol.2017170706

12. QaW W, Sun Z, Zhang Yu, Li X, Ge W, Huang Y, et al. SCCNN: A diagnosis method for hepatocellular carcinoma and intrahepatic cholangiocarcinoma based on Siamese cross contrast neural network. *IEEE Access* (2020) 8:85271–83. doi: 10.1109/ACCESS.2020.2992627

13. Ma Z, Zhou S, Wu X, Zhang H, Yan W, Sun S, et al. Nasopharyngeal carcinoma segmentation based on enhanced convolutional neural networks using multi-modal metric learning. *Phys Med Biol* (2019) 64:025005. doi: 10.1088/1361-6560/aaf5da

14. Chang YJ, Huang TY, Liu YJ, Chung HW, Juan CJ. Classification of parotid gland tumors by using multimodal MRI and deep learning. *NMR Biomed* (2021) 34:e4408. doi: 10.1002/nbm.4408

15. Xia X, Feng B, Wang J, Hua Q, Yang Y, Sheng L, et al. Deep learning for differentiating benign from malignant parotid lesions on MR images. *Front Oncol* (2021) 11:632104. doi: 10.3389/fonc.2021.632104

16. Howlett DC, Kesse KW, Hughes DV, Sallomi DF. The role of imaging in the evaluation of parotid disease. *Clin Radiol* (2002) 57:692–701. doi: 10.1053/crad.2001.0865

17. Prasad RS. Parotid gland imaging. *Otolaryngol Clin North Am* (2016) 49:285–312. doi: 10.1016/j.otc.2015.10.003

18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *2016 IEEE conference on computer vision and pattern recognition (CVPR)* (2016). PP:770–8. doi: 10.1109/CVPR.2016.90

19. Liu H, Cao H, Song E, Ma G, Xu X, Jin R, et al. A cascaded dual-pathway residual network for lung nodule segmentation in CT images. *Phys Med* (2019) 63:112–21. doi: 10.1016/j.ejmp.2019.06.003

20. Li Q, Yu B, Tian X, Cui X, Zhang R, Guo Q. Deep residual nets model for staging liver fibrosis on plain CT images. *Int J Comput Assist Radiol Surg* (2020) 15:1399–406. doi: 10.1007/s11548-020-02206-y

21. Liu C, Liu C, Lv F, Zhong K, Yu H. Breast cancer patient auto-setup using residual neural network for CT-guided therapy. *IEEE Access* (2020) 8:201666–74. doi: 10.1109/ACCESS.2020.3035809

22. Liu W, Liu X, Peng M, Chen GQ, Liu PH, Cui XW, et al. Artificial intelligence for hepatitis evaluation. *World J Gastroenterol* (2021) 27:5715–26. doi: 10.3748/wjg.v27.i34.5715

23. Simonyan K, Zisserman A. Very deep convolutional networks for Large-scale image recognition. *Comput Science* (2014) 1–1. doi: 10.48550/arXiv.1409.1556

24. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4. In: *Inception-ResNet and the impact of residual connections on learning* (2016). AAAI Press (2017) PP:4278–4284. doi: 10.1609/aaai.v31i1.11231

25. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. *Rethinking the inception architecture for computer vision*. IEEE (2016). PP:2818–26. doi: 10.1109/CVPR.2016.308

26. Sood S, McGurk M, Vaz F. Management of salivary gland tumours: United kingdom national multidisciplinary guidelines. *J Laryngol Otol* (2016) 130:S142–s149. doi: 10.1017/S0022215116000566

27. Piccioni LO, Fabiano B, Gemma M, Sarandria D, Bussi M. Fine-needle aspiration cytology in the diagnosis of parotid lesions. *Acta Otorhinolaryngol Ital* (2011) 31:1–4.

28. Dhanani R, Iftikhar H, Awan MS, Zahid N, Momin SNA. Role of fine needle aspiration cytology in the diagnosis of parotid gland tumors: Analysis of 193 cases. *Int Arch Otorhinolaryngol* (2020) 24:e508–12. doi: 10.1055/s-0040-1709111

29. Kim HJ, Kim JS. Ultrasound-guided core needle biopsy in salivary glands: A meta-analysis. *Laryngoscope* (2018) 128:118–25. doi: 10.1002/lary.26764

30. Liang YY, Xu F, Guo Y, Wang J. Diagnostic accuracy of magnetic resonance imaging techniques for parotid tumors, a systematic review and meta-analysis. *Clin Imaging* (2018) 52:36–43. doi: 10.1016/j.clinimag.2018.05.026

31. Zhang YF, Li H, Wang XM, Cai YF. Sonoelastography for differential diagnosis between malignant and benign parotid lesions: a meta-analysis. *Eur Radiol* (2019) 29:725–35. doi: 10.1007/s00330-018-5609-6

32. Zhang H, Lai H, Wang Y, Lv X, Chen C. Research on the classification of benign and malignant parotid tumors based on transfer learning and a convolutional neural network. *IEEE Access* (2021), 9:40360–71. doi: 10.1109/ACCESS.2021.3064752

33. Wang Y, Xie W, Huang S, Feng M, Ke X, Zhong Z, et al. The diagnostic value of ultrasound-based deep learning in differentiating parotid gland tumors. *J Oncol* (2022) 2022:8192999. doi: 10.1155/2022/8192999

34. National Center for Health S. Health, united states. In: *Health, united states, 2009: With special feature on medical technology*. Hyattsville (MD: National Center for Health Statistics (US (2010).

35. Isaza M, Ikezoe J, Morimoto S, Takashima S, Kadowaki K, Takeuchi N, et al. Computed tomography and ultrasonography in parotid tumors. *Acta Radiol* (1989) 30:11–1. doi: 10.1177/028418518903000103

36. Christe A, Waldherr C, Hallett R, Zbaeren P, Thoeny H. MR imaging of parotid tumors: typical lesion characteristics in MR imaging improve discrimination between benign and malignant disease. *AJNR Am J Neuroradiol* (2011) 32:1202–7. doi: 10.3174/ajnr.A2520

37. Kato H, Kanematsu M, Watanabe H, Mizuta K, Aoki M. Salivary gland tumors of the parotid gland: CT and MR imaging findings with emphasis on intratumoral cystic components. *Neuroradiology* (2014) 56:789–95. doi: 10.1007/s00234-014-1386-3

38. Golding S. Computed tomography in the diagnosis of parotid gland tumours. *Br J Radiol* (1982) 55:182–8. doi: 10.1259/0007-1285-55-651-182

39. Okahara M, Kiyosue H, Hori Y, Matsumoto A, Mori H, Yokoyama S. Parotid tumors: MR imaging with pathological correlation. *Eur Radiology* (2003) 13:L25–33. doi: 10.1007/s00330-003-1999-0

40. Sarioglu O, Sarioglu FC, Akdogan AI, Kucuk U, Arslan IB, Cukurova I, et al. MRI-Based texture analysis to differentiate the most common parotid tumours. *Clin Radiol* (2020) 75:877.e815–877.e823. doi: 10.1016/j.crad.2020.06.018

41. Zhao G, Feng Q, Chen C, Zhou Z, Yu Y. Diagnose like a radiologist: Hybrid neuro-probabilistic reasoning for attribute-based medical image diagnosis. *IEEE Trans Pattern Anal Mach Intell* (2021) PP:1–1. doi: 10.1109/TPAMI.2021.3130759

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fonc.2022.969707

OPEN ACCESS

EDITED BY
Wei Wei,
Xi'an University of Technology, China

REVIEWED BY
Qiongwen Zhang,
Sichuan University, China
Lizong Shen,
Jiangsu Provincial Hospital of Traditional Chinese Medicine, China

*CORRESPONDENCE
Ming Cheng
fccchengm@zzu.edu.cn

†These authors have contributed equally to this work and share first authorship

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

RECEIVED 15 June 2022
ACCEPTED 05 September 2022
PUBLISHED 23 September 2022

CITATION
Zhang A-q, Zhao H-p, Li F, Liang P, Gao J-b and Cheng M (2022) Computed tomography-based deep-learning prediction of lymph node metastasis risk in locally advanced gastric cancer.
Front. Oncol. 12:969707.
doi: 10.3389/fonc.2022.969707

Computed tomography-based deep-learning prediction of lymph node metastasis risk in locally advanced gastric cancer

An-qi Zhang[1†], Hui-ping Zhao[2†], Fei Li[3], Pan Liang[1,4], Jian-bo Gao[1,4] and Ming Cheng[4,5*]

[1]Department of Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2]Department of Radiology, Shaanxi Provincial People's Hospital, Xi'an, China, [3]School of Cyber Science and Engineering, Wuhan University, Wuhan, China, [4]Department of Medical Information, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [5]Henan Key Laboratory of Image Diagnosis and Treatment for Digestive System Tumor, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Purpose: Preoperative evaluation of lymph node metastasis (LNM) is the basis of personalized treatment of locally advanced gastric cancer (LAGC). We aim to develop and evaluate CT-based model using deep learning features to preoperatively predict LNM in LAGC.

Methods: A combined size of 523 patients who had pathologically confirmed LAGC were retrospectively collected between August 2012 and July 2019 from our hospital. Five pre-trained convolutional neural networks were exploited to extract deep learning features from pretreatment CT images. And the support vector machine (SVM) was employed as the classifier. We assessed the performance using the area under the receiver operating characteristics curve (AUC) and selected an optimal model, which was compared with a radiomics model developed from the training cohort. A clinical model was built with clinical factors only for baseline comparison.

Results: The optimal model with features extracted from ResNet yielded better performance with AUC of 0.796 [95% confidence interval (95% CI), 0.715-0.865] and accuracy of 75.2% (95% CI, 67.2%-81.5%) in the testing cohort, compared with 0.704 (0.625-0.783) and 61.8% (54.5%-69.9%) for the radiomics model. The predictive performance of all the radiological models were significantly better than the clinical model.

Conclusion: The novel and noninvasive deep learning approach could provide efficient and accurate prediction of lymph node metastasis in LAGC, and benefit clinical decision making of therapeutic strategy.

KEYWORDS

deep learning, locally advanced gastric cancer, lymph node metastasis, radiomics, computed tomography

Introduction

Gastric cancer (GC) is one of the most common cancers and the third leading cause of death from cancer worldwide (1). The incidence rate of gastric cancer is relatively high in Asia, South American and Europe (2–4). Locally advanced gastric cancer refers to the wall invasion deeper than the submucosa, with a high rate of lymph node metastasis (LNM) and poor clinical prognosis (5–7). Accurate evaluation on lymphatic metastasis based on preoperative computed tomography (CT) images is crucial for individual treatment of LAGC (8–10). Preoperative knowledges of LNM have important clinical significance for selecting the optimal surgical procedure (endoscopic procedures or gastrectomy plus lymph node dissection) and the need for adjuvant therapy (11–13). The National Comprehensive Cancer Network recommended CT as a first-line imaging technique for detecting LNM, but the overall accuracy is 50%-70%, which is unsatisfactory (14).

The advances in deep learning techniques provides a new field for CT imaging analysis, which could convert medical images to mineable data and generate thousands of quantitative features (15). Convolutional neural networks (CNNs) have been proved to be an effective method for improving the diagnostic accuracy of medical imaging (16–18). Due to the lack of enough annotated cases, training a CNN model from scratch for one specific clinical problem often is infeasible. An effective approach is to adopt the transfer learning technique using pre-training CNNs, which ran additional steps of pre-training on specific medical domain from the existing checkpoint. It is frequently used to alleviate the limitations of small datasets and expensive annotation (19, 20). Part of natural imaging descriptors developed for object detection have been used for lesion segmentation in medical imaging analysis (21). Another option is to use a pretrained CNNs models as the feature extractor and traditional machine learning methods as classifier, which may also have satisfactory performance in terms of prediction accuracy and computational cost (22). Handcrafted radiomics have been studied extensively for radiological diagnosis and prediction (8, 23, 24). However, the application of transfer learning to prediction of LNM in gastric cancer has not been explored.

In this study, we hypothesize that CT-based transfer learning techniques are feasible to extract deep learning features for preoperatively predicting LNM risk. To this end, our study aims to build a noninvasive measurement based on pre-trained deep learning models for the preoperatively prediction of LNM in patients with gastric cancer, making comparison with the handcrafted radiomics method. Additionally, we further explored the application value of deep learning features in predicting LNM and making treatment decisions.

Materials and methods

Patients

This retrospective study was approved by the institutional review board of our hospital, and the requirement for informed consent was waived. A total of 523 consecutive patients with gastric cancer who was treated between August 2012 and July 2019 were enrolled. The patients were enrolled based on following inclusion criteria: (a) pathologically diagnosed as local advance gastric cancer (pT2-4aNxM0); (b) all patients with gastrectomy plus lymph node dissection and CT imaging data were complete; (c) without any systematic or local treatment before CT imaging study or surgery; (d) the lesion covers at least 3 slices on CT cross section. The patients were excluded based on the following criteria: (a) invisible lesion on CT images; (b) insufficient stomach distension; (c) poor image quality for post-processing due to artifacts. The flowchart of patient selection was shown in Figure 1. We adopted computer-generated random numbers to split the training cohort (n=367, 74.40% males; mean age, 59.75 ± 10.38; range, 22-82 years) and the testing cohort (n=156, 73.98% males; mean age, 59.36 ± 9.94; range, 22-81 years). The tumor location information was got from the medical or endoscopic reports, and the clinical information was got by reviewing the medical reports.

Image process and tumor segmentation

All patients underwent contrast-enhanced CT scan and informed consent forms were signed before inspection. The CT scans were acquired with breath-hold with the patient head first supine in all of the phases for covering the whole abdomen. The details on CT acquisition parameters were described in Supplemental Material.

Tumor regions of interest (ROIs) were manually segmented CT images by two experienced radiologists using ITK-SNAP software (version 3.6.0; http://www.itksnap.org). In order to make a fair comparison with different features, we only chosen one slice with the maximum cross-sectional area of the tumor lesion by the radiologists. We randomly chosen 30 patients from training cohorts to assess the interobserver reproducibility for ROI-based radiomics features in a blinded manner. After one month, segmentation procedure was repeated to assess the intraobserver reproducibility. The features with intra-class correlation coefficient (ICC) greater than 0.75 were selected for further analysis. For deep learning features extraction, the 3 axial slices containing the delineated tumor were resized to 224× 224mm (the size for the input layer of the pretrained CNN models) with the use of a bounding box covering the radiologist contoured tumor area.

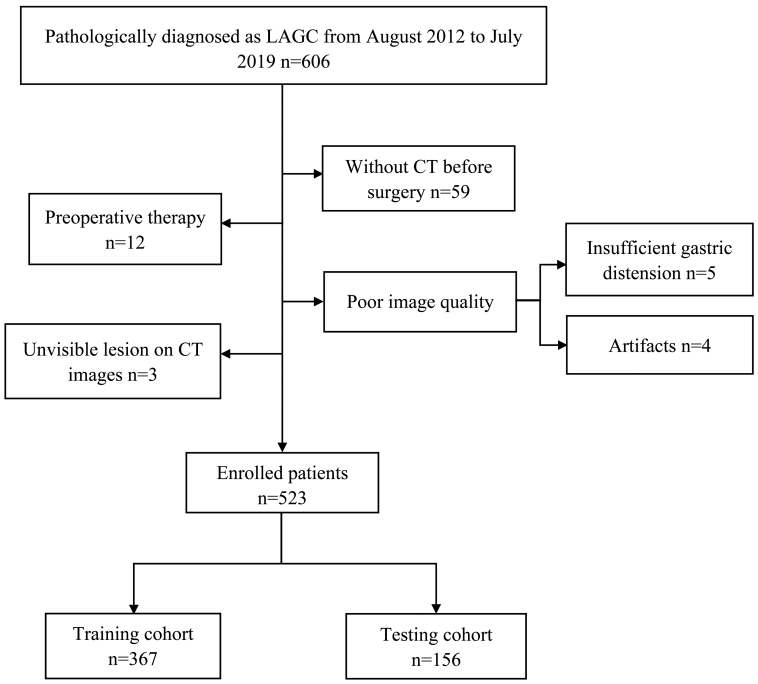


FIGURE 1
Flow chart of patient selection.

Deep learning features

We employed five commonly convolutional neural networks (ResNet (25), VGG16 (26), VGG19 (26), Xception (27) and InceptionV3 (28)) as base models to extract deep learning features automatically. These five CNNs models were pre-trained on the large-scale lightweight well-annotated biomedical image database (29). We removed the last fully connected layer at top of the network, and applied global max pooling strategies to efficiently capture the maximum values of each layer of the feature maps. Finally, we converted the feature maps to the raw values. The extracted deep features were used to construct the machine learning model. Due to the complexity of deep learning model structure, the potential mechanisms of predictive value were unclear. Additional details of deep features extraction in this study are listed in Supplementary Material. Furthermore, Gradient-weighted Class Activation Mapping technique (Grad-CAM) could generate visual explanations for any CNN-based model (30). We use this visualization technique to investigate which regions of the ROI were most important in the deep features.

Radiomics features

Image standardization was implemented before feature extraction: bi-cubic spline interpolation was used to resample the image scale in the slice to reduce the heterogeneity results from different scanners, resulting in a voxel size of 1mm×1mm×1mm (31, 32). The radiomics features were automatically extracted from each radiologist's ROIs using the Python package Pyradiomics (http://pyradiomics.readthedocs.io) (33). The radiomics features were standardized by referring to the Image Biomarker Standardization Initiative (IBSI) (34). The study was based on the reporting guidelines of IBSI. The hand-crafted radiomics features were divided into three different groups of features: shape features, histogram statistics, second order features: Gray Level Co-occurrence Matrix (GLCM), Gray Level Dependence Matrix (GLDM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Neighboring Gray Tone Difference Matrix (NGTDM). Most features mentioned above were delineated according to the IBSI, and the detailed introduction of the features were described in Supplementary Material.

Harmonization

Radiomics extracts features from medical images more precisely than general visual evaluation. However, radiomics features are affected by the acquisition protocol and reconstruction methods, thus obscuring underlying biologically important texture features. In practical clinical retrospective studies, it is impractical to standardize the parameters of different devices in advance. In order to reduce the batch effect, ComBat harmonization technique had been successfully applied to properly correct radiomic feature values from different scanner or protocol effect (35). We exploited the ComBat to pool and harmonize radiomics and deep learning features after extraction.

Feature selection and model construction

Based on the training set, we performed deep learning or radiomics feature selection and constructed model for predicting lymph node metastasis. Firstly, the z-score normalization was used for standardization. In addition, we selected top 20% best features by univariate analysis. Then, we used an embedded feature selection approach based on the least absolute shrinkage and selection operator (LASSO) algorithm to select the most predictive features. Classification model was constructed by the SVM (36). We also built a clinical model based on the clinical characteristics. The code for model construction is available on Github (https://github.com/cmingwhu/DL-LNM).

Statistical analysis

P values for differences in the clinical characteristics between cohorts were assessed by Fisher's exact test or Chi-square test for categorical variables, and the Mann-Whitney U test or independent t-test for numeric variables. Receiver operating characteristic curve (ROC) was adopted to determine the predictive performance of the related models, while the DeLong's test was adopted for comparison of AUC between each model. The AUC and 95% confidence interval (CI) were calculated. Accuracy, specificity and sensitivity were calculated to assess the diagnostic performance. The calibration of the model was evaluated by the calibration curves using the Hosmer-Lemeshow test. To assess the reproducibility of our results, we randomly divided the patients into training or testing set ten times. Subsequently, the model was reconstructed and validated repeatedly. P value < 0.05 was considered statistically significant. We used Python version 3.6 (https://www.python.org/) and R version 4.0.3 (https://www.r-project.org) to perform statistical analysis and graphic production. The packages used in this study are shown in Supplementary Material.

Results

Clinical characteristics

Figure 2 depicts the workflow processes. Of the 523 patients (mean age: 59.64 ± 10.24 years; male: 74.40%) with locally advance gastric cancer for this study, 367 patients were assigned to the training cohort, and 156 patients was assigned for testing cohort. Clinical characteristics in two cohorts are shown in Table 1. No significant difference was identified in terms of sex, age, tumor location, tumor thickness between the two cohorts (Tables 1, S1). Tumor diameter, clinical T stage, and CT-reported LN differed significantly between LNM-negative and positive group in two cohorts ($p < 0.05$). Finally, a clinical model was established (incorporating tumor diameter, CT-reported LN and clinical T stage) for predicting LNM, yielding an AUC of 0.683 and 0.756 for testing and training cohorts, respectively, as shown in Tables 2, S2.

Handcrafted radiomics model construction

851 handcrafted radiomics features were extracted, where 107 were from the original images and 744 were from the wavelet filtered images. After ComBat harmonization (35). Forty-eight features were selected, including three and forty-five from original and wavelet filtered images (Table S3). The handcrafted radiomics model get an AUC of 0.704, C-index of 0.704, accuracy of 61.8%, sensitivity of 56.5%, specificity of 73.5%, positive predictive value (PPV) of 82.4%, and negative predictive value (NPV) of 43.4% in the testing cohort, and an AUC of 0.779, C-index of 0.779, accuracy of 74.0%, sensitivity of 77.5%, specificity of 66.4%, positive predictive value (PPV) of 83.2%, and negative predictive value (NPV) of 57.9% in the training cohort in Tables 2, S2.

Deep learning model construction

For predicting LNM based on deep learning features, we compared five CNNs models which were adopted to extract deep features to optimize the prediction performance. The AUC ranged from 0.578 to 0.796 for testing cohort, and 0.804 to 0.897 for training cohort, as shown in Table 2, S2. The ResNet-SVM model containing 116 deep learning features could get the best classification performance among the five CNNs models and was superior to the radiomics model, and yielding an AUC of 0.796, C-index of 0.796, accuracy of 75.2%, sensitivity of 80.2%, specificity of 64.7%, PPV of 82.5%, NPV of 61.1% in the testing cohort in Figures 3A, B. The calibration and favorable clinical benefit could also get

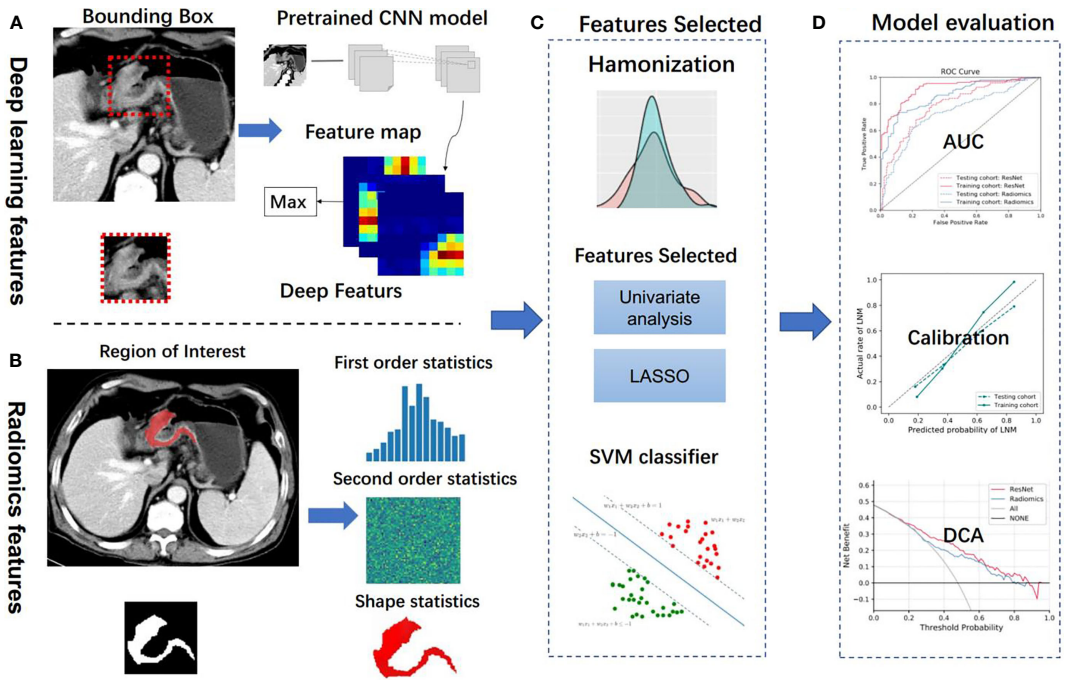


FIGURE 2
Analysis flowchart. **(A, B)** Features extraction from the deep learning method and handcrafted radiomics method. **(C)** Machine learning methods were employed in model construction. **(D)** Model evaluation. CNN, convolutional neural network; LASSO, the least absolute shrinkage and selection operator; SVM, support vector machine; AUC, area under the receiver operating characteristic curve; DCA, decision curve analysis.

TABLE 1 The clinical characteristics of patients in the training and testing cohorts.

Characteristics	Training cohort (120: 247)			Testing cohort (47: 109)		
	LNM (-)	LNM (+)	P value	LNM (-)	LNM (+)	P value
Age (mean ± SD, years)	59.98 ± 10.53	59.83 ± 10.40	0.834	60.98 ± 10.91	58.66 ± 9.46	0.182
Sex						
Female	28 (23.3)	66 (26.7)	0.569	16 (13.04)	24 (22.0)	0.168
Male	92 (76.7)	181 (73.3)		31 (86.96)	85 (78.0)	
Location						
Cardia/fundus	67	130	0.453	26	52	0.655
Body	23	51		11	26	
Antrum	29	57		10	29	
More than two-thirds of stomach	1	9		0	2	
Tumor thickness ± SD (mm)	22.65 ± 8.58	23.43 ± 7.70	0.383	21.71 ± 7.67	21.93 ± 7.46	0.865
Tumor diameter ± SD (mm)	82.60 ± 41.59	94.22 ± 51.70	0.032*	70.29 ± 30.46	90.36 ± 51.69	0.014*
Clinical T stage						
T2	13	21	0.005*	9	13	0.006*
T3	81	130		33	57	
T4a	26	96		5	39	
CT-reported LN						
Negative	90	76	<0.001*	39	37	<0.001*
Positive	30	181		8	72	

LNM, lymph node metastasis; (-), negative; (+), positive; *p < 0.05.

TABLE 2 Predictive performance of radiological or clinical models in the testing cohort.

	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
InceptionResNetV2	0. 707	65.6	67.9	60.8	78.3	47.7
	(0.653, 0.771)	(60.1, 72.7)	(55.9, 72.2)	(56.2, 73.3)	(73.0, 85.3)	(31.7, 53.8)
VGG16	0.661	61.8	65.2	58.0	63.2	60.0
	(0.540, 0.745)	(55.7, 70.6)	(60.6, 69.9)	(51.8, 65.8)	(55.5, 70.0)	(53.7, 69.4)
VGG19	0.578	49.6	40.6	68.6	72.9	35.7
	(0.507, 0.661)	(41.7, 55.1)	(40.6, 51.9)	(63.0, 75.6)	(66.0, 80.6)	(30.6, 47.9)
ResNet50	0.796	75.2	80.2	64.7	82.5	61.1
	(0.715-0.865)	(67.2, 81.5)	(75.4, 84.2)	(58.2, 71.6)	(74.9, 87.3)	(55.5, 69.3)
Xception	0.660	62.4	65.1	56.9	75.8	43.9
	(0.607, 0.759)	(56.2, 71.6)	(52.2, 69.0)	(49.8, 68.7)	(70.9, 81.1)	(40.9, 51.6)
Radiomics	0.704	61.8	56.5	73.5	82.4	43.4
	(0.625, 0.783)	(54.5, 69.9)	(50.8,62.3)	(68.8, 79.8)	(75.8,87.3)	(40.8, 52.1)
Clinical signature	0.683	68.2	67.6	67.7	70.7	53.6
	(0.632, 0.721)	(65.3, 72.1)	(63.5,70.1)	(63.1, 71.6)	(67.5, 75.2)	(50.7, 61.9)

good performance in Figures 3C, D. The number of features were adopted in model of different CNNs are listed in Table S4. Features maps from the ResNet model could indicate the locations that were important in generating the output. With the segmentation of the tumor region delineated, the informative slices (one slice with the maximum tumor area) were cropped to 224 × 224 mm using a bounding box covering the whole tumor area. The cropped images were used to generate the features from ResNet and the visualization of feature heatmaps were generated based on the Guided Grad-CAM, as shown in Figure 4. The tumoral lesion and perifocal areas in images were of great valuable for the feature pattern extraction. Then, we further analyzed the performance generated by features extracted from different layers to see whether the last layer was the most suitable to extract features. The current features extraction strategy is the best for ResNet in Table S5.

Different classifiers and features selection methods could greatly affect the prediction performance. For the features extracted from different CNNs, we compared the cross combination of multiple classifiers and feature selection methods. We find that the performances of different combinations are different, the results shown that the current combination method of classifier and extraction (ResNet-SVM) demonstrated the best discrimination ability with an AUC of 0.796 (95% CI, 0.715-0.865) for our dataset, as shown in Figure S1 and Table S6, but further generalization tests on other datasets are required. The DeLong test showed that there were significant improvements in contrast to the radiomics model and the clinical signature ($p < 0.05$), which yielded AUCs of 0.704 (95% CI, 0.625–0.783) and 0.683 (95% CI, 0.632–0.721), respectively.

Radiomics-deep learning combined model

We further integrated the radiomics and deep learning features to explore whether the predictive capability could be improved. After combination with radiomics and deep features, the prediction performance had not been improved, with a comparable AUC of 0.787 in the testing cohort in Figure S2. In addition, we further evaluated the addition of clinical factors to radiomics or deep learning features for potential improvement of prediction performance. The combination of deep and/or radiomics features with clinical features were incorporated into the model construction, the experimental results showed that combination of clinical factors could not increase the prediction performance in the testing cohort in Figures S3, S4.

Discussion

In this retrospective study, we applied deep transfer learning techniques to build a CT imaging-based prediction model for LNM prediction in gastric cancer. Our previous studies shown that the noninvasive deep learning CT image-based radiomics model was effective for LNM prediction and prognosis in GC (37). Hereby, we adopted transfer learning technique and extract deep learning features from five different pre-trained CNNs. Finally, the ResNet-SVM model could achieve better performance than the handcrafted radiomics and clinical models. In addition, different gastric cancers have different potentials for lymph node metastasis due to the heterogeneity and complexity of primary tumors. Previous studies clarified that the tumor size were independent risk factors for LNM. Our

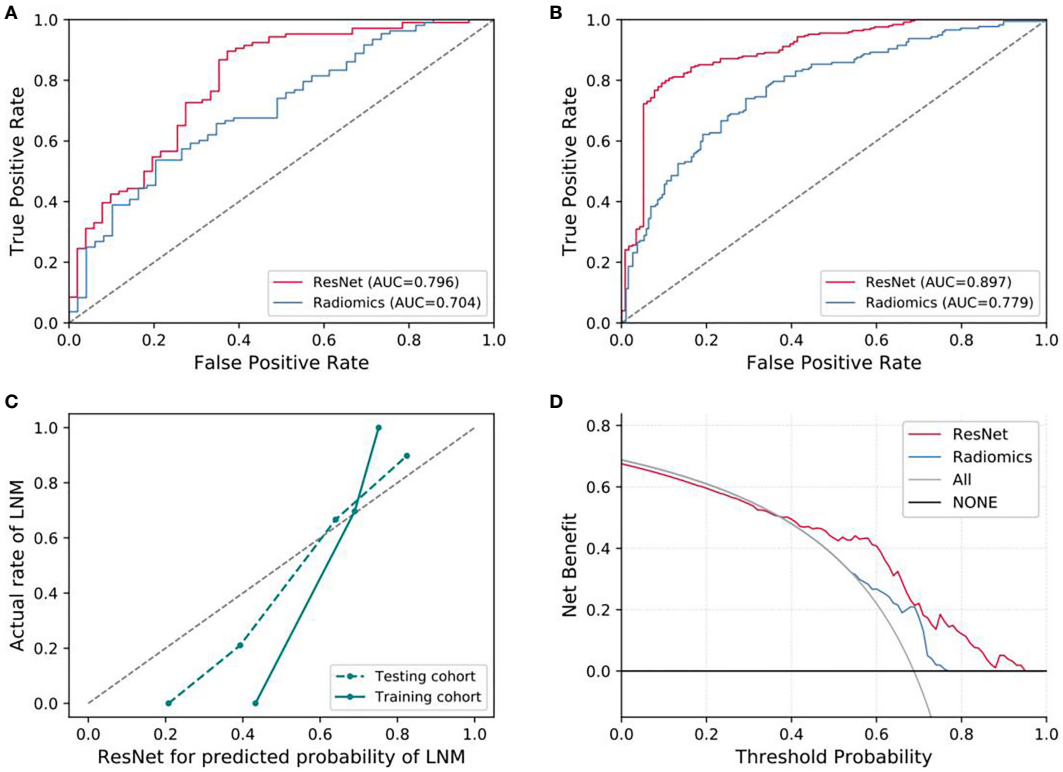


FIGURE 3
Evaluation of predictive performances for ResNet-SVM model and radiomics model. **(A)** The ROC curves showing the predictive performances of the ResNet and the radiomics model in testing cohorts. **(B)** The ROC curves showing the predictive performances of the ResNet and the radiomics model in training cohorts. **(C, D)** Curves of calibration analysis and the decision curve analysis for the ResNet and radiomics model. AUC, area under the receiver operating characteristic curve; LNM, lymph node metastasis.

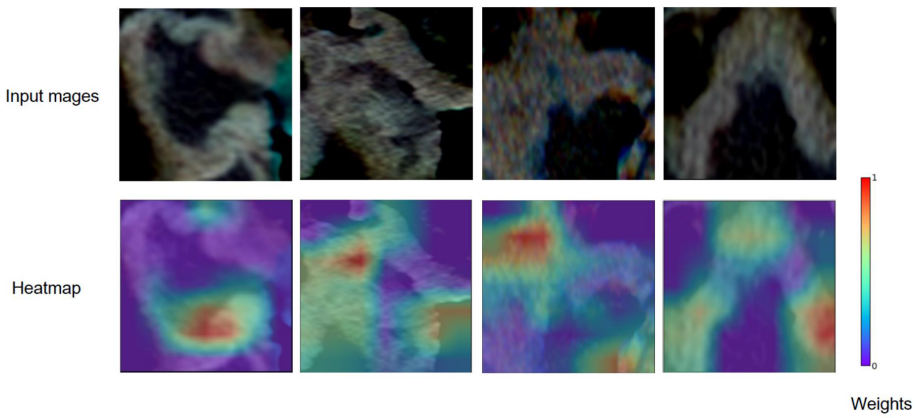


FIGURE 4
Grad-CAM visualizations for the feature heatmaps of representative patients generated from the ResNet. The right color bar indicates the scaled weights of deep features.

results are consistent with the above studies, and it is reasonable that GCs with greater tumor size tend to have a higher risk of lymph node metastasis.

As an emerging image quantification approach, radiomics has been widely used in diagnosis and prognosis of cancer patients based on medical images (5, 38, 39). Previous studies mainly focused on the characteristic manifestations of CT imaging to develop radiomic model, and did not use the transfer learning technology in the field of radiological prediction of LNM in gastric cancer. We established a CT-based model using the novel deep learning technique. Deep learning features extraction only needs to set a fixed size bounding box to tumor area, which not only improves the efficiency, but also reduces the subjectivity of manual segmentation in the radiomics procedure.

Deep learning technology has been widely used in the field of medical image processing. However, training a deep learning model from scratch is often not feasible because of various reasons: (1) the lack of a number of annotated images for one specific clinical problem. (2) reaching convergence could take too long for experiments to be worth. In the medical domain, using pre-trained CNNs as feature extractors is an effective way to alleviate these issues (19, 39–41). Transfer learning can transfer prior knowledge of image features and apply it to medical imaging with better generalization and ease of replication and testing. Our research shows that deep learning features extracted by transfer learning approach generalized well in medical tasks and achieved fairly good results. Moreover, the combination of radiomics and deep learning features did not improve the prediction performance in our study (Figures S2), which is similar to the results published by (40, 41). The reason is that the imaging features calculated from different frameworks might have different high-level dimensional characteristics, which are not suitable for feature combination.

Our study has some limitations that are worth noting. First, tumor regions of interest were manually delineated on CT images, which is high cost and laborious task. Semi-automatic or automatic segmentation method may be better. Second, although our experimental results showed good prediction performance, indicating that transfer learning could alleviate the domain difference, heterogeneity existed between various dataset. The main obstacle of this research is the lack of sufficient annotated medical images to further train the deep learning models. Such dataset could further extract more valuable features to improvement prediction performance. Third, we followed the IBSI benchmarks to filter the images after resampling, which may lead to the failure of estimating how much this would affect wavelet features to some extent. Last, this study is a single-center which is lack of external validation for the developed model, but we further randomly divided the patients into training or testing set and reconstructed and tested repeatedly ten times to evaluate the results. And, we are working to further access our model in a bigger dataset that may come from multiple centers.

Conclusion

In conclusion, our study adopted a noninvasive deep learning technique to perform prediction of LNM in GC. Compared to the handcrafted radiomics methods, the ResNet-SVM model could get better performance, and the implementation is simple and efficient without drawing the tumor contour manually. This study represented that the transfer learning strategy might also achieve good performance in medical imaging tasks without sufficient annotated medical images.

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. Requests to access these datasets should be directed to fccchengm@zzu.edu.cn.

Author contributions

H-PZ and MC: design the research. A-QZ: performed the research. A-QZ and H-PZ: collected the data. FL and PL: analyzed the data. A-QZ and MC: analyzed the data and wrote the paper. J-BG and MC: reviewed the paper. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Key Project of Science and Technology Research of Henan Province (No. 222102210112), the National Natural and Science Fund of China (No. 61802350, 81971615), National Key Research and Development Program of China (No. 2019YFC0118803).

Acknowledgments

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.969707/full#supplementary-material

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492

2. Shen L, Shan YS, Hu HM, Price TJ, Sirohi B, Yeh KH, et al. Management of gastric cancer in Asia: resource-stratified guidelines. *Lancet Oncol* (2013) 14(12): e535–47. doi: 10.1016/S1470-2045(13)70436-4

3. Smyth EC, Verheij M, Allum W, Cunningham D, Cervantes A, Arnold D. ESMO guidelines committee. gastric cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* (2016) 27(suppl 5):v38–49. doi: 10.1093/annonc/mdw350

4. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China. *CA: Cancer J Clin* (2016) 66(2):115–32. doi: 10.3322/caac.21338

5. Dong D, Fang MJ, Tang L, Shan XH, Gao JB, Giganti F, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Ann Oncol* (2020) 31 (7):912–20. doi: 10.1016/j.annonc.2020.04.003

6. Hartgrink HH, van de Velde CJ, Putter H, Bonenkamp JJ, Klein Kranenbarg E, Songun I, et al. Extended lymph node dissection for gastric cancer: who may benefit? final results of the randomized Dutch gastric cancer group trial. *J Clin Oncol* (2004) 22(11):2069–77. doi: 10.1200/JCO.2004.08.026

7. Amin MB, Edge SB, Greene FL, Schilsky RL, Washington ML, Sullivan DC, et al. AJCC cancer staging manual. *Basel*. (Switzerland: 8th ed: Springer) (2017).

8. Li J, Fang M, Wang R, Dong D, Tian J, Liang P, et al. Diagnostic accuracy of dual-energy CT-based nomograms to predict lymph node metastasis in gastric cancer. *Eur Radiol* (2018) 28(12):5241–9. doi: 10.1007/s00330-018-5483-2

9. Yamashita K, Hosoda K, Ema A, Watanabe M. Lymph node ratio as a novel and simple prognostic factor in advanced gastric cancer. *Eur J Surg Oncol* (2016) 42 (9):1253–60. doi: 10.1016/j.ejso.2016.03.001

10. Persiani R, Rausei S, Biondi A, Boccia S, Cananzi F, D'Ugo D. Ratio of metastatic lymph nodes: impact on staging and survival of gastric cancer. *Eur J Surg Oncol* (2008) 34(5):519–24. doi: 10.1016/j.ejso.2007.05.009

11. Oka S, Tanaka S, Kaneko I, Mouri R, Hirata M, Kawamura T, et al. Advantage of endoscopic submucosal dissection compared with EMR for early gastric cancer. *Gastrointest Endosc* (2006) 64(6):877–83. doi: 10.1016/j.gie.2006.03.932

12. Ajani JA, Bentrem DJ, Besh S, D'Amico TA, Das P, Denlinger C, et al. Gastric cancer, version 2.2013: featured updates to the NCCN guidelines. *J Natl Compr Canc Netw* (2013) 11(5):531–46. doi: 10.6004/jnccn.2013.0070

13. Hyung WJ, Cheong JH, Kim J, Chen J, Choi SH, Noh SH. Application of minimally invasive treatment for early gastric cancer. *J Surg Oncol* (2004) 85 (4):181–5. doi: 10.1002/jso.20018

14. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuzé S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol* (2017) 28(6):1191–206. doi: 10.1093/annonc/mdx034

15. Shen DG, Wu GR, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng* (2017) 19(1):221–48. doi: 10.1146/annurev-bioeng-071516-044442

16. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* (2018) 172(5):1122–31. doi: 10.1016/j.cell.2018.02.010

17. Li F, Chen H, Liu Z, Zhang X, Wu Z. Fully automated detection of retinal disorders by image-based deep learning. *Graefes Arch Clin Exp Ophthalmol* (2019) 257(3):495–505. doi: 10.1007/s00417-018-04224-8

18. Wakiya T, Ishido K, Kimura N, Nagase H, Kanda T, Ichiyama S, et al. CT-based deep learning enables early postoperative recurrence prediction for intrahepatic cholangiocarcinoma. *Sci Rep* (2022) 12(1):8428. doi: 10.1038/s41598-022-12604-8

19. Shin HC, Roth HR, Gao M, Le 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 Imaging* (2016) 35 (5):1285–98. doi: 10.1109/TMI.2016.2528162

20. Alzubaidi L, Fadhel MA, Al-Shamma O, Zhang JL, Santamaría J, Duan Y, et al. Towards a better understanding of transfer learning for medical imaging: a case study. *Appl Sci* (2020) 10(13):4523. doi: 10.3390/app10134523

21. Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vision* (2004) 60(2):91–110. doi: 10.1023/B:VISI.0000029664.99615.94

22. Raghu S, Sriraam N, Temel Y, Rao SV, Kubben PL. EEG Based multi-class seizure type classification using convolutional neural network and transfer learning. *Neural Netw* (2020) 124:202–12. doi: 10.1016/j.neunet.2020.01.017

23. van Rossum PSN, Xu C, Fried DV, Goense L, Court LE, Lin SH. The emerging field of radiomics in esophageal cancer: current evidence and future potential. *Transl Cancer Res* (2016) 5(4):410–23. doi: 10.21037/tcr.2016.06.19

24. Gu L, Liu Y, Guo X, Tian Y, Ye H, Zhou S, et al. Computed tomography-based radiomic analysis for prediction of treatment response to salvage chemoradiotherapy for locoregional lymph node recurrence after curative esophagectomy. *J Appl Clin Med Phys* (2021) 22(11):71–9. doi: 10.1002/acm2.13434

25. He K, Zhang X, Ren S, Sun J. (2016). Deep residual learning for image recognition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (Las Vegas, NV, USA: IEEE), June 27-30. pp. 770–8. doi: 10.1109/CVPR.2016.90

26. Simonyan K, Zisserman A. Very deep convolutional networks for Large-scale image recognition, in: 3rd International Conference on Learning Representations (ICLR), (2015) (San Diego, CA, USA: OpenReview.net), May 7-9, 2015. doi: 10.48550/arXiv.1409.1556

27. Chollet F. Xception: Deep learning with depth wise separable convolutions, (2017) in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (Honolulu, HI, USA: IEEE), July 21–26. doi: 10.1109/CVPR.2017.195

28. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. (2016). Rethinking the inception architecture for computer vision, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (Las Vegas, NV, USA: IEEE). pp. 2818–26. doi: 10.1109/CVPR.2016.308

29. Yang J, Shi R, Wei D, Liu Z, Zhao L, Ke B, et al. MedMNIST v2: A Large-scale lightweight benchmark for 2D and 3D biomedical image classification. *CoRR* (2021) abs/2110.14795. doi: 10.48550/arXiv.2110.14795

30. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks *via* gradient-based localization. *Int J Comput Vision* (2020) 128(2):336–59. doi: 10.1007/s11263-019-01228-7

31. Mackin D, Fave X, Zhang L, Yang J, Jones AK, Ng CS, et al. Correction: Harmonizing the pixel size in retrospective computed tomography radiomics studies. *PloS One* (2018) 13(1):e0191597. doi: 10.1371/journal.pone.0191597

32. Ligero M, Jordi-Ollero O, Bernatowicz K, Garcia-Ruiz A, Delgado-Muñoz E, Leiva D, et al. Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis. *Eur Radiol* (2021) 31(3):1460–70. doi: 10.1007/s00330-020-07174-0

33. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339

34. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* (2020) 295(2):328–38. doi: 10.1148/radiol.2020191145

35. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting CT radiomics. *Radiology* (2019) 291 (1):53–9. doi: 10.1148/radiol.2019182023

36. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics* (2018) 15(1):41–51. doi: 10.21873/cgp.20063

37. Li J, Dong D, Fang M, Wang R, Tian J, Li H, et al. Dual-energy CT-based deep learning radiomics can improve lymph node metastasis risk prediction for gastric cancer. *Eur Radiol* (2020) 30(4):2324–33. doi: 10.1007/s00330-019-06621-x

38. Wang S, Dong D, Zhang W, Hu H, Li H, Zhu Y, et al. Specific borrmann classification in advanced gastric cancer by an ensemble multilayer perceptron network: a multicenter research. *Med Phys* (2021) 48(9):5017–28. doi: 10.1002/mp.15094

39. Lopes UK, Valiati JF. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Comput Biol Med* (2017) 89:135–43. doi: 10.1016/j.compbiomed.2017.08.001

40. Yun J, Park JE, Lee H, Ham S, Kim N, Kim HS. Radiomic features and multilayer perceptron network classifier: a robust MRI classification strategy for distinguishing glioblastoma from primary central nervous system lymphoma. *Sci Rep* (2019) 9(1):5746. doi: 10.1038/s41598-019-42276-w

41. Hu Y, Xie C, Yang H, Ho JWK, Wen J, Han L, et al. Computed tomography-based deep-learning prediction of neoadjuvant chemoradiotherapy treatment response in esophageal squamous cell carcinoma. *Radiother Oncol* (2021) 154:6–13. doi: 10.1016/j.radonc.2020.09.014

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 13 October 2022
DOI 10.3389/fonc.2022.960178

OPEN ACCESS

EDITED BY
Shahid Mumtaz,
Instituto de Telecomunicações,
Portugal

REVIEWED BY
Zhongzhi Luan,
Beihang University, China
Attique Ur Rehman,
University of Sialkot, Pakistan

*CORRESPONDENCE
Dexing Kong
dxkong@zju.edu.cn

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

RECEIVED 02 June 2022
ACCEPTED 21 September 2022
PUBLISHED 13 October 2022

CITATION
Ling Y, Ying S, Xu L, Peng Z, Mao X, Chen Z, Ni J, Liu Q, Gong S and Kong D (2022) Automatic volumetric diagnosis of hepatocellular carcinoma based on four-phase CT scans with minimum extra information. *Front. Oncol.* 12:960178. doi: 10.3389/fonc.2022.960178

Automatic volumetric diagnosis of hepatocellular carcinoma based on four-phase CT scans with minimum extra information

Yating Ling[1], Shihong Ying[2], Lei Xu[3], Zhiyi Peng[2], Xiongwei Mao[4], Zhang Chen[5], Jing Ni[2], Qian Liu[1], Shaolin Gong[2] and Dexing Kong[1]*

[1]School of Mathematical Sciences, Zhejiang University, Hangzhou, China, [2]Department of Radiology, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China, [3]Computational Imaging and Digital Medicine, Zhejiang Qiushi Institute of Mathematical Medicine, Hangzhou, China, [4]Department of Radiology, The Hospital of Zhejiang University, Hangzhou, China, [5]Department of Radiology, Jiangcun Community Health Service Center, Hangzhou, China

Summary: We built a deep-learning based model for diagnosis of HCC with typical images from four-phase CT and MEI, demonstrating high performance and excellent efficiency.

Objectives: The aim of this study was to develop a deep-learning-based model for the diagnosis of hepatocellular carcinoma.

Materials and methods: This clinical retrospective study uses CT scans of liver tumors over four phases (non-enhanced phase, arterial phase, portal venous phase, and delayed phase). Tumors were diagnosed as hepatocellular carcinoma (HCC) and non-hepatocellular carcinoma (non-HCC) including cyst, hemangioma (HA), and intrahepatic cholangiocarcinoma (ICC). A total of 601 liver lesions from 479 patients (56 years ± 11 [standard deviation]; 350 men) are evaluated between 2014 and 2017 for a total of 315 HCCs and 286 non-HCCs including 64 cysts, 178 HAs, and 44 ICCs. A total of 481 liver lesions were randomly assigned to the training set, and the remaining 120 liver lesions constituted the validation set. A deep learning model using 3D convolutional neural network (CNN) and multilayer perceptron is trained based on CT scans and minimum extra information (MEI) including text input of patient age and gender as well as automatically extracted lesion location and size from image data. Fivefold cross-validations were performed using randomly split datasets. Diagnosis accuracy and efficiency of the trained model were compared with that of the radiologists using a validation set on which the model showed matched performance to the fivefold average. Student's *t*-test (T-test) of accuracy between the model and the two radiologists was performed.

Results: The accuracy for diagnosing HCCs of the proposed model was 94.17% (113 of 120), significantly higher than those of the radiologists, being 90.83% (109 of 120, *p*-value = 0.018) and 83.33% (100 of 120, *p*-value = 0.002). The average time analyzing each lesion by our proposed model on one Graphics

Processing Unit was 0.13 s, which was about 250 times faster than that of the two radiologists who needed, on average, 30 s and 37.5 s instead.

Conclusion: The proposed model trained on a few hundred samples with MEI demonstrates a diagnostic accuracy significantly higher than the two radiologists with a classification runtime about 250 times faster than that of the two radiologists and therefore could be easily incorporated into the clinical workflow to dramatically reduce the workload of radiologists.

KEYWORDS

computed tomography, diagnosis, hepatocellular carcinoma, deep learning, arificial intelligence

Highlights

1. The accuracy for diagnosing hepatocellular carcinomas of the proposed model and two radiologists was 94.17% (113 of 120), 90.83% (109 of 120, $p = 0.018$), and 83.33% (100 of 120, $p = 0.002$), showing significant differences.
2. The average time analyzing each lesion by our proposed model was 0.13 s, which was hundred times faster than the two radiologists.
3. The proposed model can serve as a quick and reliable "second opinion" for radiologists.

Introduction

Hepatocellular carcinoma (HCC) is the third most common malignancy worldwide, with incidence rates continuing to rise (1). CT slices often serve as an important assistive diagnostic tool for HCCs (2). According to the American Association for the Study of Liver Disease (AASLD) and the Liver Imaging Reporting and Data System (LI-RADS) reported by the American College of Radiology, the hallmark diagnostic characteristics of HCC on multi-phasic CT slices are arterial phase hyper-enhancement followed by washout appearance in the portal-venous and/or delayed phases (3, 4). Four-phase CT slices that contain non-enhanced, arterial, portal-venous, and delayed phases are recommended as the clinical standard. However, ensuring the diagnosis performance of a computer-aided system equivalent to that of radiologists with minimum extra information (MEI) about the patients for instance including only basic data about age and gender on a relatively small dataset based on four-phase CT images is still challenging in order to relieve the radiologists' workload as well as to improve the diagnosis throughout (5).

Machine learning algorithms have been widely applied in the radiological classification of various diseases and may potentially address this challenge (6–8). Recently, among different machine learnings, deep learning with convolutional neural network (CNN) have achieved state-of-the-art performances with respect to pattern recognition of images for various organs and tissues (9–15). It has been verified that CNN-based methods show high diagnostic performance in differentiation of tumors (16–20), but with most of them being limited to 2D slices, which needs manual selection. Meanwhile, it does not take advantage of 3D information that can potentially improve the diagnostic performances (21–25). Moreover, previous works (16, 17, 19, 25) for liver tumor diagnosis use three-phase CT slices, namely, non-enhanced phase, arterial phase, and transitional phase, which is between the portal-venous phase and the delay phase. However, hypointensity in the transitional phase does not qualify as "washout", which is considered a strong predictor and major criterion of HCC (3, 4). Therefore, in this study, we propose a 3D residual network (ResNet) as our basis network to explore the 3D structural information with four-phase CT images for tumor diagnosis (26).

Typically, high-performing CNN requires training on large datasets, which unfortunately are difficult to obtain especially in the medical field. As an alternative to large datasets, highly complicated clinical data collected from multi-modalities are incorporated to the CNN models (27, 28). Numerous works have discussed the auxiliary role of clinical data for HCC diagnosis, including, for example, alpha fetoprotein as a serological marker for HCCs since the 1960s (29), hepatitis B virus infection (30),

Abbreviations: HCC, hepatocellular carcinoma; HA, hemangioma; ICC, intrahepatic cholangiocarcinoma; CNN, convolutional neural network; ResNet, residual network; MExPaLe model, model fused with minimum extra information about patient and lesion; MEI, minimum extra information; T-test, Student's *t*-test.

and medical record of having non-alcoholic fatty liver diseases (31). However, those clinical data often require additional examinations. Therefore, it would be better if one only needs the patient's basic information, such as age and gender, which is crucial for liver tumor diagnosis (32–34) and makes full use of the spatial morphological information of local lesions that may be lost or downplayed in image processing.

In summary, our study aims to develop a fast-processing deep learning algorithm that exploits 3D structural with dynamic contrast information from four-phase CT scans and requires minimum patient information, i.e., age and gender, as well as automatically extracted lesion location and size from image data based on a relatively small dataset. We name the algorithm as the MExPaLe model (Model Fused with Minimum Extra Information about Patient and Lesion). The main contributions of this work are as follows:

- We propose a 3D model that feeds volumetric data as input instead of 2D CT slices to improve the diagnosis performance.
- We evaluate the diagnosis results of the basic model, which only uses non-enhanced phase CT images as input and enhanced model, which adds contrast-enhanced phase images as additional inputs. We experimentally confirm the necessity of using enhanced contrast agents in clinical workflow.
- The MExPaLe model fuses CNN and multilayer perceptron to incorporate two different modalities: image data and text data. The text data contain only information of patient gender and age, appended with the spatial morphological information of local lesions.
- The MExPaLe model demonstrates high performance and excellent efficiency. The accuracy and time efficiency for liver diagnosis of the proposed model are significantly higher than the two radiologists.

This paper is organized into four sections. In *Section 2*, we first describe the data collected in our paper, then introduce three models in this study, and finally the evaluation metrics have been presented. *Section 3* presents the results of our models and the comparison with other models and two radiologists. The discussion is provided in *Section 4*.

Materials and methods

This retrospective clinical study was approved by the review board, and the requirement for written informed consent was waived. Patients diagnosed as benign and HA through 1-year follow-up in 2018 while diagnosed as HCC and ICC after surgery or biopsy were enrolled between 2014 and 2017. Individuals without four-phase CT images were excluded, shown in Figure 1. Ultimately, a total of 601 lesions (315 HCCs) from 479 patients were selected. The details are presented in Table 1.

Data preprocessing

All CT slices were obtained with PHILIPS Brilliance iCT 256 scanner (Philips Healthcare, Netherlands). Contrast enhancement materials (Ultravist 300-3440, Bayer Schering Pharma AG, Germany) were injected. These four-phase CT images, stored as DICOM files, have a size of 512×512, and the thickness of each slice is 3 or 5 mm. The target lesions were manually labeled with 3D bounding boxes by a radiologist with 10 years of experience (XM) using software designed by Peng et al. (35) and revised if needed by a radiologist with 38 years of experience (ZY). The images were further processed by code written in the programming language Python 3.6 (https://www.python.org). We first reshaped the four-phase images to 1×1×1 mm using the cubic spline interpolation method and extracted the lesions and the surrounding 5-mm pixels by the bounding boxes. Then, the cropped 3D images were resized to a resolution of 64×64×64 voxels. The images were finally randomly selected to comprise the test data using fivefold cross-validation with the remaining images being the training data. Table 2 summarizes the distribution of each experiment.

The gender and age of the patients are the basic information recorded in the clinical system. Their contributions to HCC and non-HCC including benign, HA, and ICC diagnosis were evaluated in this study. In addition, the location and size of the lesions are inevitably lost during the common data preprocessing procedure. Therefore, we recorded the maximum normalized size and the relative location of the bounding box as our spatial morphological information during the data preprocessing. We also evaluated the contribution of spatial morphological information for HCC diagnosis.

Models

The model was built using Keras 2.2.4 (https://keras.io/) with a Tensorflow backend 1.5.0 (https://www.tensorflow.org/). For a baseline, we built a deep learning model based on the structure of 3D ResNet with 14 layers (13 convolutional layers and 1 global average pooling layer). Filter size of the first convolution layer is 5×5×5, and the following filter sizes are 3×3×3. The filter size of the global average pooling layer is 2×2×2. The basic model only uses non-enhanced phase CT images as input while the enhanced model adds contrast-enhanced phase images as additional input. For the basic and enhanced model, a fully connected layer is added following the 3D ResNet structure, whose output value represents the probability belonging to the corresponding class. The MExPaLe model contains two

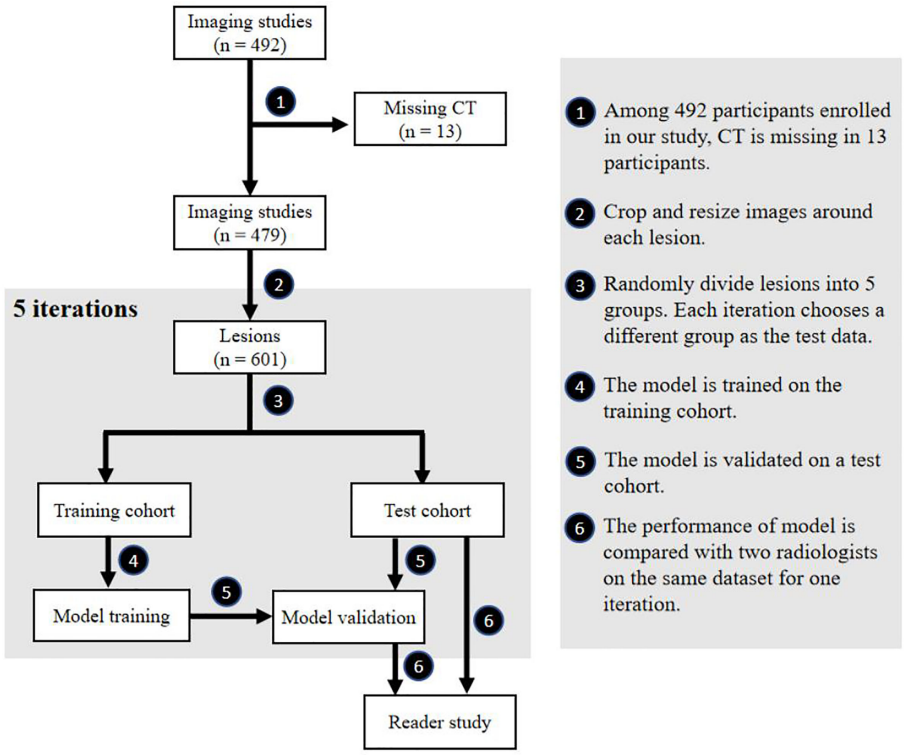


FIGURE 1
Flowchart of our study. Participant selection, model training, model testing and reader study are included in our study.

TABLE 1 Patient characteristics and demographics.

Patient characteristics	HCC	Cyst	HA	ICC	Total
Number of patients	312	37	107	41	479
Number of lesions	315	64	178	44	601
Age at imaging (mean ± std)	58 ± 11	58 ± 7	50 ± 10	59 ± 10	56 ± 11
Gender					
Male	268	28	41	29	350
Female	44	9	66	12	129

HCC, hepatocellular carcinoma; HA, hemangioma; ICC, intrahepatic cholangiocarcinoma; std, standard deviation.

TABLE 2 Distribution of the fivefold cross-validation dataset.

Experiment	E1	E2	E3	E4	E5
Training data	480	481	481	481	481
HCC	252	252	252	252	252
Non-HCC	228	229	229	229	229
Test data	121	120	120	120	120
HCC	63	63	63	63	63
Non-HCC	58	57	57	57	57
Total	601	601	601	601	601

HCC, hepatocellular carcinoma; E1–E5 denote five sets of experiments.

pathways: the CT pathway and the MEI pathway. The CT pathway has the same design as the aforementioned 3D ResNet structure but with the final classification layer removed. The MEI contains the patient age and gender exacted from the DICOM files and the relative size and location of lesions exacted from the CT pathway. MEI is text information; thus, we used a multilayer perceptron model containing two fully connected layers for this pathway. In our model, after the high-level features are flattened, image features and the text features are concatenated together. Finally, the concatenated feature vector is connected to a fully connected layer for final classification. The overview of the proposed method is shown in Figure 2.

All models use rectified linear units to help models learn non-linear features. These are used in conjunction with batch normalization and dropout to reduce overfitting. Each model was trained with a stochastic gradient descent optimizer using minibatches of eight samples. Each model was trained for 80 epochs. The training rate was initially set to 0.01, and it was reduced by half every 10 epochs.

The performance of the MExPaLe model was compared with two certified radiologists. The two radiologists (HY, with 21 years of imaging experience, and HC, with 16 years of imaging experience) did not take part in the data annotation process and were blinded to the lesion selection. For fair comparison and to simultaneously mimic the real working scenario as closely as possible, we provided four-phase CT DICOM data and the corresponding lesion 3D bounding boxes to both the MExPaLe model and radiologists. The test set for the reader study consisted of 120 randomly selected lesions in total (63 HCCs), while the remaining lesions were assigned to the training set. The time for the model from reading CT phases until classification of the lesion was recorded.

Statistics

Receiver operating characteristic (ROC) analyses were performed to calculate the area under curve (AUC) for evaluating model performance. The average accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for diagnosing each category were calculated. Student's *t*-test (T-test) using IBM SPSS Statistics 26.0 was also performed to evaluate the statistical significance of differences in comparative studies.

Results

Figure 1 shows the flowchart of our study, including participant selection, model training, model testing, and reader study. A total of 479 participants (350 men and 129 women) were enrolled in our study. The mean age ± standard deviation at enrollment was 56 years ± 11. Summaries of included participants are described in Table 1.

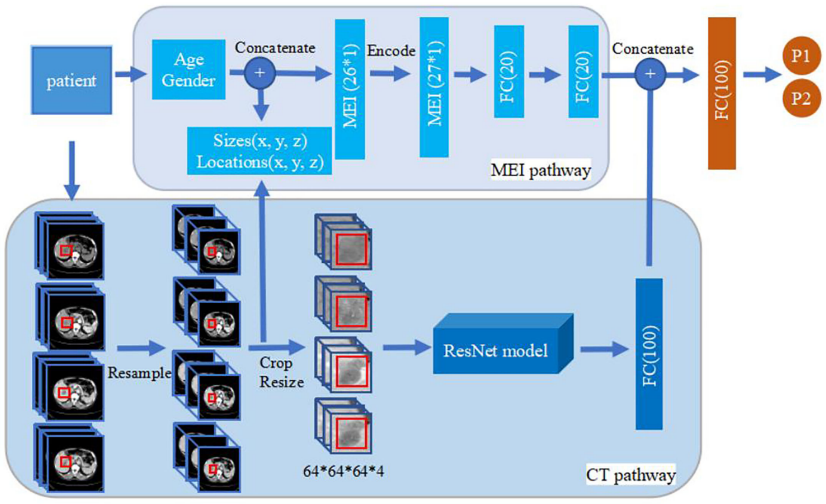


FIGURE 2
Overview of the proposed method. The upper part is MEI pathway and the lower part is the CT pathway. The 3D ResNet in CT pathway contains 14 layers (13 convolution layers, and 1 global average pooling layer). Filter size of the first convolution layer is 5×5×5, and the following filter sizes are 3×3×3. Filter size of the global average pooling layer is 2×2×2. The basic model and enhanced model only have the CT pathway. The size of image input in basic model is 64×64×64×1 while the others are 64×64×64×4. MEI, Minimum Extra Information.

Basic model and enhanced model

The diagnosis performances of the basic model and the enhanced model are shown in Table 3. Compared with the basic model, the enhanced model shows higher accuracy (17.30% higher in average, 91.68% *vs*. 74.38%, $p < 0.001$), AUC (18.47% higher in average, 95.79% *vs*. 77.32%, $p < 0.001$), sensitivity (12.06% higher in average, 94.60% *vs*. 82.54%, $p = 0.029$), specificity (23.03% higher in average, 88.45% *vs*. 65.42%, $p = 0.001$), PPV (17.34% higher in average, 90.03% *vs*. 72.69%, $p < 0.001$), and NPV (15.45% higher in average, 93.77% *vs*. 78.32%, $p = 0.008$).

The ROC curves of the basic and enhanced models with the corresponding AUC values are shown in Figure 3. The liver masses misdiagnosed by the basic model or enhanced model are shown in Figure 4. We present four-phase images of a 62-year-old man with a hemangioma and a 54-year-old man with an HCC. The major criterion of HCC such as "wash out" cannot be extracted by the model without the contrast-enhanced CT slices, which leads to the poor performance of the basic model.

MExPaLe model

In order to further improve the diagnosis, we first extracted the spatial morphological information of the local tumor during the data preprocessing process. Then, we added the patient's age and gender information, which were automatically recorded in the medical system. We finally compared the average diagnosis accuracy of models with different extra information, as shown in Figure 5. The average accuracy of the MExPaLe model was 94.18%, which was higher than that of the enhanced model (91.68%), the enhanced model with spatial morphological information

TABLE 3 Performance of basic model and enhanced model.

Parameter (%)	Basic model	Enhanced model	*p*-value*
Accuracy	74.38 (70.25–80.00)	91.68 (86.67–95.87)	**< 0.001**
AUC	77.32 (69.08–83.65)	95.79 (92.93–98.09)	**< 0.001**
Sensitivity	82.54 (69.84–95.24)	94.60 (90.47–100.00)	**0.029**
Specificity	65.42 (53.45–95.24)	88.45 (82.46–91.38)	**0.001**
PPV	72.69 (66.67–76.92)	90.03 (85.07–92.64)	**< 0.001**
NPV	78.32 (68.85–92.31)	93.77 (88.67–100.00)	**0.008**

AUC, area under curve; PPV, positive predictive value; NPV, negative predictive value. Data are median values in brackets and range in parentheses.
*p-value for differences between basic model and enhanced model, calculated with Student's t-test.
The bold values show the significant differences between basic model and enhanced model.

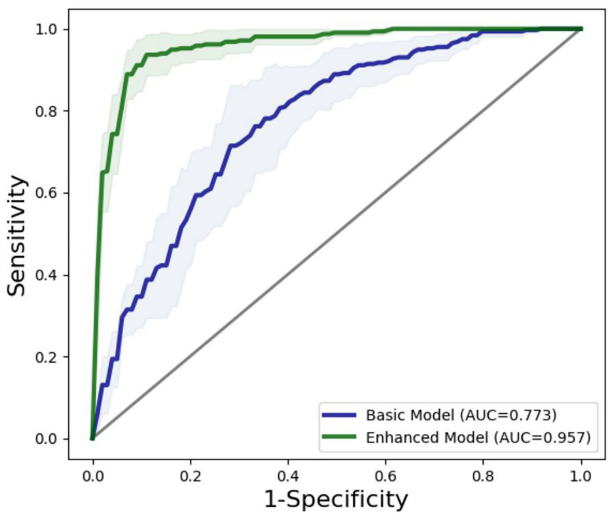


FIGURE 3
ROC curves of basic model and enhanced model. The lines reflect the average performances of the models, and the light-colored area reflects the fluctuation of the models represented by the corresponding standard deviations.

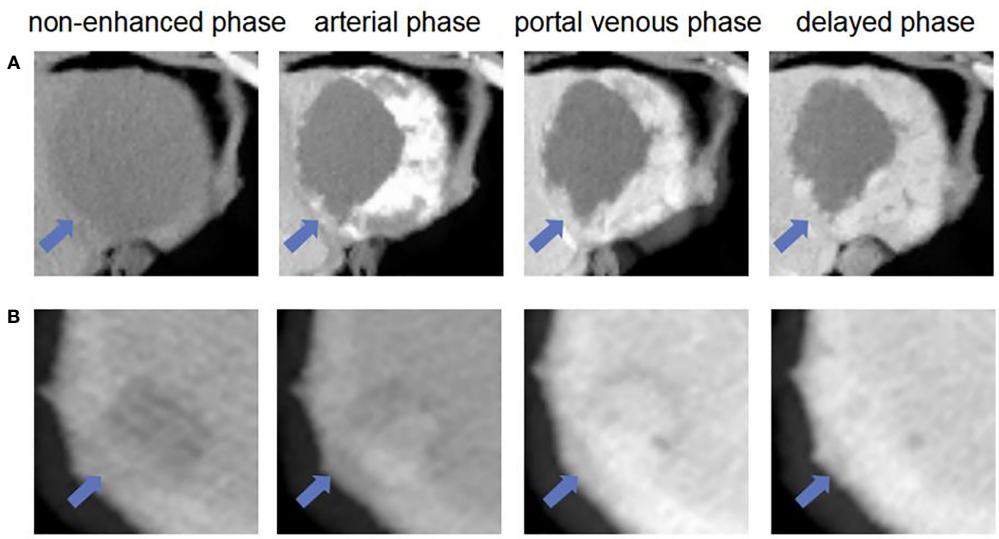


FIGURE 4
The liver masses misdiagnosed by models. **(A)** shows four phase images of a 62-year-old man with a hemangioma (arrow) that was diagnosed through one-year follow-up in 2018. The mass was correctly diagnosed as non-HCC by using enhanced model and our MExPale model. It was misdiagnosed as HCC by using basic model. **(B)** shows four phase images of 54-year-old man with a HCC (arrow) that was diagnosed after surgery. The mass was correctly diagnosed as HCC by using our MExPale model. It was misdiagnosed as HCC by using basic model and enhanced model.

(92.34%), the enhanced model with spatial morphological information and age (92.68%), and the enhanced model with spatial morphological information and sex (92.84%).

The diagnostic performance of the MExPaLe model compared with other authors is shown in Table 4. The MExPaLe model achieved an average accuracy of 94.18%, which was 4.99% higher than 2D CNN, 3.34% higher than 3D CNN, and 2.50% higher than 3D ResNet. Particularly, the MExPaLe model showed good performance in terms of specificity and NPV. The ROC curves of models are described in Figure 6A, and the confusion matrix of the MExPaLe model is described in Figure 6B. The MExPaLe model achieved an average AUC of 96.31%, which was 1.53% higher than 2D CNN, 0.31% higher than 3D CNN, and 0.52% higher than 3D

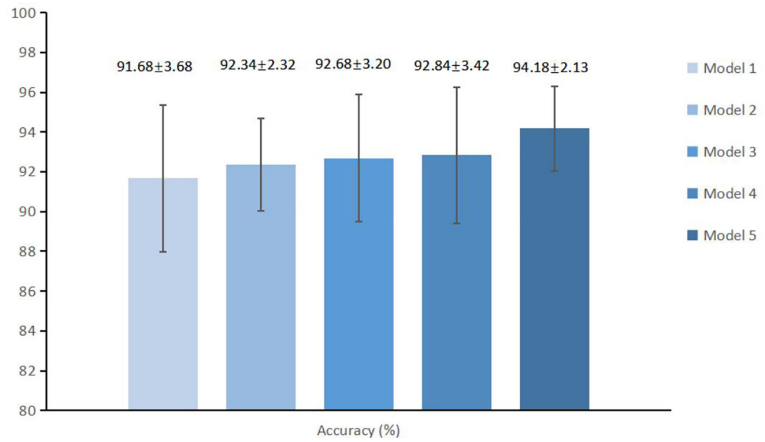


FIGURE 5
The average accuracy and standard deviations of different models. Model 1, Enhanced model; Model 2, Enhanced model with spatial morphological information; Model 3, Enhanced model with spatial morphological information and age; Model 4, Enhanced model with morphological information and gender; Model 5, MExPale model.

TABLE 4 Performance of models.

Parameter (%)	Accuracy	AUC	Sensitivity	Specificity	PPV	NPV
2D CNN (16)	89.19	94.78	89.21	89.17	90.11	88.30
3D CNN (16)	90.84	96.00	91.75	89.84	90.90	91.06
3D ResNet (25)	91.68	95.79	94.60	88.45	90.03	93.77
MExPaLe model	**94.18**	**96.31**	**98.10**	**89.85**	**91.45**	**97.70**

AUC, area under curve; PPV, positive predictive value; NPV, negative predictive value. The 3D CNN is generated from 2D CNN in (16).
The bold values show the best performance of models.

ResNet. The average ratio of true positive was 98.10%, and the average ratio of true negative was 89.85%.

Reader study

In the reader study, classification of 120 randomly selected lesions by the MExPaLe model achieved an accuracy of 94.17% (113/120). Diagnosis accuracies by radiologists from the First Affiliated Hospital of Zhejiang University (radiologist 1) and from the community primary hospital (radiologist 2) on the same lesions were 90.83% (109/120) and 83.33% (100/120), respectively (Table 5). We then randomly divided the lesions into five equal parts using T-test for statistical comparisons between the radiologists and our proposed MExPaLe model. The *p*-values comparing the MExPaLe model and radiologists 1 and 2 were 0.018 and 0.002, respectively, suggesting significant differences. The average runtime analyzing each lesion was 0.13 s for the MExPaLe model on one Graphics Processing Unit, while for the radiologists, on average 30 s and 37.5 s were needed. ROC curves of our MExPaLe model and two radiologists are shown in Figure 7. The misdiagnosed cases of the model and radiologists are described in Table 6. The coincidence degree between the MExPaLe model and radiologist 1 was 16.67% for HCC masses and 10.00% for non-HCC masses, while with radiologist 2, the coincidence degree was 25.00% for HCC masses and 22.22% for non-HCC masses. Our model showed a lower misdiagnosis rate for HCC masses compared with the two radiologists. Moreover, the performance of our model was more stable than those of the radiologists, with radiologist 1 showing high misdiagnosis for HCC masses and radiologist 2 showing high misdiagnosis for non-HCC masses. Some representative masses with varying diagnostic results from the MExPaLe model and the two radiologists are shown in Figure 8. As shown in Figure 8B, 71.43% (5/7) of the misdiagnosed cases by the model were ICC masses being misdiagnosed as HCC masses. This also constitutes the majority of misdiagnoses by the radiologists since it is hard to differentiate HCC from ICC especially owing to the low incidence rate of ICC. Therefore, by increasing the cases of ICC to balance the dataset, the model performance can be improved in the future.

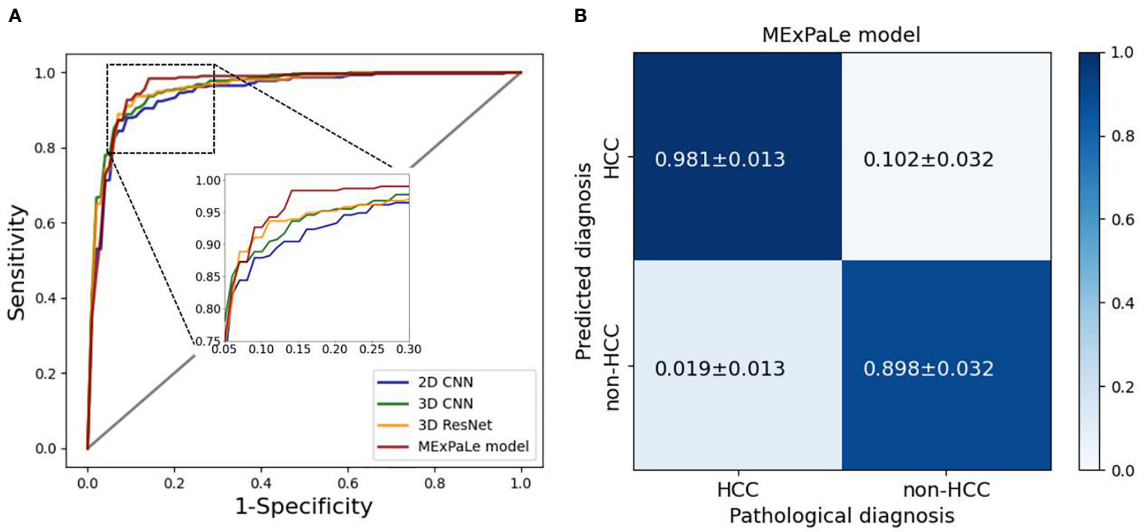


FIGURE 6
Performance of models. **(A)** ROC curves of models, **(B)** The confusion matrix of our MExPale model. HCC, hepatocellular carcinoma.

TABLE 5 Overall accuracy and times for model and radiologists' classification.

Parameter	MExPaLe model	Radiologist 1	*p*-value*	Radiologist 2	*p*-value*
Accuracy (%)	**94.17 (113/120)**	90.83 (109/120)	0.018	83.33 (100/120)	0.002
Time	**0.13 s**	30 s	–	37.5 s	–

Radiologist 1 comes from the First Affiliated Hospital of Zhejiang University, and Radiologist 2 comes from a community primary hospital.
*p-value for differences between the MExPaLe model and radiologists, calculated with Student's t-test.
The bold values show the best performance in terms of accuracy and time.

Discussion

In this work, we built a deep learning-based model, MExPaLe, for the diagnosis of liver tumor with typical images from four-phase CT and MEI, demonstrating high performance and excellent efficiency. The accuracy for diagnosing liver tumors of the proposed model and the two radiologists were 94.17% (113 of 120), 90.83% (109 of 120, $p = 0.018$), and 83.33% (100 of 120, $p = 0.002$), showing significant differences. The average time analyzing each lesion by our proposed MExPaLe model was 0.13 s, which was close to 250 times faster than that of both radiologists.

We used volumetric 3D CT patches as inputs. The 3D model can provide more relevant information to lesion classification, minimizing model variability, and it was not dependent on manual slice selection. Concerns for using the 3D model may involve possible expensive computational cost and time consumption. However, by focusing on local liver lesions and a relatively shallow model structure, we achieved sub-second runtime per case, taking four-phase CT volumetric scans as input, and therefore, it no longer becomes a practical obstacle.

In real clinical conditions, critical diagnostic features, such as hyper-enhancement and washout, are the main features used by radiologists. These features are obtained through the comparison of multi-phase CT images, necessitating the use of enhanced contrast agents to improve the diagnosis accuracy. This is also verified by our results obtained from the basic model and enhanced model, which had a median accuracy of 74.38% (range, 70.25%–80.00%) and 91.69% (range, 86.67%–95.87%), respectively, and by the statistical test.

Many works have confirmed that clinical data about the patients can improve the performance of diagnosis. However, the clinical data used in those works are often too complicated to obtain, and their processing requires additional manpower and material resources. More importantly, some clinical data can be inaccurate at the time of collection, such as family genetic history. Instead, our experiment requires only the basic information of the patient, i.e., age and gender, and minimal spatial morphological information lost during image preprocessing, which does not increase the clinical workload; therefore, it is of high practical value to be used in the clinics. The proposed MExPaLe model showed a median accuracy of 94.18% (range, 91.67%–96.67%) and a median AUC of 96.31% (range, 93.34%–98.22%). The MExPaLe model showed high specificity and NPV, attributed to the usefulness of the MEI in

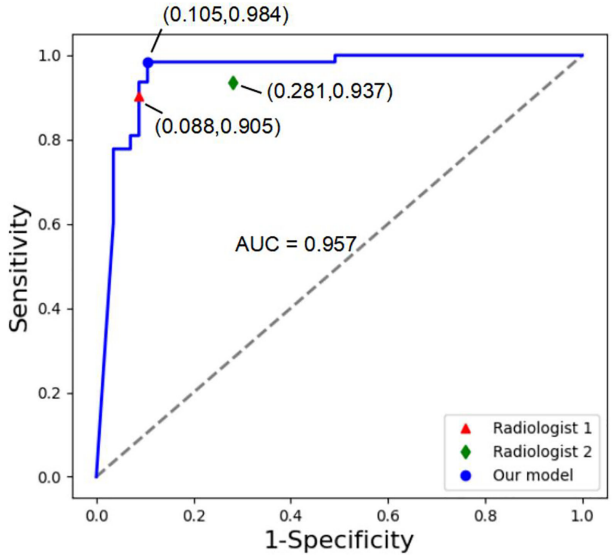


FIGURE 7
ROC curves of our MExPale model and two radiologists. Radiologist 1 comes from the First Affiliated Hospital of Zhejiang University, and Radiologist 2 comes from a community primary hospital.

TABLE 6 Misdiagnosed images for model and radiologists' classification.

Parameter	MExPaLe model	Radiologist 1	Radiologist 2
Misdiagnoses			
HCC	**1**	6	4
Non-HCC	**6**	5	16
Coincidence degree			
HCC	-	16.67% (1/**6**)	25.00% (1/**4**)
Non-HCC	–	10.00% (1/**10**)	22.22% (4/**6**)

Radiologist 1 comes from the First Affiliated Hospital of Zhejiang University, and Radiologist 2 comes from a community primary hospital. HCC, hepatocellular carcinoma. The bold values mean the number of misdiagnosed masses for our model classification.

predicting liver tumor, which made the MExPaLe model more effective than others.

Furthermore, the MExPaLe model differs from previous works in that it does not require complex-shaped ROI tracing boundaries of tumors. The location and size of a 3D bounding box around the target lesion are enough in our work. We included 5-mm extra pixels surrounding the lesions to learn more peri-tumoral information, which is necessary for enhancing tumor differentiation. Additionally, it can reduce the possible subjective bias in the image capture process and maintain tumor size information to a certain extent.

The direct comparison between the MExPaLe model and the two radiologists suggests that the MExPaLe model can serve as a reliable and quick "second opinion" for radiologists. In the diagnosis of HCCs, the accuracy of the MExPaLe model was higher than that of the chief radiologist at a first-tier research hospital and the radiologist from a community primary hospital, both with statistical significances. Furthermore, the runtime of the MExPaLe model per case for liver tumor diagnosis was close to 250 times faster compared with the radiologists, suggesting that the use of the MExPaLe model can greatly improve the diagnosis throughput in the clinics.

While these results are promising, several limitations should be acknowledged regarding this study. Because of the limited number of imaging studies, we were restricted to a cross-validation experimental design. It would be better if we can incorporate an additional test dataset, and ideally an external dataset to consolidate the usefulness of our model in the clinical diagnosis of HCCs. Another limitation is that only four typical primary liver cancer types were available with the exclusion of other relevant cancers types including metastatic liver cancers.

In conclusion, we proposed a model for the diagnosis of liver tumor. The MExPaLe model, which has incorporated four-phase CT volumes and the MEI, achieves the highest prediction accuracy of 94.18% (range, 91.67%–96.67%) and an AUC of 96.31% (range, 93.34%–98.22%). It is superior to both the basic model and the

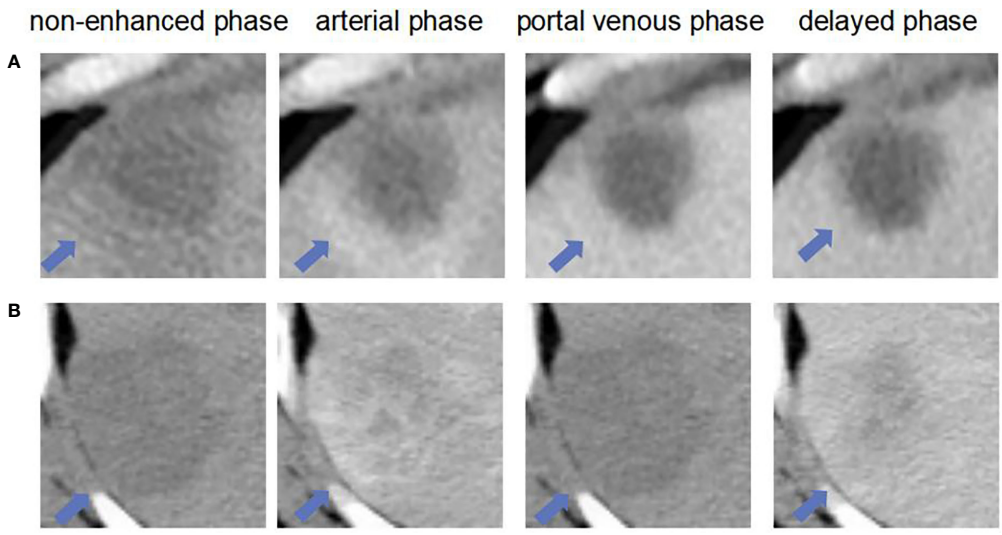


FIGURE 8
The liver masses misdiagnosed by model and two radiologists. **(A)** shows four phase images of a 59-year-old man with a HCC (arrow) that was diagnosed after surgery. The mass was misdiagnosed diagnosed as non-HCC by and our MExPale model and both two radiologists. **(B)** shows four images of a 64-year-old man with a ICC (arrow) that was diagnosed after surgery. The mass was misdiagnosed diagnosed as HCC by our MExPale model and both two radiologists.

enhanced model. It is about 250 times more time-efficient compared with the radiologists for liver tumor diagnosis, taking only 0.13 s. The architectural design of the MExPaLe model may be applicable to more multi-phase CT-based diagnosis projects to provide high-quality patient care in a time-efficient manner.

Data availability statement

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

Ethics statement

This study was reviewed and approved by Clinical Research Ethics Committee of the First Affiliated Hospital, Zhejiang University School of Medicine. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

Literature research, YL. Project supervision, DK. Data annotation, ZP and XM. Experiment, YL. Clinical studies, YL, LX, ZP, XM, SG, ZC, JN, QL and SY. Data analysis, YL, LX, ZC, JN, QL and SY. Statistical analysis, YL. Manuscript writing, YL. Manuscript revision, LX, YL contributed equally to this work with SY. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China, Grant Nos. 12090020 and 12090025.

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

References

1. Rebecca LS, Kimberly DM, Ahmedin J. Tumor statistics 2019. *CA Cancer J Clin* (2019) 69:7–34. doi: 10.3322/caac.21551

2. Freddie B, Jacques F, Isabelle S, Rebecca LS, Lindsey AT, Ahmedin J. Global tumor statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 tumors in 185 countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492

3. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American association for the study of liver diseases. *Hepatol* (2018) 68:723–50. doi: 10.1002/hep.29913

4. Tang A, Bashir MR, Corwin MT, Cruite I, Dietrich CF, Do RKG, et al. Evidence supporting LI-RADS major features for CT- and MR imaging–based diagnosis of hepatocellular carcinoma: A systematic review. *Radiology* (2018) 286:29–48. doi: 10.1148/radiol.2017170554

5. Ayuso C, Rimola J, Vilana R, Burrel M, Darnell A, García-Criado Á, et al. Diagnosis and staging of hepatocellular carcinoma (HCC): current guidelines. *Eur J Radiol* (2018) 101:72–81. doi: 10.1016/j.ejrad.2018.01.025

6. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169

7. Acharya UR, Koh JE, Hagiwara Y, Tan JH, Gertych A, Vijayananthan A, et al. Automated diagnosis of focal liver lesions using bidirectional empirical mode decomposition features. *Comput Biol Med* (2018) 94:11–8. doi: 10.1016/j.compbiomed.2017.12.024

8. Xu X, Zhang H-L, Liu Q-P, Sun S-W, Zhang J, Zhu F-P, et al. Radiomic analysis of contrast-enhanced ct predicts microvascular invasion and outcome in hepatocellular carcinoma. *Hepatol* (2019) 70(6):1133–44. doi: 10.1016/j.jhep.2019.02.023

9. Krizhevsky A, Sutskever I, Hinton G. *ImageNet classification with deep convolutional neural networks*. Communications of the ACM (2017). New York, NY, United States: IEEE (2017) 60(6): 84–90. doi: 10.1145/3065386

10. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015) 521(7553):436–44. doi: 10.1038/nature14539

11. Aboutalib SS, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu SD. Deep learning to distinguish recalled but benign mammography images in breast cancer screening. *Clin Cancer Res* (2018) 24:5902–9. doi: 10.1158/1078-0432.CCR-18-1115

12. Xi IL, Zhao Y, Wang R, Chang M, Purkayastha S, Chang K, et al. Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging. *Clin Cancer Res* (2020) 26:1944–52. doi: 10.1158/1078-0432.CCR-19-0374

13. Ozdemir O, Russell RL, Berlin AA. A 3D probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose CT scans. *IEEE Trans Med Imaging* (2019) 39(5):1419–29. doi: 10.1109/TMI.2019.2947595

14. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA-Journal Am Med Assoc* (2016) 316:2402–10. doi: 10.1001/jama.2016.17216

15. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* (2017) 542:115–+. doi: 10.1038/nature21056

16. Koichiro Y, Hiroyuki A, Osamu A, Shigeru K. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: A preliminary study. *Radiol* (2018) 286:887–96. doi: 10.1148/radiol.2017170706

17. Todoroki Y, Iwamoto Y, Lin L, Hu H, Chen Y-W. Automatic detection of focal liver lesions in multi-phase CT images using a multi-channel & multi-scale CNN, conference proceedings: Annual international conference of the IEEE engineering in medicine and biology society. *IEEE Eng Med Biol Soc Annu Conf* (2019) 2019:872–5. doi: 10.1109/EMBC.2019.8857292

18. Frid-Adar M, Klang E, Amitai M, Goldberger J, Greenspan H. *Synthetic data augmentation using GAN for improved liver lesion classification*. In: 2018 IEEE 15th international symposium on biomedical imaging. Washington, DC, USA: IEEE (2018). p. 289–93. doi: 10.1109/ISBI.2018.8363576

19. Liang D, Lin LF, Hu HJ, Zhang QW, Chen QQ, Iwamoto Y, et al. Combining convolutional and recurrent neural networks for classification of focal liver lesions in multi-phase CT images. In: *Med image computing comput assisted intervention - miccai 2018*, vol. Pt Ii. (2018). Springer, Cham p. 666–75. doi: 10.1007/978-3-030-00934-2_74

20. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med* (2018) 15:e1002686. doi: 10.1371/journal.pmed.1002686

21. Hamm CA, Wang CJ, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. *Eur Radiol* (2019) 29:3338–47. doi: 10.1007/s00330-019-06205-9

22. Song Y, Yu Z, Zhou T, Teoh JYC, Lei B, Choi KS, et al. Learning 3d features with 2d cnns via surface projection for ct volume segmentation. In: *MICCAI*. Springer, Cham (2020). p. 176–86. doi: 10.1007/978-3-030-59719-1_18

23. Carreira J, Zisserman A. Quo vadis, action recognition? A new model and the kinetics dataset, In: *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA: IEEE (2017). p. 6299–308. doi: 10.1109/CVPR.2017.502

24. Hara K, Kataoka H, Satoh Y. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In: 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE (2018). p. 6546–55. doi: 10.1109/CVPR.2018.00685

25. Zhou J, Wang W, Lei B, Ge W, Huang Y, Zhang L, et al. Automatic detection and classification of focal liver lesions based on deep convolutional neural networks: a preliminary study. *Front Oncol* (2021) 10:581210. doi: 10.3389/fonc.2020.581210

26. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE (2016). p. 770–8. doi: 10.1109/CVPR.2016.90

27. Liu R, Pan D, Xu Y, Zeng H, He Z, Lin J, et al. A deep learning–machine learning fusion approach for the classification of benign, malignant, and intermediate bone tumors. *Eur Radiol* (2021) 32:1371–83. doi: 10.1007/s00330-021-08195-z

28. Gao R, Zhao S, Aishanjiang K, Cai H, Wei T, Zhang Y, et al. Deep learning for differential diagnosis of malignant hepatic tumors based on multi-phase contrast-enhanced CT and clinical data. *J Hematol Oncol* (2021) 14(1):1–7. doi: 10.1186/s13045-021-01167-2

29. Johnson P. Role of alpha-fetoprotein in the diagnosis and management of hepatocellular carcinoma. *Gastroenterol Hepatol* (1999) 14:32–6. doi: 10.1046/j.1440-1746.1999.01873.x

30. El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterol* (2012) 142:1264–73. doi: 10.1053/j.gastro.2011.12.061

31. Kanwal F, Kramer JR, Mapakshi S, Natarajan Y, Chayanupatkul M, Richardson PA, et al. Risk of hepatocellular cancer in patients with non-alcoholic fatty liver disease. *Gastroenterol* (2018) 155:1828–37. doi: 10.1053/j.gastro.2018.08.024

32. Bosch FX, Ribes J, Díaz M, Cléries R. Primary liver cancer: worldwide incidence and trends. *Gastroenterol* (2004) 127:5–16. doi: 10.1053/j.gastro.2004.09.011

33. Wu EM, Wong LL, Hernandez BY, Ji J-F, Jia W, Kwee SA, et al. Gender differences in hepatocellular cancer: disparities in nonalcoholic fatty liver disease/steatohepatitis and liver transplantation. *Hepatoma Res* (2018) 4:66. doi: 10.20517/2394-5079.2018.87

34. Janevska D, Chaloska-Ivanova V, Janevski V. Hepatocellular carcinoma: risk factors, diagnosis and treatment. *Open Access Maced J Med Sci* (2015) 3:732. doi: 10.3889/oamjms.2015.111

35. Peng J, Wang J, Kong D. A new convex variational model for liver segmentation. *Proceedings of the 21st International Conference on Pattern Recognition*. Tsukuba, Japan: IEEE (2012).

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 31 October 2022
DOI 10.3389/fonc.2022.925903

OPEN ACCESS

EDITED BY
Wei Wei,
Xi'an University of Technology, China

REVIEWED BY
Michal Kawulok,
Silesian University of Technology, Poland
Leonhard Müllauer,
Medical University of Vienna, Austria
Philipp Stroebel,
University Medical Center Göttingen, Germany

*CORRESPONDENCE
Pengyu Wang
Y10180292@mail.ecust.edu.cn
Dingrong Zhong
748803069@qq.com
Jie Liu
20112040@bjtu.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 22 April 2022
ACCEPTED 11 October 2022
PUBLISHED 31 October 2022

CITATION
Zhang H, Chen H, Qin J, Wang B, Ma G, Wang P, Zhong D and Liu J (2022) MC-ViT: Multi-path cross-scale vision transformer for thymoma histopathology whole slide image typing.
Front. Oncol. 12:925903.
doi: 10.3389/fonc.2022.925903

MC-ViT: Multi-path cross-scale vision transformer for thymoma histopathology whole slide image typing

Huaqi Zhang[1†], Huang Chen[2†], Jin Qin[1], Bei Wang[2], Guolin Ma[3], Pengyu Wang[4*], Dingrong Zhong[2*] and Jie Liu[1*]

[1]School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China, [2]Department of Pathology, China-Japan Friendship Hospital, Beijing, China, [3]Department of Radiology, China-Japan Friendship Hospital, Beijing, China, [4]School of Information Science and Engineering, East China University of Science and Technology, Shanghai, China

Objectives: Accurate histological typing plays an important role in diagnosing thymoma or thymic carcinoma (TC) and predicting the corresponding prognosis. In this paper, we develop and validate a deep learning-based thymoma typing method for hematoxylin & eosin (H&E)-stained whole slide images (WSIs), which provides useful histopathology information from patients to assist doctors for better diagnosing thymoma or TC.

Methods: We propose a multi-path cross-scale vision transformer (MC-ViT), which first uses the cross attentive scale-aware transformer (CAST) to classify the pathological information related to thymoma, and then uses such pathological information priors to assist the WSIs transformer (WT) for thymoma typing. To make full use of the multi-scale (10×, 20×, and 40×) information inherent in a WSI, CAST not only employs parallel multi-path to capture different receptive field features from multi-scale WSI inputs, but also introduces the cross-correlation attention module (CAM) to aggregate multi-scale features to achieve cross-scale spatial information complementarity. After that, WT can effectively convert full-scale WSIs into 1D feature matrices with pathological information labels to improve the efficiency and accuracy of thymoma typing.

Results: We construct a large-scale thymoma histopathology WSI (THW) dataset and annotate corresponding pathological information and thymoma typing labels. The proposed MC-ViT achieves the Top-1 accuracy of 0.939 and 0.951 in pathological information classification and thymoma typing, respectively. Moreover, the quantitative and statistical experiments on the THW dataset also demonstrate that our pipeline performs favorably against the existing classical convolutional neural networks, vision transformers, and deep learning-based medical image classification methods.

Conclusion: This paper demonstrates that comprehensively utilizing the pathological information contained in multi-scale WSIs is feasible for thymoma typing and achieves clinically acceptable performance. Specifically,

the proposed MC-ViT can well predict pathological information classes as well as thymoma types, which show the application potential to the diagnosis of thymoma and TC and may assist doctors in improving diagnosis efficiency and accuracy.

KEYWORDS

thymoma typing, histopathology whole slide image, vision transformer, cross-correlation attention, multi-scale feature fusion

Introduction

Thymic epithelial tumors (i.e., thymomas) are uncommon and primary anterior mediastinum neoplasms derived from the thymic epithelium. According to the histological classification standard, the World Health Organization (WHO) distinguishes thymomas (types A, AB, B1, B1+B2, B2, B2+B3, and B3) from thymic carcinoma (TC) (1, 2). Considering that thymoma may gradually develop into TC, thymoma typing is crucial to assist doctors in diagnosis and prognosis (3). The morphological diagnosis of thymoma has traditionally posed difficulties for histopathologists since thymoma has great histological variability and intratumoral heterogeneity (4, 5), and it is difficult to conceptualize a cogent and easily reproducible morphological classification standard. Currently, based on the schema of WHO, the morphological classification of thymic epithelial neoplasms is described as follows: Type A thymoma usually consists of the spindle or ovoid-shaped cells with bland nuclei, scattered chromatin, and inconspicuous nucleoli arranged in solid sheets with few or no lymphocytes in the tumor. By comparison, type B thymoma may display coarse lobulation delineated by fibrous septa. Type B1 thymoma contains dense lymphocyte neoplastic with scant neoplastic epithelial cells, which are composed of oval cells with pale round nuclei and small nucleoli. In type B2 thymoma, the neoplastic thymic epithelial cells are increased in number and appear as scattered plump cells among equivalent mixed lymphocytes. The epithelial cells are large and polygonal, which have obvious vesicular nuclei and central prominent nucleoli, and show a tendency to palisade around vessels and fibrous septa. Here, dilated perivascular spaces are commonly existed. Type B3 thymoma corresponds to the lobular growth pattern of a smoothly contoured tumor composed predominantly of epithelial cells having a round or polygonal shape and clear cytoplasm. Note that perivascular spaces with epithelial palisading are prominent, and lymphocytes are almost always interspersed among the tumor cells. In addition, type AB thymoma has features of type A thymoma that are admixed with foci showing features of type B thymoma. TC exhibits clear-cut cytological atypia and a set of cytoarchitectural features no longer specific to the thymus (6).

At present, the diagnosis of thymoma and TC basically relies on the visual observation of WSIs by histopathologists. With the rapid development of deep learning technology, we aim to develop a computer-assisted diagnosis (CAD) system to provide doctors with more histopathological information to assist the diagnosis and prognosis. More specifically, we can achieve the initial screening of WSIs through an efficient CAD system (7–10) to assist doctors in obtaining the detailed thymoma pathological information and the accurate thymoma typing results. Over the past few years, convolutional neural networks (CNNs) have shown excellent performance in most computer vision tasks including medical image processing. However, many studies (11–13) have gradually discovered some inherent limitations of CNNs, such as the difficulty in modeling long-range dependencies and the local receptive field. To better modeling global feature relations, some scholars extend the transformer from the natural language processing field to the computer vision field, and then propose high-performance vision transformers (ViTs) including Swin-T (12), PVT (13), LeViT (14), TNT (15), T2T-ViT (16), IPT (17), and Uformer (18) to serve various high-level and low-level vision tasks. In addition, there are also some ViT variants developed to achieve medical image processing, such as GasHis-ViT (19) for histopathology image normal and abnormal classification, and Swin-Unet (20) and AFTer-Unet (21) for multi-organ CT image segmentation. However, in digital pathology workflow, existing ViTs are difficult to effectively utilize for thymoma histopathology WSI typing due to the following two problems (1): Affected by the implementation mechanism of multi-head self-attention (MSA), current ViTs usually have large computational costs; thus, it is unsuitable to directly process the full-scale WSI with millions of resolutions (2). Although many existing ViTs can effectively model global and local feature relations, most of them fail to employ the complementary between multi-scale or multi-resolution features. Considering that thymoma histopathology WSIs have the inherent multi-scale information, for example, a WSI

includes three magnification versions in terms of 10 × , 20 × , and 40 × . Moreover, the local pathological information of a WSI has close correspondences with the thymoma type. Therefore, we can address such problems by comprehensively employing the above-mentioned two types of information to design ViT.

In this paper, we propose a multi-path cross-scale vision transformer (MC-ViT) to achieve thymoma histopathology WSI typing. MC-ViT contains two core components, the first one named cross attentive scale-aware transformer (CAST), which takes the multi-scale patches from the same WSI as inputs and then predicts corresponding pathological information classes (spindle thymic epithelial cells, B1 thymic epithelial cells, B2 thymic epithelial cells, B3 thymic epithelial cells, fibrous septa, erythrocyte, lymphocyte, perivascular space, medullary differentiated areas, and tumor) to serve thymoma typing. Unlike the standard ViT (11), the proposed CAST constructs multiple paths to separately process 10 × , 20 ×, and 40 × WSI patches for capturing potential pathological information in different receptive field features. In general, 10 × WSI patches contain more information about the medullary differentiated areas and fibrous septa, 20 × WSI patches are mainly related to the perivascular space and lymphocyte, and 40 × WSI patches can better reflect the properties of the erythrocyte and thymic epithelial cells. To comprehensively utilize such pathological information, we also propose a cross-correlation attention module (CAM) to fuse multi-scale features in the main path of CAST. The second component is the WSIs transformer (WT), which is designed to classify the thymoma type of WSIs. Here, we propose to use the fixed number of multi-scale WSI patches to represent a full-scale WSI, and introduce the pathological information labels of these WSI patches as priors to improve the interpretability and accuracy of thymoma typing. Specifically, we concatenate the low-level features of multi-scale WSI patches and corresponding pathological information labels to form a 1D feature matrix as the input, and then predict the thymoma type (A, AB, B1, B1+B2, B2, B2+B3, B3, or C) by WT. Based on this design, we achieve 95.1% thymoma typing accuracy using a lightweight model with only a three-stage transformer encoder. Finally, this paper constructs a large-scale thymoma histopathology WSI (THW) dataset, which contains 129 hematoxylin & eosin (H&E)-stained WSIs with the pathological information and thymoma typing annotations.

The thymoma diagnosis workflow is illustrated in Figure 1, and the main contributions can be summarized as follows:

- We propose an MC-ViT, which is the first transformer architecture designed for thymoma histopathology WSI typing.
- We develop a CAST with a cross-correlation attention mechanism, which can fully leverage the multi-scale information inherent in WSIs to achieve pathological information classification.
- We achieve the end-to-end thymoma histopathology WSI typing. The proposed WSIs transformer takes pathological information labels as priors to convert a WSI into a 1D feature matrix as the network input,

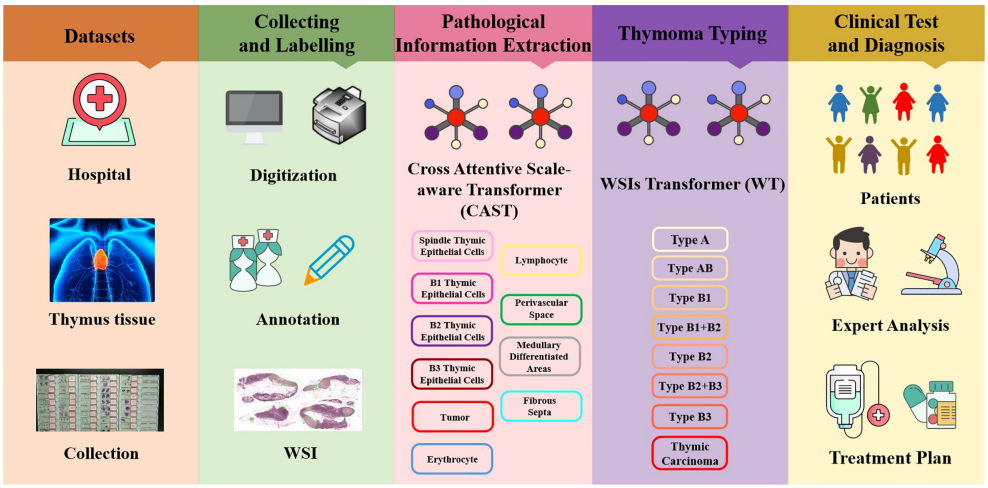


FIGURE 1
Illustration of the thymoma diagnosis workflow. Firstly, we collect the clinical data from the China–Japan Friendship Hospital to construct the thymoma histopathology WSI (THW) dataset. Then, histopathologists are invited to manually label the WSIs of the THW dataset as eight thymoma types with 10 classes of local pathological information. Next, we propose the cross attentive scale-aware transformer (CAST) for pathological information classification, which can guide the WSI transformer (WT) to achieve accurate thymoma histopathology WSI typing. Finally, according to the predicted results of the network, doctors can more efficiently and accurately diagnose thymoma and TC.

which solves the computing complexity problem caused by full-scale WSI.
- We publish a large-scale thymoma histopathology WSI dataset with 323 H&E-stained WSIs from 129 patients and annotate the pathological information classes and thymoma types.

Related works

Vision transformers

Starting with AlexNet (22), deep CNNs serve as the mainstream backbone networks in computer vision for many years. However, many studies (11–13) point out that CNNs are unsuitable to model long-range dependencies in the data. Recently, with the development of the non-local self-attention, the transformer (23) and its variants (12–15, 17, 18) show excellent performance on many computer vision tasks and the potential to replace CNNs. For example, ViT (11) adopts the classical transformer architecture [23] to achieve image classification; it first splits an image into non-overlapping patches and then regards these patches as input tokens for network training. To reduce the model complexity of the vision transformer, Swin-T (12) proposes an efficient shifted-window-based self-attention, and adopts two successive Swin transformer blocks to model non-local feature relations. For achieving dense prediction tasks (e.g, instance segmentation and object detection), Wang et al. (13) design the Pyramid Vision Transformer (PVT) and the Spatial-Reduction Attention (SRA) to effectively reduce resource consumption and computational costs of using transformer. Moreover, in high-level vision tasks, LeViT (14) develops an alternately residual block and employs the attention bias to replace traditional absolute position embeddings for achieving competitive performance. After that, transformer in transformer (TNT) (15) combines the patch-level and pixel-level transformer blocks; thus, this architecture can effectively represent the feature relations between and within regions. In low-level vision tasks, Chen et al. (17) not only construct a large-scale benchmark based on the ImageNet dataset, but also design an image processing transformer (IPT) to serve various image restoration tasks including image super-resolution, denoising, and deraining. Then, Uformer (18) presents a hierarchical U-shaped transformer architecture with skip connections like U-Net (24). By combining the depth-wise convolution in basic transformer blocks, Uformer can capture long-range and short-range dependencies (global and local information) simultaneously. However, the above vision transformers fail to comprehensively consider the multi-scale information of an image. In this paper, we further propose an MC-ViT, which can effectively extract and employ multi-scale features to improve network performance.

Attention mechanism

In deep learning-based methods, the attention mechanism can enhance important features as well as suppress redundant features, thereby improving the network performance on various computer vision tasks. In general, attention mechanisms are mainly divided into three classes according to different modes of action (1): channel attention, (2) spatial attention, and (3) self-attention. In addition to the self-attention mechanism mentioned above, it is worth noting that the Squeeze-and-Excitation (SE) module (25) is the first plug-and-play channel attention mechanism, which can model the cross-channel interdependence to enhance the useful channels of features. Motivated by the SE module, selective kernel network (SKNet) (26) presents to use the multi-scale information with different receptive fields to adjust the weights of the channel attention. Subsequently, Woo et al. (27) design a convolutional block attention module, which not only proposes spatial attention to enhance important feature locations by aggregating neighborhood information, but also combines spatial attention and the channel attention for achieving attention complementarity. The similar spatial attention is also used in the Attention-UNet (28). In addition, triplet attention (29) and tensor element self-attention (30) can establish the cross-dimension feature interactions for achieving multi-view spatial attention. More recently, to model the attention across multi-scale features, cross-MPI (31) presents to use the batch-wise multiplication to explicitly correlate input features and corresponding multi-depth planes. Different from the above methods, we develop an efficient cross-correlation attention module in CAST; this attention mechanism can model the spatial-level multi-scale feature relations and then enhance the multi-scale fusion features at each transformer block. Extensive experiments also demonstrate that the proposed CAM is effective to improve the network performance on the thymoma typing task.

Materials and methods

Patients and dataset

In this study, all content, including the informed consent of patients, received approval from the Institutional Ethics Review Committee of the China–Japan Friendship Hospital. Specifically, we collected 323 H&E-stained whole slides from 129 thymoma and TC patients, and show the clinical information of such patients in Table 1. Afterwards, we produced corresponding thymoma histopathology WSIs by scanning these slides through the high-throughput digital scanner Shenzhen Shengqiang Technol. Co. Ltd (Slide Scan SystemSQS-600P). Each WSI has three magnification scales in terms of 10×, 20× and 40× with

TABLE 1 Clinical information of patients.

Basic information of patients			Thymoma typing information of patients							
Male	Female	Age	A	AB	B1	B1+B2	B2	B2+B3	B3	TC
61	68	17–81	12	30	15	18	20	9	19	6

resolutions of 0.57 μm/pixel , 0.29 μm/pixel , and 0.14 μm/pixel , respectively. To obtain accurate thymoma typing annotations, we invited experienced pathologists to label such WSIs as eight thymoma types, as shown in Figure 2, namely, type A, type AB, type B1, type B1+B2, type B2, type B2+B3, type B3, and TC. Considering the morphological continuum characteristic of thymomas, it remains a challenge to effectively distinguish the types B1, B2, and B3 thymomas. At present, the manual annotation of thymomas is mainly dependent on the experience and subjective judgment of pathologists, so there is usually a certain difference between the annotation results of different pathologists. To improve the annotation quality of the training set, the invited pathologists use the collective discussion to determine the type of each patient, and during the annotation, they check the corresponding immunohistochemistry (IHC)-stained WSI of each H&E-stained WSI to define a more accurate thymoma type. In addition, different pathological information related to thymoma typing is also labeled on WSIs, including spindle thymic epithelial cells, B1 thymic epithelial cells, B2 thymic epithelial cells, B3 thymic epithelial cells, fibrous septa, erythrocyte, lymphocyte, perivascular space, medullary differentiated areas, and tumor. For some indistinguishable classes like thymic epithelial cells (B1, B2, and B3), we provide the corresponding IHC-stained WSIs, which can help us locate thymic epithelial cells and calculate the ratio between epithelial cells and lymphocytes in local WSI regions. Concretely, the number of lymphocytes is more than that of epithelial cells in B1 thymoma WSIs, the number of lymphocytes is close to that of epithelial cells in B2 thymoma WSIs, and the number of lymphocytes is lower than that of epithelial cells in B3 thymoma WSIs. Furthermore, there are still slight differences in the nuclear heterogeneity, cell size, and chromatin for thymic epithelial cells (B1, B2, and B3). The above properties can also assist pathologists in distinguishing the thymoma type of a WSI. In this way, we consider the epithelial cells in B1, B2, or B3 thymoma WSIs as B1, B2, or B3 thymic epithelial cells. As shown in Figure 2, a total of 10 classes of pathological information can be used to train the proposed MC-ViT for improving the accuracy of thymoma typing. After that, we denote these labeled data as the thymoma histopathology WSI dataset, where 243 WSIs are selected to train the proposed pipeline and 80 other WSIs are used as the test set. Among them, each WSI is divided into 3,000 non-overlapping patches with three resolutions (64×64, 128×128, and 256×256) for network training. By constructing this large-scale dataset, we can effectively achieve the thymoma histopathology WSI typing to further assist doctors in diagnosing thymoma or TC.

Overall architecture

Thymoma typing is a complex and challenging digital pathology workflow. As shown in Figure 2, doctors usually need to comprehensively consider different local pathological information from multi-scale (10×, 20×, and 40× magnifications) WSIs to confirm the thymoma type.

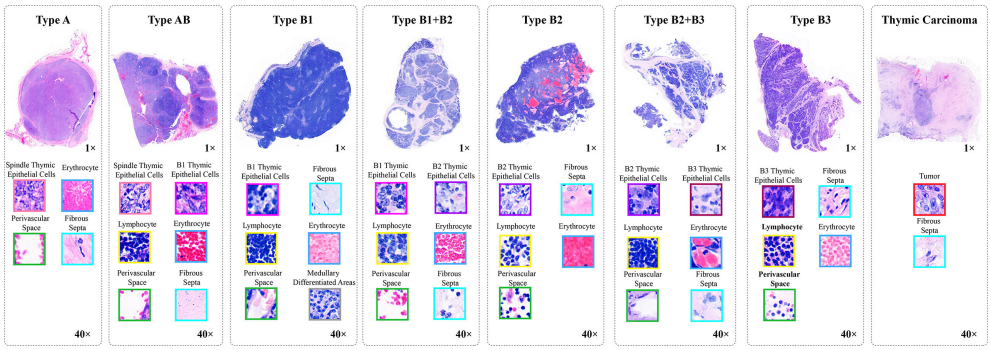


FIGURE 2
The pathological information against the thymoma types. Concretely, there are 10 pathological information classes (spindle thymic epithelial cells, B1 thymic epithelial cells, B2 thymic epithelial cells, B3 thymic epithelial cells, fibrous septa, erythrocyte, lymphocyte, perivascular space, medullary differentiated areas, and tumor) and eight types (A, AB, B1, B1+B2, B2, B2+B3, B3 and TC).

Therefore, taking local pathological information as priors can effectively achieve the deep learning-based thymoma histopathology WSI typing. In this paper, we propose an MC-ViT and show its overall architecture in Figure 3. Concretely, the proposed MC-ViT consists of two sub-networks: (1) the CAST for pathological information classification, and (2) the WSIs transformer for thymoma typing.

In the concrete implementation, each WSI is firstly split into 10×, 20× and 40× magnification WSI patches with sizes of $H/2\times W/2\times 3$, $H\times W\times 3$, and $2H\times 2W\times 3$, respectively, where multi-scale WSI patches at the same position on a WSI can form a group of network inputs, and H and W represent the height and width of WSI patches. The first sub-network CAST is designed as a three-branch structure, where the local-guided branch (LGB) and the global-guided branch (GGB) can extract the local and global receptive field features from 40× and 10× WSI patches, respectively, and the feature aggregation branch (FAB) takes 20× WSI patches as inputs. In the above branches, we first use a patch splitting layer to split and flatten input WSI patches into non-overlapping 1D features, and then adopt a linear embedding layer to project these 1D features to the expected dimensions, like Swin-T (12), where each group of 1D features can be regarded as a "token". After that, we utilize three well-established transformer architectures including Swin-T (12), PVT (13), and ViT (11) to construct LGB, FAB, and GGB, respectively, for adapting multi-scale input features. Concretely, each branch is built as a hierarchical structure with three stages, LGB, FAB, and GGB, which respectively use the window-based multi-head self-attention (W-MSA), the spatial reduction attention (SRA), and the multi-head self-attention (MSA) to build basic transformer blocks as shown in Figure 4, and adopt the patch splitting layer with 4×4 kernel size to achieve two times down-sampling for token sequences to produce hierarchical representations. The configurations of each network branch are illustrated in Table 2. Different from LGB and GGB, to effectively predict pathological information classes of input WSI patches, the FAB fuses multi-scale (multiple receptive fields) features from different branches at each transformer block. Here, we carefully design a cross-correlation attention module, which can establish the spatial-level relations between multi-scale features with potential pathological information, thereby promoting the multi-scale feature fusion in the transformer.

The second sub-network WT is a simple but effective three-stage transformer encoder. For a WSI, we randomly select a fixed number of WSI patches, and then through the CAST to produce the multi-scale embeddings and the pathological information labels of these WSI patches. Specifically, we first concatenate the multi-scale embeddings of each WSI patch with the corresponding pathological information label at the channel dimension, and then connect the concatenated features of WSI patches at the node dimension. To this end, each WSI can be encoded into a feature matrix $M\in\mathbb{R}^{m\times 769}$ with pathological information priors to train the proposed WT, where M indicates that each WSI is divided into M small patches. In the WT, we use classical transformer blocks (11) with absolute position encodings to process the input feature matrices for thymoma typing. In addition, converting a 2D full-scale WSI to a 1D feature matrix can significantly reduce the computational costs of the transformer.

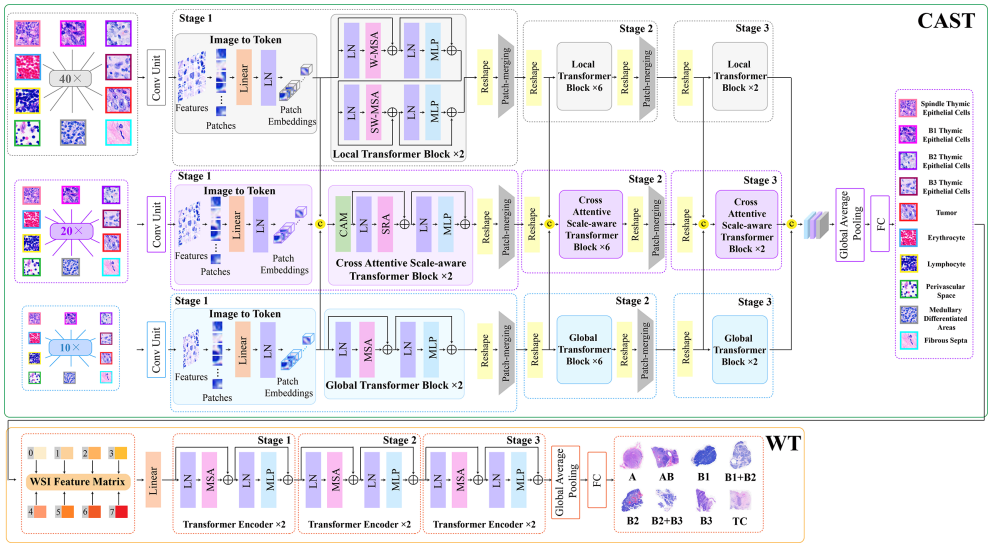


FIGURE 3
The architecture of the proposed multi-path cross-scale vision transformer (MC-ViT), which consists of the cross attentive scale-aware transformer (CAST) for pathological information classification and the WSIs transformer (WT) for thymoma typing.

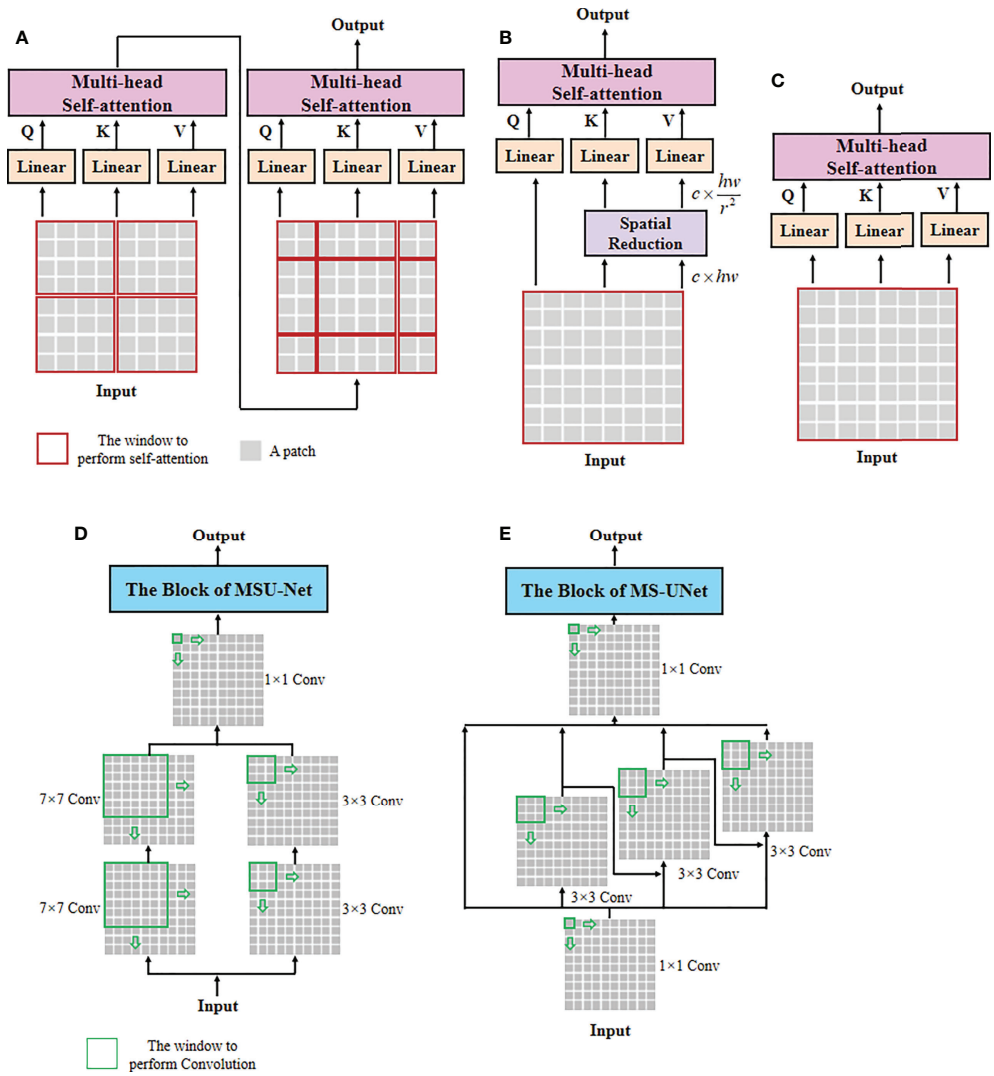


FIGURE 4
The architectures of self-attention and multi-scale convolution. **(A)** Window-based multi-head self-attention (W-MSA), **(B)** spatial reduction attention (SRA), **(C)** multi-head self-attention (MSA), **(D)** multi-scale convolution of MSU-Net, and **(E)** multi-scale convolution of MS-UNet.

Cross attentive scale-aware transformer

Unlike general natural and medical images (32, 33), a WSI usually has three magnification scales in terms of 10 × , 20 ×, and 40 × . To effectively utilize different scale WSIs for modeling multi-scale feature relations, we propose a CAST consisting of three kinds of basic transformer blocks, namely, the global transformer block, the CAST block, and the local transformer block. As shown in Figure 4, the proposed CAST is also different from existing advanced multi-scale U-Net architectures. For examples, Su et al. (34) design MSU-Net that uses scale-specific convolutions (1×1, 3×3, and 7×7) to capture multi-scale features (see Figure 4D). Kushnure et al. (35) construct MS-UNet to process the split feature channels to produce multi-scale representations (see Figure 4E). However, locally connected convolutions are not enough to extract sufficient global information, which limits the receptive fields of both MSU-Net and MS-UNet. In contrast, the proposed CAST can capture richer global information by three different non-local self-attention mechanisms, and fully leverage multi-scale WSIs (10×, 20× and 40×) rather than only the multi-scale features from a scale-specific WSI. Then, considering that the above transformer blocks have different receptive fields, the clinical observation process for thymoma histopathology WSIs can be effectively simulated in the proposed CAST. Concretely, in GGB, the global transformer block has similar configurations to that of the classical transformer block (11), which contains an MSA, a multi-layer perception (MLP), and two layer normalizations

TABLE 2 The network configurations of the proposed MC-ViT, where *P*, *C*, *N*, and *E* indicate the patch size, the channel number of the output, the head number of transformer block, and the expansion ratio of MLP, respectively.

		Stage	Branch	Input size	Patch merging	Transformer encoder	Output size
MC-ViT	CAST	Stage 1	LGB (40×)	256^2×3	$P = 8, C = 128$	$\begin{bmatrix} N = 2 \\ E = 8 \end{bmatrix} \times 2$	32^2×128
			FAB (20 ×)	128^2×3	$P = 4, C = 128$	$\begin{bmatrix} N = 2 \\ E = 8 \end{bmatrix} \times 2$	32^2×128
			GGB (10 ×)	64^2×3	$P = 2, C = 128$	$\begin{bmatrix} N = 2 \\ E = 8 \end{bmatrix} \times 2$	32^2×128
		Stage 2	LGB (40 ×)	32^2×128	$P = 2, C = 256$	$\begin{bmatrix} N = 4 \\ E = 4 \end{bmatrix} \times 2$	16^2×256
			FAB (20 ×)	32^2×128	$P = 2, C = 256$	$\begin{bmatrix} N = 4 \\ E = 4 \end{bmatrix} \times 2$	16^2×256
			GGB (10 ×)	32^2×128	$P = 2, C = 256$	$\begin{bmatrix} N = 4 \\ E = 4 \end{bmatrix} \times 2$	16^2×256
		Stage 3	LGB (40 ×)	16^2×256	$P = 2, C = 512$	$\begin{bmatrix} N = 8 \\ E = 4 \end{bmatrix} \times 2$	8^2×512
			FAB (20 ×)	16^2×256	$P = 2, C = 512$	$\begin{bmatrix} N = 8 \\ E = 4 \end{bmatrix} \times 2$	8^2×512
			GGB (10 ×)	16^2×256	$P = 2, C = 512$	$\begin{bmatrix} N = 8 \\ E = 4 \end{bmatrix} \times 2$	8^2×512
	WT	Stage 1	–	512×769	–	$\begin{bmatrix} N = 12 \\ E = 4 \end{bmatrix} \times 2$	512×769
		Stage 2	–	512×769	–	$\begin{bmatrix} N = 12 \\ E = 4 \end{bmatrix} \times 2$	512×769
		Stage 3	–	512×769	–	$\begin{bmatrix} N = 12 \\ E = 4 \end{bmatrix} \times 2$	512×769

(LNs) before MSA and MLP with the GELU non-linear layers (23). The calculation process in the global transformer block is

$$\tilde{A}_i = \mathrm{MSA}(\mathrm{LN}(A_{i-1})) + A_{i-1}, \quad (1)$$

$$A_i = \mathrm{MLP}\left(\mathrm{LN}(\tilde{A}_i)\right) + \tilde{A}_i, \quad (2)$$

where A_{i-1} and A_i are the input and output features of the i th global transformer block, and $\tilde{A}_i$ denotes the output of intermediate features by the MSA.

Then, in LGB, the local transformer block continues the advantages of Swin-T (12), which replaces the MSA with the window-based multi-head self-attention, and employs two successive Swin transformer blocks to achieve cross-window connections. The concrete configurations are shown in Figure 3; compared with MSA, W-MSA focuses more on modeling the feature relations in non-overlapping local windows, which not only effectively promotes the extraction of local information, but also significantly reduces the computations of transformer blocks. The local transformer block can be computed as

$$\tilde{B}_i = \mathrm{W-MSA}(\mathrm{LN}(B_{i-1})) + B_{i-1}, \quad (3)$$

$$\hat{B}_i = \mathrm{MLP}\left(\mathrm{LN}(\tilde{B}_i)\right) + \tilde{B}_i, \quad (4)$$

$$\bar{B}_i = \mathrm{SW-MSA}\left(\mathrm{LN}(\hat{B}_i)\right) + \hat{B}_i, \quad (5)$$

$$B_i = \mathrm{MLP}(\mathrm{LN}(\bar{B}_i)) + \bar{B}_i, \quad (6)$$

Where B_{i-1} and B_i are the input and output features of the i th local transformer block, SW-MSA is the multi-head self-attention with the shifted windowing configuration, and $\tilde{B}_i$, $\hat{B}_i$, and $\bar{B}_i$ represent the intermediate features output by MSA, the first MLP, and SW-MSA, respectively. Referring to Swin-T (12), we adopt the relative position bias to compute W-MSA and SW-MSA, which can be expressed as

$$\text{Self}-\text{attention}(Q,K,V)=\text{Softmax}(QK^T/\sqrt{d}+R)V, \quad (7)$$

where Q , K, and V are the query, key, and value matrices, respectively.

Finally, in FAB, we propose the cross-correlation attention module to combine with the spatial-reduction attention (13) to construct the cross attentive scale-aware (CAS) transformer block. Specifically, each CAS transformer block is composed of a CAM, an SRA, an MLP, and two LNs. Different from the global and local transformer blocks, we first adopt a CAM to aggregate and enhance the multi-scale features A_i , B_i, and C_i from different branches. With this design, the spatial-level feature relations can be supplemented and the representation of potential pathological information can be boosted effectively. Then, MLP can update the multi-scale fusion features captured by SRA accompanied with LNs for stable training and rapid convergence. The CAS transformer block can be formulated as

$$\tilde{C}_i=\text{SRA}(\text{LN}(\text{CAM}([A_{i-1},B_{i-1},C_{i-1}])))+C_{i-1}, \quad (8)$$

$$C_i=\text{MLP}\left(\text{LN}(\tilde{C}_i)\right)+\tilde{C}_i, \quad (9)$$

Where C_{i-1} and C_i are the input and output features of the i th CAS transformer block, and $\tilde{C}_i$ denotes the intermediate features output by the SRA.

In addition, after the last transformer block of each stage, we use a 4 × 4 patch splitting (unfolding) layer $PS(\cdot)$ to down-sample the reshaped features, and a linear embedding layer $FC(\cdot)$ to project the down-sampled features to the expected dimension for producing hierarchical representations

$$A_i/B_i/C_i=FC(PS(reshape(A_i/B_i/C_i))). \quad (10)$$

In the proposed CAST, after fusing and updating each stage's multi-scale features, we use the last fully connected layer with softmax of FAB to predict the pathological information classes of input WSI patches. During the test process, the predicted pathological information labels and the extracted multi-scale embeddings from the same WSI are connected as an input feature matrix to feed the subsequent WT.

WSI transformer

Benefiting from the prediction for pathological information labels and the encoding for multi-scale embeddings by the first sub-network CAST, we can construct an efficient WT with a three-stage transformer encoder to further achieve thymoma histopathology WSI typing. As shown in Figure 3, after concatenating the pathological information labels and multi-scale embeddings to convert a full-scale WSI to a simple input feature matrix, the computations of WT are significantly reduced. Specifically, each stage contains two classical transformer blocks (11); the head number N of MSA and the expansion ratio E of MLP in each transformer block are set as 12 and 4, respectively. In addition, we not only introduce absolute position encodings, but also replace class tokens with a global average pooling layer and a fully connected layer (36) to improve the accuracy of thymoma typing. The network configurations of the proposed CAST and WT are shown in Table 2.

Cross-correlation attention module

Converting an image into a sequence of tokens will result in the spatial information loss, and most existing vision transformers (12, 13, 15, 16, 31) fail to consider the spatial-level relations between features. To address this issue, we propose a cross-correlation attention module to effectively establish the spatial-level relations between multi-scale features as well as achieving the multi-scale feature fusion. As shown in Figure 5, CAM can comprehensively consider different receptive field features with global and local information, and enhance the multi-scale fusion features through a spatial attention map generated by the cross-correlation attention mechanism. Considering that multi-scale features completely contain potential pathological information, the proposed CAM can further improve the accuracy of potential pathological classification. Specifically, CAM first concatenates the features $A_i\in\mathbb{R}^{c\times hw}$, $B_i\in\mathbb{R}^{c\times hw}$, and $C_i\in\mathbb{R}^{c\times hw}$ from GGB, LGB, and FAB, respectively, to generate the features $G\in\mathbb{R}^{3\times c\times hw}$, and then reshapes its size to $3\times c\times h\times w$. After a 1 × 1 convolution, we can get the spatial-level features $f_1\in\mathbb{R}^{3\times c\times h\times w}$. Moreover, we reshape the features $C_i\in\mathbb{R}^{c\times hw}$ as another spatial-level features $f_2\in\mathbb{R}^{c\times 1\times h\times w}$ and cross-correlate features $f_1\in\mathbb{R}^{3\times c\times h\times w}$ and $f_2\in^{c\times 1\times h\times w}$ by a batch-wise multiplication to establish the relations between multi-scale features for producing the attention map Att

$$Att=\sigma(conv_1(\text{reshape}([A_i,B_i,C_i]))\otimes\text{reshape}(C_i)), \quad (11)$$

Where $\sigma(\cdot)$ indicates the Sigmoid activation function, $conv_1$ $(\cdot)$ is the 1 × 1 convolution, $[\cdot,\cdot,\cdot]$ and reshape$(\cdot)$ denote the feature concatenation and reshape operations, and $\otimes$ is the batch-wise matrix multiplication.

After that, we split the attention map $Att\in\mathbb{R}^{3\times h\times w}$ into three individual attention maps $Att_A\in\mathbb{R}^{1\times h\times w}$, $Att_B\in\mathbb{R}^{1\times h\times w}$, and $Att_C\in\mathbb{R}^{1\times h\times w}$ to enhance corresponding spatial-level features $f_A\in\mathbb{R}^{c\times h\times w}$, $f_B\in\mathbb{R}^{c\times h\times w}$, and $f_C\in\mathbb{R}^{c\times h\times w}$ by the element-wise multiplication. Here, the features f_A , f_B , and f_C related to different receptive field information are split from the spatial-level features $G\in\mathbb{R}^{3\times c\times h\times w}$. Then, we re-aggregate the enhanced spatial-level features by a 3 × 3 convolution to generate the features $f_3\in\mathbb{R}^{c\times h\times w}$. The final output $F\in\mathbb{R}^{c\times hw}$ of CAM can be obtained by a reshape operation, and the above process is expressed as

$$F=\text{reshape}(conv_3([Att_A\,\dot{o}\,f_A,Att_B\,\dot{o}\,f_B,Att_C\,\dot{o}\,f_C])), \quad (12)$$

where $\dot{o}$ denotes the element-wise multiplication.

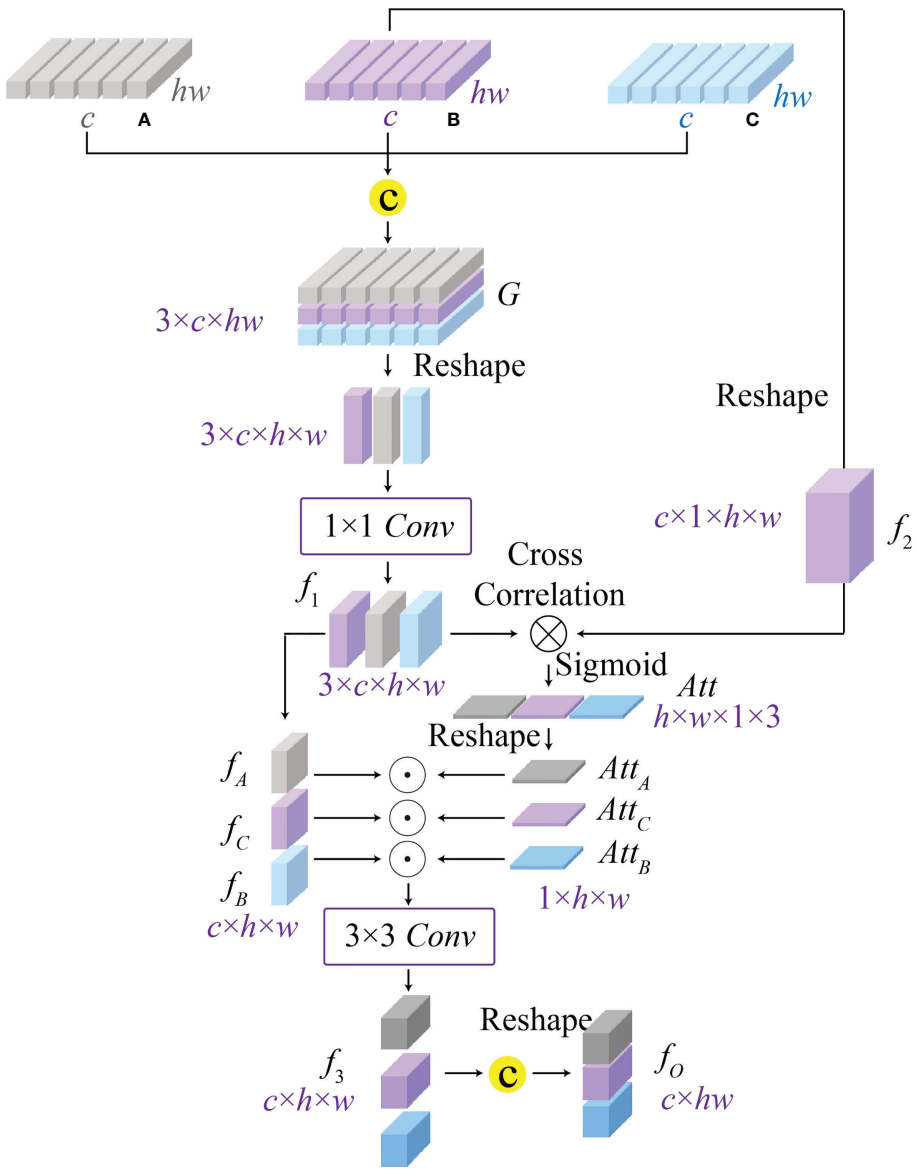


FIGURE 5
The architecture of the proposed cross-correlation attention module (CAM), which can model the spatial-level relationship between multi-scale features **(A–C)** from the global-guided, the local-guided, and the feature aggregation branches for achieving the multi-scale feature fusion.

Loss functions

The first sub-network CAST can classify input WSI patches into 10 pathological information classes, including spindle thymic epithelial cells, B1 thymic epithelial cells, B2 thymic epithelial cells, B3 thymic epithelial cells, fibrous septa, erythrocyte, lymphocyte, perivascular space, medullary differentiated areas, and tumor. Specifically, we use the cross-entropy loss (22) to train the proposed CAST, which can minimize the distance between predicted probabilities and corresponding ground truths by the following expression

$$\mathcal{L}_{\text{CAST}} = -\sum_{k=1}^{K} y_k \log\ (p_k), \tag{13}$$

where K is the number of pathological information classes, p_k represents the predicted probability that an input WSI patch belongs to the k th pathological information class, and y_k is its ground truth.

After that, the second sub-network WT can predict the thymoma type of the input feature matrix for achieving thymoma typing. Concretely, there are eight thymoma types (A, AB, B1, B1+B2, B2, B2+B3, B3, and C) in our task. Similarly,

we also adopt the cross-entropy loss to optimize this multi-classification task as

$$\mathcal{L}_{\mathrm{WT}} = -\sum_{t=1}^{T} Y_t \log\ (P_t), \tag{14}$$

where T is the number of thymoma types, P_t represents the predicted probability that a input feature matrix belongs to the t th thymoma type, and Y_t is its ground truth.

Experimental results and analysis

Implementation details

The proposed MC-ViT is programmed by PyTorch 1.9.0 and all experiments are conducted on a server with Intel (R) Core (TM) i9-10850K CPU (5.0 GHz) and NVIDIA GeForce RTX 3090 GPU (24GB). In our concrete implementation, the Adam optimizer with momentums $\bar{e}ta_1$ = 0.9 and β_2=0.999 is used to optimize both CAST and WT. For the proposed CAST, there are 160 epochs in network training with batch size 64 and the initial learning rate 2e -3. Moreover, the proposed WT is trained in 160 epochs using batch size 8 and the initial learning rate 1e -3. In Figure 6, we report the training loss and accuracy against training epochs to show the effectiveness and convergence of the proposed CAST and WT.

Evaluation metrics

To comprehensively evaluate the performance of the proposed CAST for pathological information classification and the performance of the proposed WT for thymus typing, we introduce eight well-established metrics, namely, recall (37) (Rec), Top-1 accuracy (Top-1 Acc), mean accuracy (38) (MAcc), precision (37) (Pre), F-measure (38) (F1), receiver operating characteristic (ROC) curve, area under the curve (AUC), and confusion matrix (CM), and three statistical metrics, namely, sensitivity and specificity with the 95% confidence interval (CI) and the two-sided McNemar's tests (39) (test statistic and asymptotic Sig.). For the first five metrics, the larger values indicate a classification method has better performance. The AUC is defined as the area surrounded by the coordinate axis and the ROC curve, where a large AUC value denotes a high classification accuracy.

Evaluation for pathological information classification

This subsection first compares the proposed CAST with four well-known vision transformers, ViT (11), TNT (15), LeViT (14), and CrossViT (40); two classical CNNs, ResNet-101 (41) and DenseNet-121 (42); and four state-of-the-art medical image classification methods, GuSA-Net (43), ROPsNet (44), CPWA-Net (45), and IL-MCAM (46). The quantitative results on the proposed THW dataset are shown in Table 3; compared with existing advanced classification methods, the proposed CAST achieves 0.016, 0.012, 0.015, and 0.007 improvements in terms of Rec, Top-1 Acc, Macc, and F1, respectively. In general, some classical transformer-based and CNN-based methods, such as ViT, TNT, LeViT, and ResNet-101, fail to achieve satisfactory classification results, which could be attributed to the fact that these methods ignore capturing and utilizing the inherent multi-scale information in WSIs. In contrast, state-of-the-art IL-MCAM and CrossViT achieve better classification accuracy since both of them are built as the multi-scale network architecture. It is noteworthy that GuSA-Net is the improvement of DenseNet; thus, its classification performance is slightly better than that of DenseNet. Currently, in most

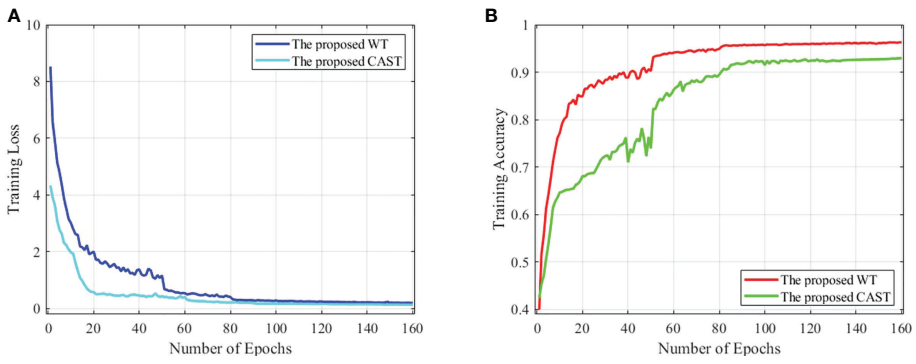


FIGURE 6
(A) The training loss against training epochs and **(B)** the training accuracy against training epochs on the THW dataset.

TABLE 3 Quantitative comparisons (Rec, Top-1 Acc, MAcc, Pre, and F1) for pathological information classification on the THW dataset.

Methods	Pathological Information Classification				
	Rec	Top-1 Acc	Macc	Pre	F1
(ICLR'2021) ViT (11)	0.813	0.834	0.804	0.810	0.811
(NIPS'2021) TNT (15)	0.814	0.828	0.813	0.817	0.815
(ICCV'2021) LeViT (14)	0.821	0.857	0.833	0.827	0.824
(ICCV'2021) CrossViT (40)	0.897	0.886	0.860	0.867	0.882
(CVPR'2016) ResNet-101 (41)	0.819	0.836	0.802	0.808	0.813
(CVPR'2017) DenseNet-121 (42)	0.873	0.848	0.834	0.860	0.867
(TMI'2020) GuSA-Net (43)	0.918	0.927	0.909	0.925	0.921
(TMI'2021) ROPsNet (44)	0.874	0.892	0.886	0.882	0.878
(JBHI'2021) CPWA-Net (45)	0.817	0.832	0.821	0.813	0.815
(CBM'2022) IL-MCAM (46)	0.906	0.918	0.912	0.903	0.904
CAST (Ours)	0.934	0.939	0.927	0.922	0.928

Red and blue contents represent the best and suboptimal results, respectively.

clinical cases, doctors need to comprehensively observe the multi-scale (10 × , 20 ×, and 40 ×) local patches of a WSI to determine its pathological information classes, and then diagnose the corresponding thymoma type. The proposed CAST effectively simulates the above process by taking multi-scale WSI patches as inputs and fusing multi-scale features in each stage. As a result, we successfully achieve an improvement of 0.015 on MAcc compared with the state-of-the-art IL-MCAM, and about 0.023 average improvement on other evaluation metrics.

To further verify the effectiveness of the proposed CAST, we illustrate the ROC curve and AUC of each pathological information class in the left part of Figure 7. It can be seen that the proposed CAST performs well on six classes, namely, erythrocyte, lymphocyte, spindle thymic epithelial cells, B1 thymic epithelial cells, B2 thymic epithelial cells, and B3 thymic epithelial cells. For the other three pathological information classes, where fibrous septa and perivascular space can be distinguished on H&E-stained WSIs according to the color and position information of existing cells (like fibroblasts and erythrocytes), medullary differentiated areas are usually distinguishable on IHC-stained WSIs. Compared with the above five pathological information classes, our pipeline achieves slightly poor but still competitive classification results for fibrous septa, perivascular space, and medullary differentiated areas. In summary, the proposed CAST effectively distinguishes each pathological information class using only H&E-stained WSIs. Since the classification results of pathological information are closely related to thymoma types, CAST can assist the subsequent WT for thymoma typing.

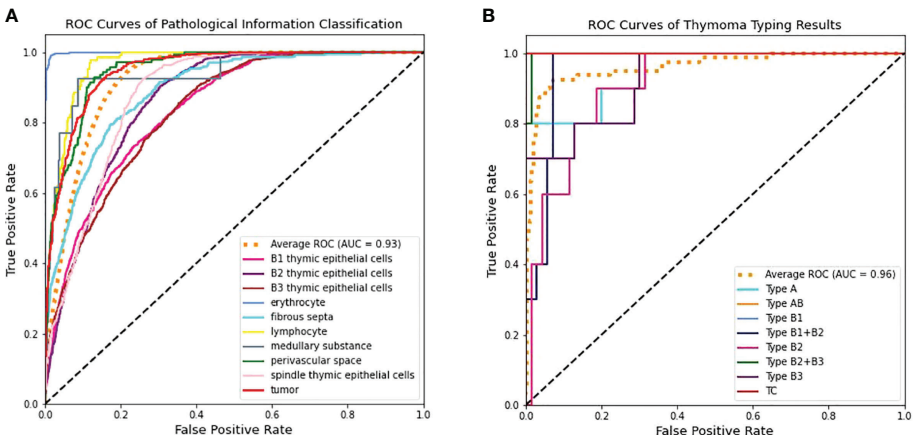


FIGURE 7
The ROC curve and AUC on the THW dataset. **(A)** Pathological information classification and **(B)** thymoma typing.

Thymoma typing evaluation

Clinically, the thymoma type of a WSI is reflected by multiple pathological information; hence, theoretically, the high-precision classification results for pathological information are helpful to improve the accuracy of thymoma typing. To demonstrate the above content, we respectively use the classification methods ViT, TNT, LeViT, CrossViT, ResNet-101, DenseNet-121, GuSA-Net, and IL-MCAM to predict the pathological information labels and the uniform size embeddings of each WSI. By concatenating these labels and embeddings to produce input feature matrices to train WT, we can denote corresponding comparison methods as ViT+WT, TNT+WT, LeViT+WT, CrossViT+WT, ResNet-101+WT, DenseNet-121+WT, GuSA-Net+WT, ROPsNet+WT, CPWA-Net+WT, and IL-MCAM+WT. Their predicted results are shown in Table 4; we can observe that the proposed MC-ViT (CAST+WT) achieves the best classification accuracy, especially on Top-1 Acc (about 0.017 improvement) and F1 (about 0.016 improvement). The ROC curve and AUC of each thymoma type are also shown in the right part of Figure 7, which further proves that the proposed MC-ViT is effective to classify various thymoma types. Based on the above quantitative analysis, we can conclude that the pathological information labels provided by CAST help to achieve thymoma histopathology WSI typing, and the quality of such labels and embeddings determines the typing accuracy.

In addition, we statistically analyze the performance of these comparison methods and our MC-ViT by computing the sensitivity and specificity with 95% CI, and the two-sided McNemar's tests (test statistic and asymptotic Sig.). As can be seen from Table 5, the proposed MC-ViT achieves completely correct typing results (sensitivity) for types AB, B1, B2, and C, and the competitive average 0.875 sensitivity (95% CI: 0.528–0.970) and 0.982 specificity (95% CI: 0.911–0.992) for the thymoma typing task. Moreover, the two-sided McNemar's tests (average 1.810 test statistic and 0.42996 asymptotic Sig.) further show the statistical significance of our predicted results,

TABLE 4 Quantitative comparisons (Rec, Top-1 Acc, Macc, Pre, and F1) for thymoma typing on the THW dataset.

Methods	Thymoma Typing				
	Rec	Top-1 Acc	MAcc	Pre	F1
(ICLR'2021) ViT (11)+WT	0.832	0.839	0.820	0.824	0.828
(NIPS'2021) TNT (15)+WT	0.825	0.861	0.836	0.839	0.852
(ICCV'2021) LeViT (14)+WT	0.844	0.868	0.843	0.849	0.846
(ICCV'2021) CrossViT (40)+WT	0.903	0.899	0.875	0.879	0.891
(CVPR'2016) ResNet-101 (41)+WT	0.841	0.831	0.819	0.825	0.833
(CVPR'2017) DenseNet-121 (42)+WT	0.902	0.863	0.842	0.865	0.883
(TMI'2020) GuSA-Net (43)+WT	0.931	0.937	0.916	0.934	0.923
(TMI'2021) ROPsNet (44)+WT	0.881	0.898	0.890	0.896	0.888
(JBHI'2021) CPWA-Net (45)+WT	0.848	0.856	0.843	0.846	0.847
(CBM'2022) IL-MCAM (46)+WT	0.921	0.928	0.915	0.908	0.914
CAST+WT (Ours)	0.948	0.951	0.942	0.931	0.939

Red and blue contents represent the best and suboptimal results, respectively.

TABLE 5 Quantitative comparisons (sensitivity and specificity with 95% CI, test statistic, and asymptotic Sig.) for thymoma typing on the THW dataset.

Types	Sensitivity (95% CI)	Specificity (95% CI)	Test Statistic	Asymptotic Sig.
A	0.800 (0.442–0.965)	1.000 (0.935–1.000)	0.500	0.47950
AB	1.000 (0.655–1.000)	1.000 (0.935–1.000)	–	–
B1	1.000 (0.655–1.000)	0.986 (0.912–0.999)	0.000	1.00000
B1+B2	0.800 (0.442–0.965)	1.000 (0.935–1.000)	0.500	0.47950
B2	1.000 (0.655–1.000)	0.871 (0.765–0.936)	7.111	0.00766
B2+B3	0.800 (0.442–0.965)	1.000 (0.935–1.000)	0.500	0.47950
B3	0.600 (0.274–0.863)	1.000 (0.935–1.000)	2.250	0.13361
TC	1.000 (0.655–1.000)	1.000 (0.935–1.000)	–	–
Average	0.875 (0.528–0.970)	0.982 (0.911–0.992)	1.810	0.42996

which have slight differences from the expert-annotated ground truths.

Discussion

In this section, we discuss the effectiveness of the proposed multi-scale (multi-path) transformer architecture and the cross-correlation attention mechanism. Concretely, we define eight ablation models: (1) Single-branch Swin-T Transformer (SSwT): SSwTonly has LGB for processing 40 × WSIs; (2) Single-branch Pyramid Vision Transformer (SPVT): SPVT only has FAB for processing 20 × WSIs; (3) Single-branch Vision Transformer (SViT): SViT only has GGB for processing 10 × WSIs; (4) CAST without (w/o) GGB: this model has LGB and FAB for processing 40 × and 20 × WSIs; (5) CAST w/o FAB: this model has LGB and GGB for processing 40 × and 10 × WSIs; (6) CAST w/o LGB: this model has FAB and GGB for processing 20 × and 10 × WSIs; (7) CAST w/o CAM: this model contains three paths but without CAM; and (8) the proposed CAST. For fair comparisons, the training dataset and implementation details remain unchanged, and corresponding experimental results are exhibited in the following subsections.

Ablation study for multiple paths of transformer

Firstly, we evaluate the effectiveness of multiple paths in the proposed CAST, where LGB, FAB, and GGB are ablated and adopted respectively to demonstrate their contributions. The quantitative results of seven ablation models, SSwT, SPVT, SViT, CAST w/o GGB, CAST w/o FAB, CAST w/o LGB, and CAST, are listed in Table 6. It can be seen that ablating GGB reduces the accuracy to classify the medullary differentiated areas and fibrous septa, ablating LGB weakens the performance to distinguish different thymic epithelial cells, and ablating FAB causes unsatisfactory results to recognize the perivascular space and lymphocyte. In summary, simultaneously adopting three paths in CAST to process multi-scale WSIs can achieve more excellent performance compared with using a single path or dual paths, and FAB brings the largest improvement to pathological information classification.

Ablation study for multi-scale transformer architecture

Next, we compare CAST w/o CAM and SPVT to verify the effectiveness of the proposed multi-scale transformer architecture. Specifically, CAST w/o CAM adopts three transformer branches to process 10×, 20× and 40× WSI patches, respectively, while replacing the proposed CAM by the traditional feature concatenation to achieve multi-scale feature fusion. SPVT only uses a single branch with the SRA-based transformer blocks to train 20 × WSI patches. The quantitative comparisons (Top-1Acc, Macc, and F1) on the THW dataset are reported in Figure 8A, and it can be seen that using the multi-scale transformer architecture brings significant performance improvements for pathological information classification. In addition, Figure 9 shows their confusion matrices, which further demonstrate that comprehensively considering the multi-scale information in WSIs can reduce the confusion between similar thymoma types.

Ablation study for cross-correlation attention mechanism

Finally, we evaluate the contribution of the proposed CAM to show its effectiveness on pathological information

TABLE 6 Ablation study (Acc and MAcc) for multiple paths in the proposed CAST on the THW dataset.

Acc of Each Class	SSwT	SPVT	SViT	CAST w/o GGB	CAST w/o FAB	CAST w/o LGB	CAST
Spindle Thymic Epithelial Cells	0.859	0.871	0.844	0.866	0.918	0.897	0.923
B1 Thymic Epithelial Cells	0.887	0.876	0.854	0.895	0.886	0.905	0.915
B2 Thymic Epithelial Cells	0.893	0.881	0.861	0.903	0.892	0.911	0.920
B3 Thymic Epithelial Cells	0.890	0.879	0.858	0.898	0.887	0.909	0.916
Fibrous Septa	0.859	0.877	0.896	0.915	0.919	0.902	0.924
Erythrocyte	0.912	0.918	0.904	0.925	0.926	0.933	0.941
Lymphocyte	0.861	0.896	0.849	0.902	0.924	0.927	0.938
Perivascular Space	0.865	0.894	0.842	0.898	0.917	0.929	0.936
Medullary Differentiated Areas	0.857	0.883	0.905	0.918	0.921	0.905	0.927
Tumor	0.887	0.886	0.882	0.897	0.905	0.908	0.929
MAcc	0.877	0.886	0.870	0.902	0.910	0.913	0.927

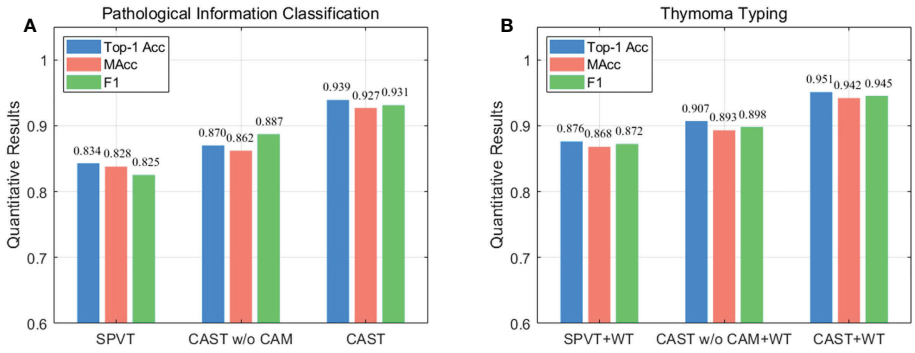


FIGURE 8
Ablation study (Top-1 Acc, MAcc, and F1) on the THW dataset, where w/o represents without such component. **(A)** Pathological information classification and **(B)** thymoma typing.

classification and thymoma typing. Corresponding quantitative results are shown in Figure 8, and from the overall integration of evaluation metrics Top-1 Acc, Macc, and F1, we can observe that adopting CAM to aggregate multi-scale features significantly improves the precision for pathological information classification and thymus typing. On the other hand, the confusion matrices about thymoma typing are reported in Figure 9; although CAST w/o CAM+WT outperforms SPVT +WT, some highly similar thymoma types are still difficult to distinguish, such as B1, B1+B2, and B2 types. By comparison, the proposed MC-ViT (CAST+WT) achieves better thymoma typing results. Overall, these ablation studies show that the accurate pathological information labels are beneficial for boosting thymoma typing accuracy, and the proposed

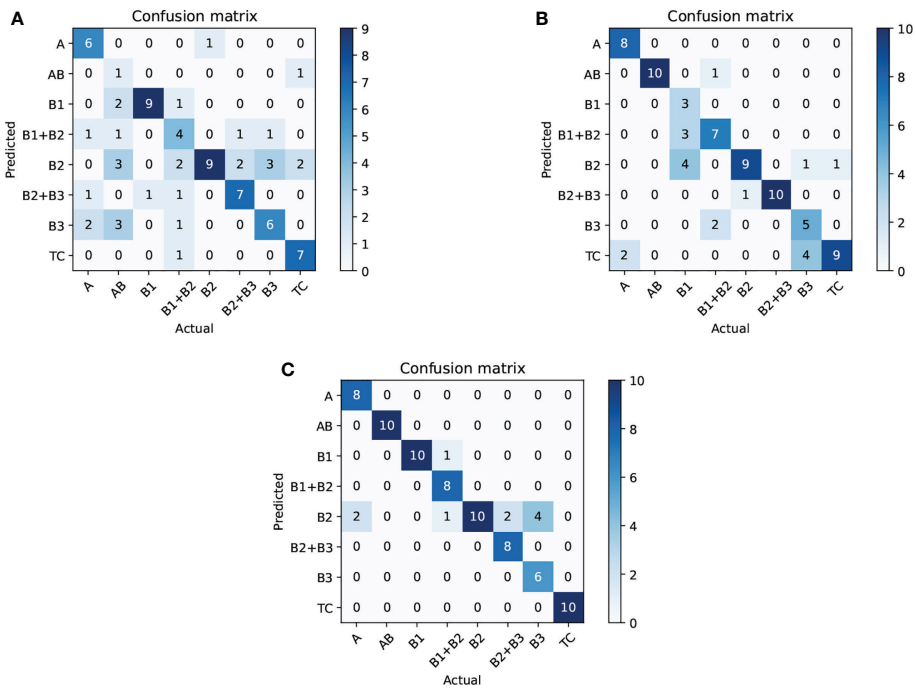


FIGURE 9
Ablation study (confusion matrix) for thymoma typing on the THW dataset; subfigures **(A–C)** are SPVT+WT, CAST w/o CAM+WT, and CAST+WT, where w/o represents without such component.

CAM is effective to improve pathological information classification results.

Unsupervised method for thymoma typing

In CAD, unsupervised methods are mainly used for processing unlabeled or incompletely labeled data, and they can automatically determine the total class of input data and then achieve the classification task. Traditional unsupervised methods include clustering and dimensionality reduction, and deep learning-based unsupervised methods include domain adaptation and contrastive learning. Compared with traditional methods, most deep learning-based methods have superior performance but require minor annotation information to assist network training, which means they fail to achieve full unsupervised classification. For example, domain adaptation methods require a small labeled dataset as the source domain to achieve the unsupervised classification of the target domain. Contrastive learning methods need to define the similarity between samples through pretext tasks, and then classify these data in a self-supervised (unsupervised) way. In general, supervised methods perform favorably against full unsupervised methods. In this work, we introduce a classical full unsupervised method (47) for thymoma typing, which is the combination of CNN and k -means clustering. We find that this method cannot successfully distinguish types B1, B1+B2, B2, B2 +B3, and B3; however, it still shows high potential when only classifying three types A, B, and TC (0.659 Top-1 Acc). Hence, we think that full unsupervised methods are more suitable for a simple classification of unlabeled data; they can provide certain diagnosis information for doctors while effectively reducing the time consumed by manual annotation. By comparison, supervised methods can better achieve precise thymoma typing when having sufficient labeled data.

Conclusions

In this paper, we propose an MC-ViT for achieving thymoma histopathology WSI typing. Aiming at full-scale WSIs that are difficult to train by deep learning-based methods, the proposed MC-ViT is designed as a twofold transformer architecture to separately predict the pathological information labels of WSI patches and the thymoma type of a WSI, where the former effectively fuses complementary multi-scale information to produce accurate pathological information priors, and the latter successfully converts the full-scale WSI to the low-cost feature matrix to achieve efficient network training by introducing such priors. In addition, we propose a cross-correlation attention mechanism to enhance and fuse multi-scale features with global and local receptive fields. Considering that CAM well establishes the spatial-level feature relations in the transformer, our thymoma typing results achieve further improvements. Extensive experiments also show that our MC-ViT outperforms most existing advanced transformer-based and CNN-based methods on the proposed THW dataset with 323 WSIs. In future works, we look forward to incorporating CT images and histopathology WSIs for achieving the multi-modal information fusion-based thymoma typing, which may further assist doctors to improve the efficiency and accuracy of thymoma and TC diagnosis. In addition, we will make the network outputs the soft labels (the probability of a WSI belongs to types B1, B2, and B3) instead of the hard labels (the class of a WSI belongs to type B1, B2, or B3) for thymoma WSIs with B1, B2, and B3 types, thereby providing more reasonable diagnosis information for doctors.

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.

Ethics statement

This study was approved by the institutional ethics review committee of the China–Japan Friendship Hospital. Written informed consent to participate in this study was provided by the participant's legal guardian.

Author contributions

HZ proposed the algorithm and wrote the manuscript. HC provided the guidance and labeled the pathological information classes and thymoma types of WSIs. JQ collected the dataset from the China–Japan Friendship Hospital. BW, GM, and DZ verified the medical research significance of this study. PW designed the figures and experiments, and revised the manuscript. JL provided the financial support and guided the study. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Key Research and Development Program of China (No. 2017YFA0700401), the National Natural Science Foundation of China (No.

KKA309004533, 81571836), and the Fundamental Research Funds for the Central Universities (2021YJS036).

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.

References

1. Marx A, Chan JK, Chalabreysse L, Dacic S, Detterbeck F, French CA, et al. The 2021 WHO classification of tumors of the thymus and mediastinum: What is new in thymic epithelial, germ cell, and mesenchymal tumors? *J Thorac Oncol* (2021) 17:200–13. doi: 10.1016/j.jtho.2021.10.010

2. Borczuk AC. *WHO classification of tumours: Thoracic tumours* (International agency for research on cancer (IARC) publications). (2021).

3. Scorsetti M, Leo F, Trama A, D'Angelillo R, Serpico D, Macerelli M, et al. Thymoma and thymic carcinomas. *Crit Rev oncolo/hematol* (2016) 99:332–50. doi: 10.1016/j.critrevonc.2016.01.012

4. Venuta F, Anile M, Diso D, Vitolo D, Rendina EA, De Giacomo T, et al. Thymoma and thymic carcinomas. *Eur J cardio-thoracic Surg* (2010) 37:13–25. doi: 10.1016/j.ejcts.2009.05.038

5. Han X, Gao W, Chen Y, Du L, Duan J, Yu H, et al. Relationship between computed tomography imaging features and clinical characteristics, masaoka–koga stages, and world health organization histological classifications of thymoma. *Front Oncol* (2019) 9:1041. doi: 10.3389/fonc.2019.01041

6. Luo T, Zhao H, Zhou X. The clinical features, diagnosis and management of recurrent thymoma. *J Cardiothorac Surg* (2016) 11:140. doi: 10.1186/s13019-016-0533-9

7. Zormpas-Petridis K, Failmezger H, Raza SEA, Roxanis I, Jamin Y, Yuan Y. Superpixel-based conditional random fields (SuperCRF): Incorporating global and local context for enhanced deep learning in melanoma histopathology. *Front Oncol* (2019) 10:1045. doi: 10.3389/fonc.2019.01045

8. Zormpas-Petridis K, Noguera R, Ivankovic DK, Roxanis I, Jamin Y, Yuan Y. SuperHistopath: a deep learning pipeline for mapping tumor heterogeneity on low-resolution whole-slide digital histopathology images. *Front Oncol* (2021) 9:586292. doi: 10.3389/fonc.2020.586292

9. Liu Y, Li X, Zheng A, Zhu X, Liu S, Hu M, et al. Predict ki-67 positive cells in H&E-stained images using deep learning independently from IHC-stained images. *Front Mol Biosci* (2020) 7:183. doi: 10.3389/fmolb.2020.00183

10. Xie J, Liu R, Luttrell IVJ, Zhang C. Deep learning based analysis of histopathological images of breast cancer. *Front Genet* (2019) 10:80. doi: 10.3389/fgene.2019.00080

11. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale, in: *Proc. Int. Conf. Learn. Represent. (ICLR)*, (Vienna, Austria: OpenReview.net).

12. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows, in: *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, (Montreal, QC, Canada: IEEE). pp. 10012–22. doi: 10.1109/ICCV48922.2021.00986

13. Wang W, Xie E, Li X, Fan DP, Song K, Liang D, et al. (2021). Pyramid vision transformer: A versatile backbone for dense prediction without convolutions, in: *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, (Montreal, QC, Canada: IEEE). pp. 568–78. doi: 10.1109/ICCV48922.2021.00061

14. Graham B, El-Nouby A, Touvron H, Stock P, Joulin A, Jégou H, et al. (2021). LeViT: a vision transformer in ConvNet's clothing for faster inference, in: *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, (Montreal, QC, Canada: IEEE). pp. 12259–69. doi: 10.1109/ICCV48922.2021.01204

15. Han K, Xiao A, Wu E, Guo J, Xu C, Wang Y. Transformer in transformer. *Proc Adv Neural Inform. Process Syst (NIPS)* (2021) 34:15908–19.

16. Yuan L, Chen Y, Wang T, Yu W, Shi Y, Jiang ZH, et al (2021). Tokens-to-token ViT: Training vision transformers from scratch on ImageNet, in: *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, (Montreal, QC, Canada: IEEE). pp. 558–67. doi: 10.1109/ICCV48922.2021.00060

17. Chen H, Wang Y, Guo T, Xu C, Deng Y, Liu Z, et al. (2021). Pre-trained image processing transformer, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (Nashville, TN, USA: IEEE). pp. 12299–310. doi: 10.1109/CVPR46437.2021.01212

18. Wang Z, Cun X, Bao J, Liu J. (2022). Uformer: A general U-shaped transformer for image restoration, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (New Orleans, LA, USA: IEEE). doi: 10.48550/arXiv.2106.03106

19. Chen H, Li C, Li X, Wang G, Hu W, Li Y, et al. GasHis-transformer: A multi-scale visual transformer approach for gastric histopathology image classification. *arXiv preprint arXiv:2104.14528* (2021). doi: 10.48550/arXiv.2104.14528

20. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. Swin-unet: Unet-like pure transformer for medical image segmentation. *arXiv preprint arXiv:2105.05537* (2021). doi: 10.48550/arXiv.2105.05537

21. Yan X, Tang H, Sun S, Ma H, Kong D, Xie X. After-unet: Axial fusion transformer unet for medical image segmentation. *Proc IEEE Winter Conf Appl Comput Vis (WACV)* (2022), 3971–81. doi: 10.1109/WACV51458.2022.00333

22. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Proc Adv Neural Inform. Process Syst (NIPS)* (2012) 60:84–90.

23. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. (2017). Attention is all you need. *Proc Adv Neural Inform. Process Syst (NIPS)* 2017:5998–6008.

24. Ronneberger O, Fischer P, Brox T. (2015). U-Net: Convolutional networks for biomedical image segmentation, in: *Proc. Int. Conf. Med. Image Comput. Computer-Assisted Interv. (MICCAI)*, (Munich, Germany: Springer). pp. 234–41. doi: 10.1007/978-3-319-24574-4_28

25. Hu J, Shen L, Sun G. (2018). Squeeze-and-excitation networks, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (Salt Lake City, UT, USA: IEEE). pp. 7132–41. doi: 10.1109/CVPR.2018.00745

26. Li X, Wang W, Hu X, Yang J. (2019). Selective kernel networks, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (Long Beach, CA, USA: IEEE). pp. 510–9. doi: 10.1109/CVPR.2019.00060

27. Woo S, Park J, Lee JY, Kweon IS. (2018). CBAM: Convolutional block attention module, in: *Proc. European Conf. Comput. Vis. (ECCV)*, (Munich, Germany: Springer). pp. 3–19. doi: 10.1007/978-3-030-01234-2_1

28. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention U-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* (2018). doi: 10.48550/arXiv.1804.03999

29. Misra D, Nalamada T, Arasanipalai AU, Hou Q. (2021). Rotate to attend: Convolutional triplet attention module, in: *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, (Waikoloa, HI, USA: IEEE). pp. 3139–48. doi: 10.1109/WACV48630.2021.00318

30. Babiloni F, Marras I, Slabaugh G, Zafeiriou S. (2020). TESA: Tensor element self-attention via matricization, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (Seattle, WA, USA: IEEE). pp. 13945–54. doi: 10.1109/CVPR42600.2020.01396

31. Zhou Y, Wu G, Fu Y, Li K, Liu Y. (2021). Cross-MPI: Cross-scale stereo for image super-resolution using multiplane images, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (Nashville, USA: IEEE). pp. 14842–51. doi: 10.1109/CVPR46437.2021.01460

32. Chen CF, Fan Q, Mallinar N, Sercu T, Feris R. (2018). Big-little net: An efficient multi-scale feature representation for visual and speech recognition, in: *Proc. Int. Conf. Learn. Represent. (ICLR)*, (Vancouver, BC, Canada: OpenReview.net).

33. Jiang K, Wang Z, Yi P, Chen C, Huang B, Luo Y, et al (2020). Multi-scale progressive fusion network for single image deraining, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (Seattle, WA, USA: IEEE), Vol. 2020). pp. 8346–55. doi: 10.1109/CVPR42600.2020.00837

34. Su R, Zhang D, Liu J, Cheng C. MSU-net: Multi-scale U-net for 2D medical image segmentation. *Front Genet* (2021) 140:639930. doi: 10.3389/fgene.2021.639930

35. Kushnure DT, Talbar SN. MS-UNet: A multi-scale UNet with feature recalibration approach for automatic liver and tumor segmentation in CT images. *Comput Med Imag Grap.* (2021) 89:101885. doi: 10.1016/j.compmedimag.2021.101885

36. Chu X, Tian Z, Zhang B, Wang X, Wei X, Xia H, et al. Conditional positional encodings for vision transformers. *arXiv preprint arXiv:2102.10882* (2021). doi: 10.48550/arXiv.2102.10882

37. Xue D, Zhou X, Li C, Yao Y, Rahaman MM, Zhang J, et al. An application of transfer learning and ensemble learning techniques for cervical histopathology image classification. *IEEE Access* (2020) 8:104603–18. doi: 10.1109/ACCESS.2020.2999816

38. Naylor P, Laé M, Reyal F, Walter T. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Trans Med Imag* (2018) 38:448–59. doi: 10.1109/TMI.2018.2865709

39. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer* (2019) 119:11–7. doi: 10.1016/j.ejca.2019.05.023

40. Chen CFR, Fan Q, Panda R. (2021). CrossViT: Cross-attention multi-scale vision transformer for image classification, in: *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, .(Montreal, QC, Canada: IEEE) pp. 357–66. doi: 10.1109/ICCV48922.2021.00041

41. He K, Zhang X, Ren S, Sun J. (2016). Deep residual learning for image recognition, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (Las Vegas, NV, USA: IEEE). pp. 770–8. doi: 10.1109/CVPR.2016.90

42. Huang G, Liu Z, van der Maaten L, Weinberger KQ. (2017). Densely connected convolutional networks, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, (Honolulu, HI, USA: IEEE). pp. 4700–8. doi: 10.1109/CVPR.2017.243

43. Yang H, Kim JY, Kim H, Adhikari SP. Guided soft attention network for classification of breast cancer histopathology images. *IEEE Trans Med Imag* (2020) 39:1306–15. doi: 10.1109/TMI.2019.2948026

44. Peng Y, Zhu W, Chen Z, Wang M, Geng L, Yu K, et al. Automatic staging for retinopathy of prematurity with deep feature fusion and ordinal classification strategy. *IEEE Trans Med Imag* (2021) 40:1750–62. doi: 10.1109/TMI.2021.3065753

45. Feng R, Liu X, Chen J, Chen DZ, Gao H, Wu J. A deep learning approach for colonoscopy pathology WSI analysis: accurate segmentation and classification. *IEEE J Biomed Health Informat* (2021) 25:3700–8. doi: 10.1109/JBHI.2020.3040269

46. Chen H, Li C, Li X, Rahaman MM, Hu W, Li Y, et al. IL-MCAM: An interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach. *Comput Biol Med* (2022) 143:105265. doi: 10.1016/j.compbiomed.2022.105265

47. Caron M, Bojanowski P, Joulin A, Douze M. (2018). Deep clustering for unsupervised learning of visual features, in: *Proc. European Conf. Comput. Vis. (ECCV)*, (Munich, Germany: Springer). pp. 132–49. doi: 10.1007/978-3-030-01264-9_9

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 08 December 2022
DOI 10.3389/fonc.2022.968784

OPEN ACCESS

EDITED BY
Shahid Mumtaz,
Instituto de Telecomunicações, Portugal

REVIEWED BY
Chong-Ke Zhao,
Fudan University, China
Abdul Rehman Javed,
Air University, Pakistan
Thippa Reddy Gadekallu,
VIT University, India

*CORRESPONDENCE
Chengliang Yin
chengliangyin@163.com
Wei Kang
kanve822@hotmail.com
Wenle Li
drlee0910@163.com

†These authors have contributed equally to this work

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

RECEIVED 14 June 2022
ACCEPTED 21 October 2022
PUBLISHED 08 December 2022

CITATION
Li W, Hong T, Fang J, Liu W, Liu Y, He C, Li X, Xu C, Wang B, Chen Y, Sun C, Li W, Kang W and Yin C (2022) Incorporation of a machine learning pathological diagnosis algorithm into the thyroid ultrasound imaging data improves the diagnosis risk of malignant thyroid nodules. *Front. Oncol.* 12:968784. doi: 10.3389/fonc.2022.968784

Incorporation of a machine learning pathological diagnosis algorithm into the thyroid ultrasound imaging data improves the diagnosis risk of malignant thyroid nodules

Wanying Li[1], Tao Hong[2†], Jianqiang Fang[3,4†], Wencai Liu[5†], Yuwen Liu[6†], Cunyu He[4], Xinxin Li[4], Chan Xu[4], Bing Wang[4], Yuanyuan Chen[7], Chenyu Sun[8], Wenle Li[9*], Wei Kang[10*] and Chengliang Yin[11*]

[1]Center for Management and Follow-up of Chronic Diseases, Xianyang Central Hospital, Xianyang, China, [2]Pediatric Surgery Ward, Fuwai Hospital Chinese Academy of Medical Sciences, Shenzhen, China, [3]Ultrasound Interventional Department, Xianyang Central Hospital, Xianyang, China, [4]Clinical Medical Research Center, Xianyang Central Hospital, Xianyang, China, [5]Department of Orthopaedic Surgery, The First Affiliated Hospital of Nanchang University, Nanchang, China, [6]Department of Chronic Disease and Endemic Disease Control Branch, Xiamen Municipal Center for Disease Control and Prevention, Xiamen, China, [7]School of Statistics, RENMIN University of China, Beijing, China, [8]AMITA Health Saint Joseph Hospital Chicago, Chicago, IL, United States, [9]State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics and Center for Molecular Imaging and Translational Medicine, School of Public Health, Xiamen University, Xiamen, China, [10]Department of Mathematics, Physics and Interdisciplinary Studies, Guangzhou Laboratory, Guangzhou, Guangdong, China, [11]Faculty of Medicine, Macau University of Science and Technology, Macao, Macao SAR, China

Objective: This study aimed at establishing a new model to predict malignant thyroid nodules using machine learning algorithms.

Methods: A retrospective study was performed on 274 patients with thyroid nodules who underwent fine-needle aspiration (FNA) cytology or surgery from October 2018 to 2020 in Xianyang Central Hospital. The least absolute shrinkage and selection operator (lasso) regression analysis and logistic analysis were applied to screen and identified variables. Six machine learning algorithms, including Decision Tree (DT), Extreme Gradient Boosting (XGBoost), Gradient Boosting Machine (GBM), Naive Bayes Classifier (NBC), Random Forest (RF), and Logistic Regression (LR), were employed and compared in constructing the predictive model, coupled with preoperative clinical characteristics and ultrasound features. Internal validation was performed by using 10-fold cross-validation. The performance of the model was measured by the area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, F1 score, Shapley additive explanations (SHAP) plot, feature importance, and correlation of features. The best cutoff

value for risk stratification was identified by probability density function (PDF) and clinical utility curve (CUC).

Results: The malignant rate of thyroid nodules in the study cohort was 53.2%. The predictive models are constructed by age, margin, shape, echogenic foci, echogenicity, and lymph nodes. The XGBoost model was significantly superior to any one of the machine learning models, with an AUC value of 0.829. According to the PDF and CUC, we recommended that 51% probability be used as a threshold for determining the risk stratification of malignant nodules, where about 85.6% of patients with malignant nodules could be detected. Meanwhile, approximately 89.8% of unnecessary biopsy procedures would be saved. Finally, an online web risk calculator has been built to estimate the personal likelihood of malignant thyroid nodules based on the best-performing ML-ed model of XGBoost.

Conclusions: Combining clinical characteristics and features of ultrasound images, ML algorithms can achieve reliable prediction of malignant thyroid nodules. The online web risk calculator based on the XGBoost model can easily identify in real-time the probability of malignant thyroid nodules, which can assist clinicians to formulate individualized management strategies for patients.

KEYWORDS

thyroid nodules, malignant, machine learning, predictive model, web calculator

Introduction

The incidence of sonographically detected thyroid nodules is increasing in individuals; approximately 50% to 68% can be detected in healthy individuals. Most of these nodules are benign and asymptomatic (1–3), and only about 8% to 16% are malignant nodules (4–6). Due to the complexity and diversity of thyroid nodules, it is challenging for doctors to distinguish which nodules harbor clinically relevant malignancies (7). For more than 30 years, ultrasound and fine-needle aspiration (FNA) cytology were the traditional diagnostic methods as the cornerstones in the clinical management of patients with thyroid nodules (8).

FNA provides the most effective and practical diagnostic information for evaluating whether a nodule is malignant to reach a definitive diagnosis, which has traditionally been used to meet this purpose (9, 10). However, approximately 50% of all biopsied nodules proved to be benign and grew indolent with non-aggressive behavior (6, 11, 12). Moreover, biopsies in one out of seven thyroid nodules may not yield final cytological results and usually require repeated biopsies or additional evaluation (13). Obviously, it is not cost-effective to submit all these lesions to FNA.

As a non-invasive, low-cost, and convenient technique for thyroid nodule detection, ultrasound is widely accepted as the preferred imaging method for the diagnosis and monitoring of thyroid nodules. Therefore, ultrasonography has been considered as having a greater role in determining the need for FNA and follow-up planning (2, 7). In order to improve the accuracy of ultrasound-based diagnosis, various available ultrasound-based risk stratification systems have already been proposed by many national and international thyroid associations, such as the ACR TIRADS, the French TIRADS, the Korea-TIRADS, and the EU-TIRADS (14). The most commonly used thyroid nodule classification system is the Thyroid Imaging Reporting and Data System (TIRADS) developed by the American College of Radiology (ACR). However, the limitations of these systems include that subjective assessment of nodules (15) is inferior to the personal judgment by experts (8) and that different classification systems for the same thyroid nodules may yield varying results (16), which cannot be ignored. There is an urgent need to develop an improved and reliable diagnostic method to distinguish benign and malignant thyroid nodules, which could help reduce the number of unnecessary biopsies or diagnostic surgery without jeopardizing the detection of clinically relevant malignant thyroid nodules.

A predictive model based on machine learning (ML) algorithms, designed to "learn" from clinical and sonographic datasets and predict the nature of thyroid nodules, is in some cases more robust than human experts (17), and as a result, ML algorithms have been widely used to classify thyroid nodules

objectively (18–20). However, previous studies have classified thyroid nodules by analyzing thyroid ultrasound images. The purpose of the present study was to develop ML-ed models for predicting malignant lumps based on the database of clinical characteristics and ultrasound features of thyroid nodules confirmed by pathological examination in Chinese populations. Compared with using only image analysis, our ML-ed predictive models not only integrated ultrasound features but also included clinical features of patients with thyroid nodules, which may be more comprehensive and convenient, especially for clinicians and patients. It can carry out individualized treatment and management based on the received ultrasound reports. The new model obtained could be used to predict the malignant risk of thyroid nodules in individuals online *via* a web calculator.

Materials and methods

The retrospective study followed the tenets of the Declaration of Helsinki and was approved by the Ethics Committee of Xianyang Central Hospital (No. 2022-IRB-68). All the study participants provided written informed consent, which waived the requirement for informed patient consent because data for all subjects were anonymized.

Collection of patients

A total of 9,999 consecutive patients with thyroid nodules who underwent FNA cytology or surgical procedure at Xianyang Central Hospital from the year 2000 to 2020 were included in our study. All included participants met the following inclusion criteria: 1) a single thyroid nodule with a diameter of 5–50 mm, 2) complete clinical and ultrasonic data, and 3) all nodules with definite pathological confirmation. The exclusion criteria were 1) undistinguishable coalescent thyroid lesions and 2) pathology provided ambiguous diagnostic findings for their nodules. Please see Figure S1.

Collection of ultrasound data

Ultrasound images of thyroid glands and the surrounding areas were acquired by ultrasound machine with a linear array probe at Xianyang Central Hospital. The ultrasound images were performed independently by two ultrasonologists, with a senior ultrasonologist making the final decision on controversial patients. The following features of each nodule, including the size, shape, composition, echogenicity, margin, echogenicity, and cervical lymph node status, were carefully measured and recorded. Images of the thyroid are obtained according to ACR accreditation standards. Ultrasound features were divided according to the ACR TIRADS (3), and each feature had a corresponding score. The higher the score, the greater the malignant tendency. In the processing of statistical analysis, the ultrasound characteristics of each nodule were replaced with the corresponding scores in the ACR TIRADS. For example, taller-than-wide will be assigned 3 points, so we wrote the Arabic numeral 3 instead of taller-than-wide in Table 1.

The benign and malignant pathology of all thyroid nodules in all participants was confirmed by FNA or surgery. All pathological results were examined blindly and separately by two pathologists, with a final decision made by a senior pathologist.

Analysis strategy

In order to maximize the predictive performance and ultimately reduce overfitting, we used the least absolute shrinkage and selection operator (lasso) regression analysis to screen variables, followed by logistic analysis to identify independent risk factors for malignant nodules.

A total of six ML algorithms were developed in this study, including Decision Tree (DT), Extreme Gradient Boosting (XGBoost), Gradient Boosting Machine (GBM), Naive Bayes Classifier (NBC), Random Forest (RF), and Logistic Regression (LR), to predict malignant thyroid nodules based on the variables with multivariable logistic regression p-value less than 0.05. Models have been validated internally by using 10-fold cross-validation. Subsequently, the area under the receiver operating characteristic curve (AUC) values, accuracy, precision, recall, and F1 score have been calculated to measure and compare the performance of each model.

Because many machine learning algorithms are considered functional black boxes, their internal processes are not well understood. Given this issue, various interpretability methods have been proposed to assess the influence of variables on the predicted results (21, 22). For instance, the relative importance of variables, the Shapley additive explanations (SHAP) method, and the heat map of the correlation of features were employed to further visualize the interpretation of ML-ed models at the feature level. An optimal cutoff value for clinical application was determined by probability density functions (PDFs). Clinical utility curves (CUCs) were plotted to compare the net benefits of different thresholds.

The demographic and clinical characteristics of all included patients were analyzed by t-test and chi-square test *via* SPSS Statistics software (version 26.0, SPSS Inc., Chicago, IL, USA). Continuous and categorical variables are expressed as mean ± SD and frequency in this study. p-Values <0.05 were considered statistically significant with 95% confidence intervals (CIs) applied for all analyses. R software was applied for developing predictive models *via* the "rms" package and establishing a web risk calculator *via* the "shiny" package.

Results

Clinical and ultrasound characteristics

Of all 9,999 participants with thyroid nodules, 500 (50%) harbored malignant nodules, while 500 (50%) had benign disease based on the pathological diagnosis. We collected clinical features (age and gender) and recorded image features (thyroid nodule location, size, shape, composition, echogenicity, margin, echogenicity, and cervical lymph node status).

Selection of features

Eight of 15 variables were screened by lasso analysis into logistic regression analysis, and all statistically significant factors in the univariate logistic regression analysis were included in the multivariate logistic regression analysis. Finally, age, margin, shape, echogenic foci, echogenicity, and lymph nodes were identified as independent predictors of thyroid cancer. There was no significant statistical difference in the nodule location (laterality) and composition in the differentiation of benign and malignant thyroid nodules. The results of the univariate and multivariate analyses are demonstrated in Table 1.

Demographic baseline

A cohort of patients from Xianyang Central Hospital in China was enrolled in this study. Results of the t-test and the chi-square test indicated there was no statistically significant difference between the training and the validation cohorts at a 0.05 significance level.

TABLE 1 The results of univariate and multivariable logistic regression.

Characteristics	Univariate logistic regression			Multivariable logistic regression		
	OR	CI	p	OR	CI	p
Age	0.97	0.95–0.99	0.002	0.97	0.95–1	0.027
Composition						
1	Ref	Ref	Ref	Ref	Ref	Ref
2	4.8	1–23.03	0.05	NA	NA	NA
Echogenic.Foci						
0	Ref	Ref	Ref	Ref	Ref	Ref
1	1.21	0.54–2.67	0.644	0.75	0.28–1.97	0.557
2	1.48	0.09–24.16	0.781	1.35	0.06–30.07	0.848
3	7.54	3.85–14.76	<0.001	4.12	1.87–9.05	<0.001
Echogenicity						
1	Ref	Ref	Ref	Ref	Ref	Ref
2	9.21	3.95–21.49	<0.001	4.72	1.76–12.68	0.002
3	3.81	0.54–27.08	0.181	3.89	0.47–32.19	0.208
Laterality						
Left	Ref	Ref	Ref	Ref	Ref	Ref
Right	0.91	0.56–1.48	0.7	NA	NA	NA
Middle	3.53	0.95–13.2	0.061	NA	NA	NA
Lymph.Nodes						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	6.9	2.8–16.98	<0.001	5.48	1.97–15.27	0.001
Margin						
0	Ref	Ref	Ref	Ref	Ref	Ref
2	6.83	3.31–14.08	<0.001	3.87	1.71–8.77	0.001
3	8.71	3.46–21.91	<0.001	4.61	1.64–12.95	0.004
Shape						
0	Ref	Ref	Ref	Ref	Ref	Ref
3	3.83	2.08–7.06	<0.001	2.86	1.38–5.93	0.005

Composition (1, mixed cystic and solid; 2, solid or almost completely solid). Echogenic.Foci (0, none or large comet-tail artifacts; 1, macrocalcifications; 2, peripheral (rim) calcifications; 3, punctate echogenic foci). Echogenicity (1, hyperechoic or isoechoic; 2, hypoechoic; 3, very hypoechoic). Margin (0, smooth or ill-defined; 2, lobulated or irregular; 3, extra-thyroidal extension). Shape (0, wider-than-tall; 3, taller-than-wide). NA, Not Available.

Development and validation of ML-ed models

The six predictors identified in the differentiation of malignant and benign thyroid nodules were used to construct ML-ed models, including LR, NBC, DT, RF, GBM, and XGBoost, respectively, for predicting malignant thyroid nodules. The predictive performance of the six ML-ed models is shown in Figure 1. The whole cohort used 10-fold cross-validation in this study. All models had shown good performance in predicting malignant nodules. Their AUC values of XGBoost, LR, NBC, DT, RF, and GBM were 0.829, 0.821, 0.825, 0.759, 0.821, and 0.822, respectively, in the 10-fold cross-validation. The XGBoost model indicated the best performance than any of the others. Meanwhile, the XGBoost model also achieved the highest accuracy of 0.65 and precision of 0.63, as shown in Figure 2. Thus, the XGBoost was identified as our final predictive model in this study.

Explanation of model

To further illustrate the models at the feature level, a SHAP summary diagram was plotted to demonstrate how these features affect the presence of malignant thyroid nodules. The SHAP values for each feature plotted for each sample are shown in Figure 3. We concluded that margin, shape, echogenic foci, echogenicity, and lymph nodes exerted negative effects on predicting the risk of malignant thyroid nodules, whereas the risk of malignancy increases with age.

Additionally, we ranked the importance of features in Figure 4 in order to explore the extent to which each independent risk factor contributed to the model. Although there were slight differences in the importance of each variable across models, the margin contributed most to the prediction of malignant nodules in most models. In the XGBoost model, the relative importance of variables decreased in the following order: margin, echogenic foci, lymph nodes, age, shape, and echogenicity. The correlation heat map indicated there was no linear correlation between the variables, and they harbor independent predictive power in clinical practice (Figure 5).

Application of model

As illustrated in Figure 6, we recommend a threshold probability of 51% as the optimum cutoff value for the probability of malignant nodules. In this situation, we could detect 85.6% (red area under the blue line) of malignant nodules, while the number of biopsy procedures for benign nodules would be reduced by 89.8% (yellow area under the red line) in Figure 7. Finally, in order to facilitate the practical application of the model in clinical work, we embedded the best predictive model into a web risk calculator (Figure 8) that can easily derive the probability and risk stratification of patients with malignant nodules in real time. Figure S1 shows the flow chart of our current study.

Discussion

As a highly prevalent disease, the incidence of thyroid nodules in China is 20%–35% (23), of which 7%–15% are malignant (1). It is a challenge for clinicians to distinguish malignant from benign nodules. Hence, we propose to develop a predictive model based on machine learning for assessing the malignancy risk of thyroid nodules in the Chinese population. To our knowledge, this is the first ML-ed predictive model to

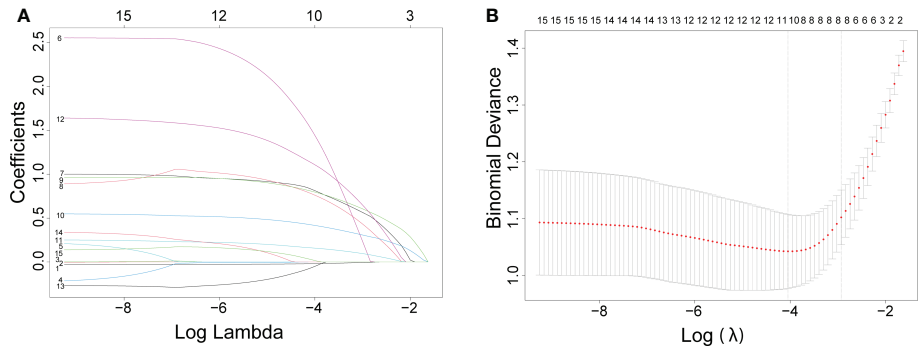


FIGURE 1
The results of the least absolute shrinkage and selection operator (lasso) regression analysis. The coefficients of all variables are reduced to 0 from instability to stability in **(A)** and obtain the model coefficient of λ value that minimizes the model deviation by cross-validation curve in **(B)**.

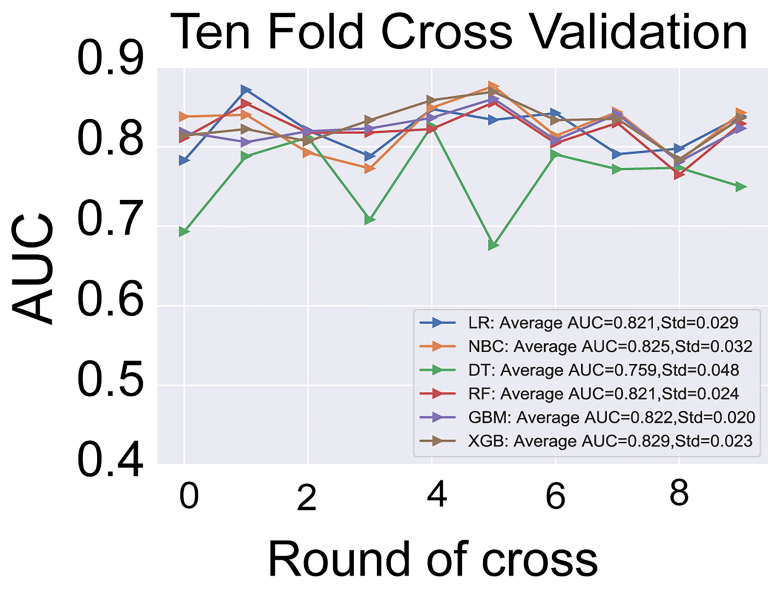


FIGURE 2
The results of 10-fold cross-validation in the six models of LR, NBC, DT, RF, GBM, and XGBoost. The average AUC of XGBoost model is the highest one. LR, Logistic Regression; NBC, Naive Bayes Classifier; DT, Decision Tree; RF, Random Forest; GBM, Gradient Boosting Machine; XGBoost, Extreme Gradient Boosting; AUC, area under the receiver operating characteristic curve.

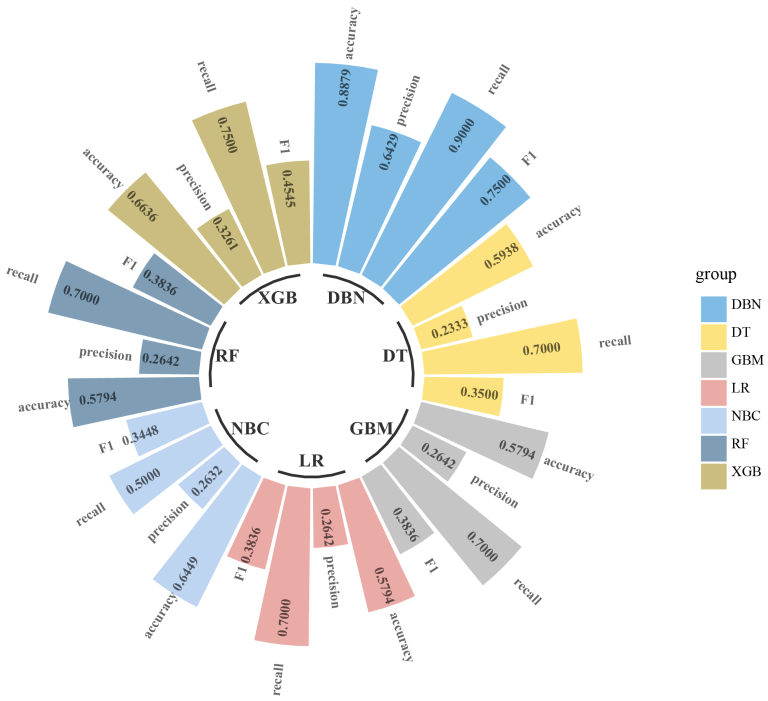


FIGURE 3
Circular bar plot. The performance of six models has been evaluated by five criteria of AUC, accuracy, precision, recall, and F1. AUC, area under the receiver operating characteristic curve.

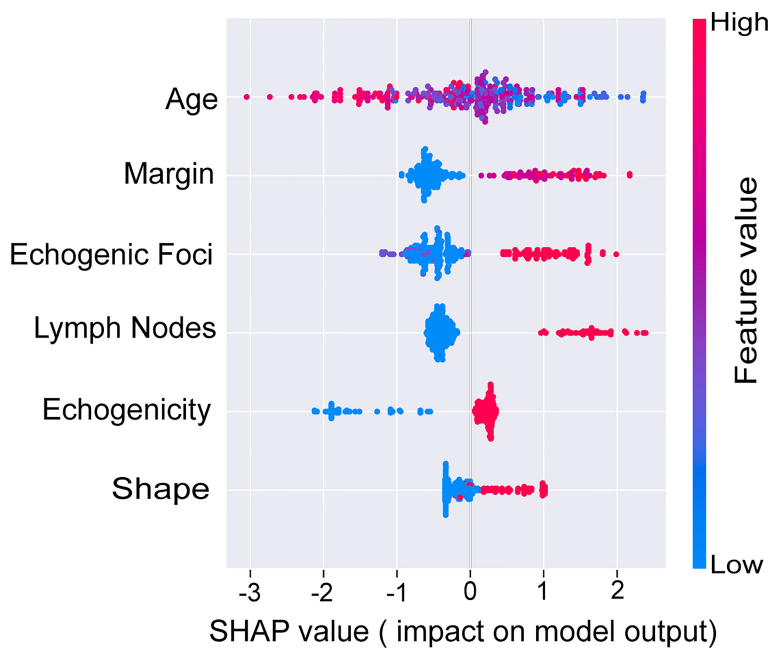


FIGURE 4
SHAP values of the selected features. The higher the SHAP value of each variable, the more impact and contribution to the model. SHAP, Shapley additive explanations.

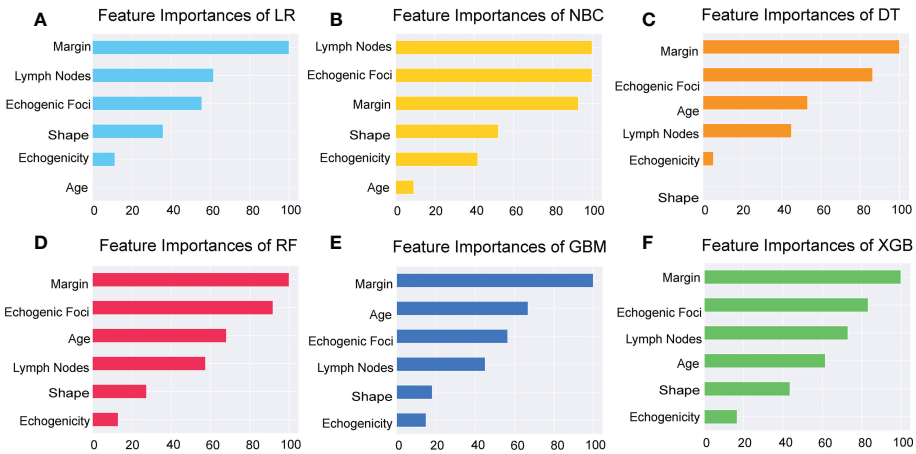


FIGURE 5
Importance of the selected features. Importance of each feature had been demonstrated and compared in the six models of LR, NBC, DT, RF, GBM, and XGBoost. LR, Logistic Regression; **(A)** NBC, Naive Bayes Classifier; **(B)** DT, Decision Tree; **(C)** RF, Random Forest; **(D)** GBM, Gradient Boosting Machine; **(E)** XGBoost, Extreme Gradient Boosting **(F)**.

predict malignant thyroid nodules by integrating clinical and ultrasound features. Our model was constructed using six variables, including clinical characteristics (age) and ultrasound features (margin, shape, echogenic foci, echogenicity, and lymph nodes). Our findings indicated that the proposed model could detect malignant thyroid nodules accurately and reduce unnecessary biopsies by estimating risk stratification. Finally, through a convenient and practical web application, our model can assist doctors and patients to carry out precise and individualized management of thyroid nodules.

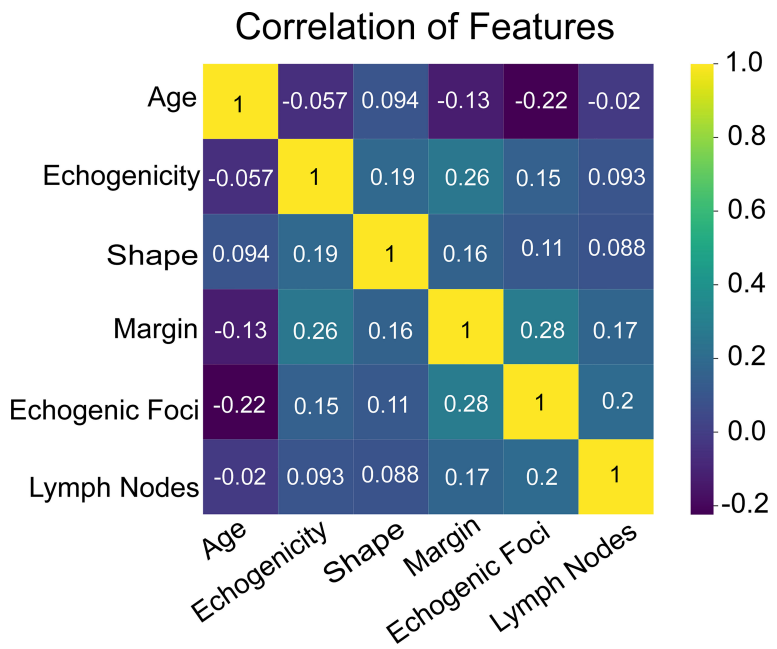


FIGURE 6
The heat map of correlation of the selected features, including age, echogenicity, shape, margin, echogenic foci, and lymph nodes.

We found that age played an adverse role in the risk of malignancy as a predictive parameter in our model. This finding was consistent with that of Chen et al., where there was no increased risk of malignancy in those aged 28–63 in the same population of a Chinese cohort (24). In addition, a similar finding was revealed in a US cohort of 196 patients with malignant FNA cytology, and the incidence of malignant thyroid nodules in patients under 45 years old was twice that in those over 45 years old (8.1% *vs.* 4.0%, $p < 0.001$) (25). Similarly, Italian scholars have reported that cytology suspicious or indicative of papillary thyroid cancer is associated with younger age (26). However, our result was contrary to what Belfiffiore et al. reported that thyroid cancer is more common in elderly patients (27). Different views on the association between age and the incidence of malignant thyroid nodules deserve further exploration.

Previous findings unveiled that the abnormal cervical lymph nodes may indicate malignant nodule metastasis (28), which is consistent with our study. It is reported that 30%–80% of patients with thyroid cancer have cervical lymph node metastasis (29). Cervical lymph nodes are usually not palpable because of their deep location and small size. Ultrasound has demonstrated its high sensitivity and specificity for the assessment of non-palpable lymph nodes (30). Cervical lymph nodes may enlarge as a result of a benign process, such as reactive hyperplasia due to inflammation in submandibular and upper cervical nodes (29). However, most investigators agree on the sonographic features of metastatic lymph nodes in thyroid cancer, including cystic degeneration, a rounded shape, loss of echogenic hilum, hypoechoic or hyperechoic mass, and calcification (31–33). The cervical lymph node is the first metastatic site of malignant nodules. Thyroid nodules should be highly suspected as malignant when abnormal lymph nodes are observed.

Meanwhile, margin ranked first in the importance of features and contributed the most to our model. We found that lobulated or irregular margins and extensive extrathyroidal extension detected by ultrasound increased the risk of malignancy risk in nodules. A lobulated or irregular margin is defined as a spiculated or jagged edge. Some studies have revealed that an irregular or microlobulated margin suggests malignancy (34, 35). Extensive extrathyroidal extension refers to a frank invasion of adjacent soft tissue or vascular, which is a highly reliable characteristic of malignancy and also has a negative effect on prognosis (36).

Furthermore, another feature we found that increased the risk of malignant nodules was the shape of the nodules. As first observed by Kim et al. and subsequently confirmed in a series of studies (35, 37–40), a lump with a shape taller-than-wide is another useful predictor of malignancy. These results may be associated with the growth pattern. It is found that the growth of benign nodules remains within normal tissue planes, so the shape of benign nodules can be ovoid to round, whereas malignant nodules grow centrifugally through the normal tissue plane (38, 39). In the ACR TIRADS, taller-than-wide was assigned 3 points in the TIRADS, and our results have

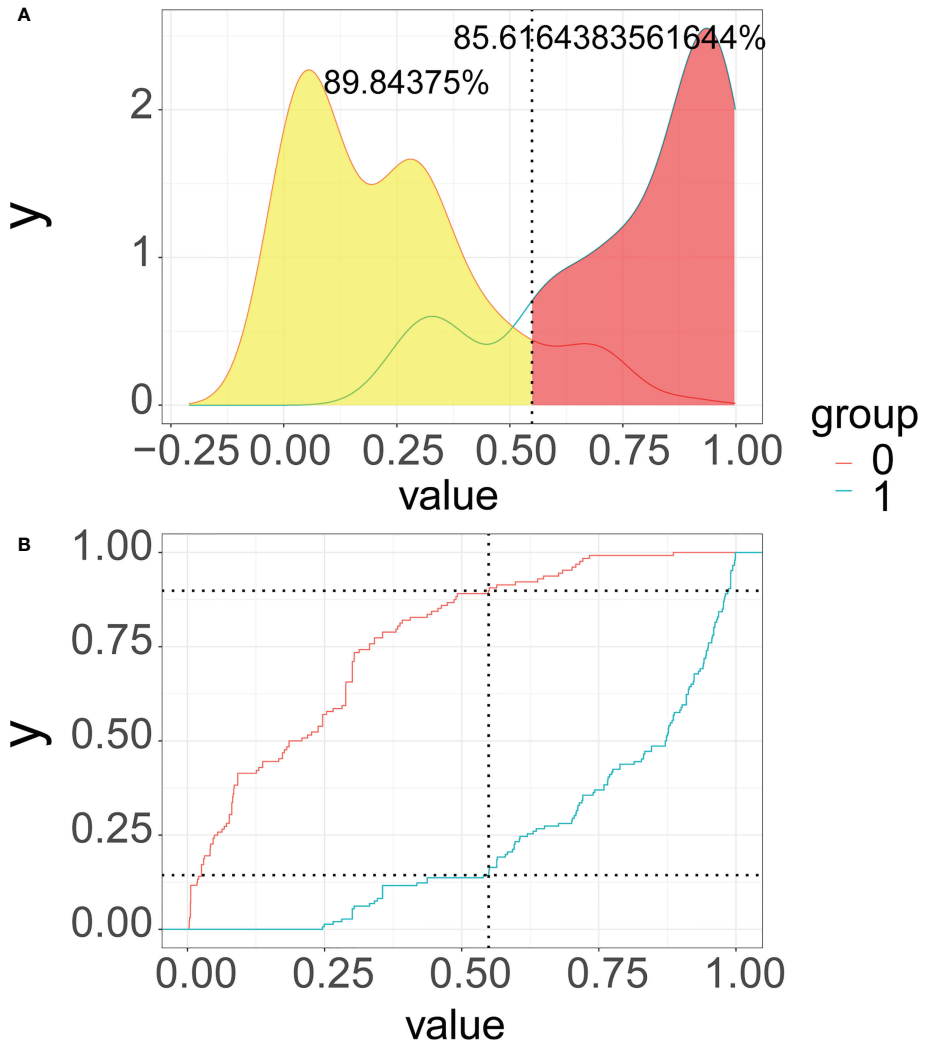


FIGURE 7
Probability density functions **(A)** and **c**linical utility curves **(B)** of the predictive model (0, benign modules group; 1, malignant nodules group). It is 85.6% (red area under the blue line) of malignant nodules, and the number of biopsy procedures for benign nodules was reduced by 89.8% (yellow area under the red line).

confirmed the high-risk role of shape in the malignancy of thyroid nodules.

Notwithstanding, echogenic foci, and echogenicity were regarded as predictive variables in both our predictive model and the ACR TIRADS; there were different opinions on assessing thyroid nodules on some features. Our results showed the presence of macrocalcifications and peripheral (rim) calcifications had no statistical difference between malignant and benign nodules. However, macrocalcifications and peripheral (rim) calcifications could be assigned 1 and 2 points in the TIRADS, respectively. Macrocalcifications refer to coarse echogenic foci accompanied by acoustic shadowing. Evidence in published data describing their correlation with an increased malignancy risk is weak (41); additionally, the relationship between macrocalcifications and nodules lacking other malignant characteristics is also mixed (42, 43). Peripheral (rim) calcifications lie along all or part of the nodule's margin. Compared with macrocalcifications, they are more strongly correlated with malignancy (41), but several studies suggested that their correlation with malignancy is variable (43).

The statistical results of logistic regression showed that patients with punctate echogenic foci had a higher tendency to develop malignant nodules. Punctate echogenic foci are smaller in size and less shadowed than macrocalcifications and may correspond to the psammomatous calcifications associated with papillary cancers in the solid components of thyroid nodules. Histologically, punctate echogenic foci are smaller and less shadowed than macrocalcifications, which are considered

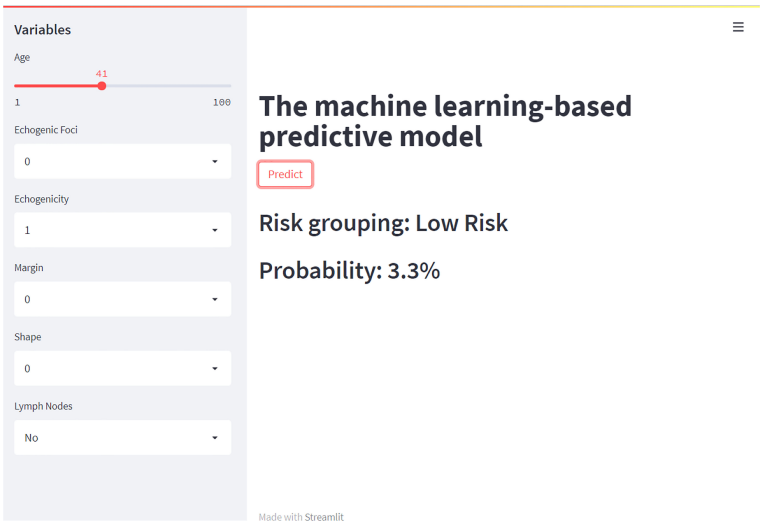


FIGURE 8
The application of the web risk calculator for patients with malignant nodules.

highly positive associations with malignancy, especially in combination with other suspicious features (3, 5).

Echogenicity refers to a nodule's reflectivity relative to adjacent tissue. Except for the thyroid parenchyma, which is usually used as reference tissue, the neck strap muscles with very low echogenicity are also used as the basis for comparison. Previously, several studies investigated that a higher degree of hypoechogenicity was highly suggestive of malignancy, with a specificity of 92%–94% (37, 44). Interestingly, a higher degree of hypoechogenicity harbors no statistical significance for predicting malignant nodules based on our results of multivariable logistic regression. These results need more evidence to verify.

Another important finding of our paper was composition, which was not an independent predictor of malignant nodules. Thyroid nodules that are cystic or almost completely cystic have no score in the ACR TIRADS because they are highly correlated with benign cytology, and only 13%–26% of thyroid cancers harbor a cystic component (29, 44–46). Spongiform, composed predominantly (>50%) of small cystic spaces, is considered a sign of benignity with high specificity (44). In our study, we found that no patient had cystic or almost completely cystic or spongiform ultrasound features. Additionally, according to ACR TIRADS, mixed cystic and solid, and solid or almost completely solid are the risk factors for malignant nodules, with scores of 1 and 2, respectively (3). Solid nodules with an eccentric configuration and acute angle are suspected to be malignant (47), whereas these conclusions were not observed in our study.

The differences between our study and the risk stratification system also illustrate the inadequacy of the classification system to evaluate thyroid nodules, such as interobserver variation/a subjective assessment of the nodules (8, 48). Therefore, it is necessary to add clinical data to further improve the accuracy and objectivity of the predictive model. As Chen et al. described in their literature, the predictive power of the new model was superior to that of ACR TIRADS when age is included. Furthermore, our study cohort enrolled retrospectively Chinese patients from a single medical center; these differences therefore may be due to demographic differences and healthcare disparities between patients in the USA and China.

Therefore, these differences may be due to the mismatch between the classification system and the current medical situation in China. It is more rational to apply a risk stratification system according to the population. Accordingly, Zhou et al. formulated Chinese guidelines for ultrasound malignancy risk stratification of thyroid nodules (C-TIRADS) that are specific to China's national and medical conditions (23).

We offered clinicians and patients an online web application for estimating the risk of a malignant thyroid nodule using the XGBoost model, which combined six variables including age, lymph nodes, margin, shape, echogenic foci, and echogenicity. By inputting the corresponding personalized parameters of patients, visitors can quickly obtain the corresponding malignancy risk. The link is as follows: https://share.streamlit.io/liuwencai6/thyroid_final/main/thyroid_final.py.

Depending on the cutoff value in the PDF and CUC, we recommended 51% as the threshold probability of the next management strategy and risk stratification. In that case, about 85.6% of patients with malignant thyroid nodules can be detected, FNA was recommended, and careful follow-up and possibly early surgery should be considered. Moreover, we could also save approximately 89.8% of unnecessary biopsy procedures in low-risk populations (malignant risk ≤51%). This result is

consistent with the main goal of all currently available sonographic risk stratification systems, that is, to eliminate unnecessary thyroid biopsies without endangering the diagnosis of clinically malignant nodules (15). We believe that the incorporation of our predictive model into clinical practice will improve the diagnostic accuracy of malignant thyroid nodules and minimize the number of unnecessary FNA in low-risk patients with thyroid nodules. Compared with existing models (49, 50), the performances are good but various due to differences in population and dataset.

There are several limitations of this existing study. First, the retrospective nature of this study may have resulted in potential bias. Second, the ultrasound features were read and provided by sonographers rather than captured directly from the ultrasonic images, which may cause bias in data quality due to extraction and interpretation. We strongly recommend that the machine learning model be used to extract the features from ultrasonic images directly and from several types of machines in future studies. In addition, all patients and ultrasound assessments are derived from a single medical center, which may restrict the accuracy; large-scale multicenter cohorts and external validation would be more forceful. Finally, we classified the features of nodules by ACR TIRADS, rather than the Chinese-TIRADS proposed by the Chinese professional society, to evaluate ultrasound parameters (23). We will expand our cohort and dataset in a further study to optimize our model and algorithm in the future.

Conclusion

In conclusion, our study yielded a machine learning-based model combining age with ultrasound parameters, including shape, margin, echogenic foci, echogenicity, and lymph nodes, to predict the presence of malignant thyroid nodules. Our model showed good performance and was embodied in a web risk calculator to estimate the risk of malignant thyroid nodules.

Data availability statement

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

Ethics statement

This study has been approved by the Ethics Committee of Xianyang Central Hospital, Ethics No. 2022-IRB-68.

Author contributions

CLY, WK, and WLL designed the article. YWL, CX, and BW collected and evaluated the data. WYL and WLL wrote the first draft of the manuscript. All authors reviewed the manuscript. All authors contributed to the interpretation of the results. WYL, JQF and WCL wrote the final draft of the manuscript. YWL CYH and YYC read and approved the final version of the manuscript. All authors contributed to the article and approved the submitted version.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.968784/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Flow Chart.

References

1. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* (2016) 26(1):1–133. doi: 10.1089/thy.2015.0020

2. Gharib H, Papini E, Garber JR, Face D, Face R, Hegedus L, et al. American Association of clinical endocrinologists, American college of endocrinology, and associazione Medici endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules–2016 update. *Endocr Pract* (2016) 22(5):622–39. doi: 10.4158/EP161208.GL

3. Tessler FN, Middleton WD, Grant EG, William D, Grant Edward G. ACR thyroid imaging, reporting and data system (TI-RADS): White paper of the ACR TI-RADS committee. *J Am Coll Radiol* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046

4. Burman KD, Wartofsky L. CLINICAL PRACTICE. thyroid nodules. *N Engl J Med* (2015) 373(24):2347–56. doi: 10.1056/NEJMcp1415786

5. Rago T, Vitti P. Risk stratification of thyroid nodules: From ultrasound features to TIRADS. *Cancers (Basel)* (2022) 14(3):717. doi: 10.3390/cancers14030717

6. Smith-Bindman R, Lebda P, Feldstein VA, Sellami D, Goldstein RB, Brasic N, et al. Risk of thyroid cancer based on thyroid ultrasound imaging characteristics: results of a population-based study. *JAMA Intern Med* (2013) 173(19):1788–96. doi: 10.1001/jamainternmed.2013.9245

7. Lamartina L, Deandreis D, Durante C, Filetti S. ENDOCRINE TUMOURS-imaging in the follow-up of differentiated thyroid cancer- current evidence and future perspectives for a risk-adapted approach. *Eur J Endocrinol* (2016) 175(5): R185–202. doi: 10.1530/EJE-16-0088

8. Solymosi T, Hegedus L, Bonnema SJ, Frasoldati A, Jambor L, Kovacs GL, et al. Ultrasound-based indications for thyroid fine-needle aspiration: Outcome of a TIRADS-based approach versus operators' expertise. *Eur Thyroid J* (2021) 10 (5):416–24. doi: 10.1159/000511183

9. Gharib H, Papini E, Paschke R, Duick DS, Valcavi R, Hegedüs L, et al. American Association of clinical endocrinologists, associazione Medici endocrinologi, and EuropeanThyroid association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. *Endocr Pract* (2010) 16 Suppl 1:1–43. doi: 10.4158/EP.16.3.468

10. Singh Ospina N, Brito JP, Maraka S, Espinosa de Ycaza AE, Rodriguez-Gutierrez R, Gionfriddo MR, et al. Diagnostic accuracy of ultrasound-guided fine needle aspiration biopsy for thyroid malignancy: systematic review and meta-analysis. *Endocrine* (2016) 53(3):651–61. doi: 10.1007/s12020-016-0921-x

11. Davies L, Welch HG. Current thyroid cancer trends in the united states. *JAMA Otolaryngol Head Neck Surg* (2014) 140(4):317–22. doi: 10.1001/jamaoto.2014.1

12. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda system for reporting thyroid cytopathology: a meta-analysis. *Acta Cytol* (2012) 56(4):333–9. doi: 10.1159/000339959

13. Ali SZ, Siperstein A, Sadow PM, Golding AC, Kennedy GC, Kloos RT, et al. Extending expressed RNA genomics from surgical decision making for cytologically indeterminate thyroid nodules to targeting therapies for metastatic thyroid cancer. *Cancer Cytopathol* (2019) 127(6):362–9. doi: 10.1002/cncy.22132

14. Swan KZ, Thomas J, Nielsen VE, Jespersen ML, Bonnema SJ. External validation of AIBx, an artificial intelligence model for risk stratification, in thyroid nodules. *Eur Thyroid J* (2022) 11(2):e210129. doi: 10.1530/ETJ-21-0129

15. Grani G, Lamartina L, Ascoli V, Bosco D, Biffoni M, Giacomelli L, et al. Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: Toward the "Right" TIRADS. *J Clin Endocrinol Metab* (2019) 104(1):95–102. doi: 10.1210/jc.2018-01674

16. Huang BL, Ebner SA, Makkar JS, Bentley-Hibbert S, McConnell RJ, Lee JA , et al. A multidisciplinary head-to-Head comparison of American college of radiology thyroid imaging and reporting data system and American thyroid association ultrasound risk stratification systems. *Oncologist* (2020) 25(5):398–403. doi: 10.1634/theoncologist.2019-0362

17. Oberije C, Nalbantov G, Dekker A, Boersma L, Borger J, Reymen B, et al. A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: a step toward individualized care and shared decision making. *Radiother Oncol* (2014) 112(1):37–43. doi: 10.1016/j.radonc.2014.04.012

18. Buda M, Wildman-Tobriner B, Hoang JK, Thayer D, Tessler FN, Middleton WD, et al. Management of thyroid nodules seen on US images: Deep learning may match performance of radiologists. *Radiology* (2019) 292(3):695–701. doi: 10.1148/radiol.2019181343

19. Thomas J, Haertling T. AIBx, artificial intelligence model to risk stratify thyroid nodules. *Thyroid* (2020) 30(6):878–84. doi: 10.1089/thy.2019.0752

20. Guan Q, Wang Y, Du J, Qin Y, Lu H, Xiang J, et al. Deep learning based classification of ultrasound images for thyroid nodules: a large scale of pilot study. *Ann Transl Med* (2019) 7(7):137. doi: 10.21037/atm.2019.04.34

21. Souza LR, Colonna JG, Comodaro JM, Naveca FG. Using amino acids co-occurrence matrices and explainability model to investigate patterns in dengue virus proteins. *BMC Bioinf* (2022) 23(1):80. doi: 10.1186/s12859-022-04597-y

22. Lundberg S, Lee SI. A unifed approach to interpreting model predictions. *Adv Neural Inf Process Syst* (2017) 30:4765–74. doi: 10.5555/3295222.3295230

23. Zhou J, Yin L, Wei X, Zhang S, Song Y, Luo B, et al. Chinese Guidelines for ultrasound malignancy risk stratification of thyroid nodules: the c-TIRADS. *Endocrine* (2020) 70(2):256–79. doi: 10.1007/s12020-020-02441-y

24. Chen L, Zhang J, Meng L, Lai Y, Huang W. A new ultrasound nomogram for differentiating benign and malignant thyroid nodules. *Clin Endocrinol (Oxf)* (2019) 90(2):351–9. doi: 10.1111/cen.13898

25. Bessey LJ, Lai NB, Coorough NE, Chen H, Sippel RS. The incidence of thyroid cancer by fine needle aspiration varies by age and gender. *J Surg Res* (2013) 184(2):761–5. doi: 10.1016/j.jss.2013.03.086

26. Rago T, Fiore E, Scutari M, Santini F, Di Coscio G, Romani R, et al. Male Sex, single nodularity, and young age are associated with the risk of finding a papillary thyroid cancer on fine-needle aspiration cytology in a large series of patients with nodular thyroid disease. *Eur J Endocrinol* (2010) 162(4):763–70. doi: 10.1530/EJE-09-0895

27. Belfiore A, La Rosa GL, La Porta GA, Giuffrida D, Milazzo G, Lupo L, et al. Cancer risk in patients with cold thyroid nodules: relevance of iodine uptake, sex, age, and multinodularity. *Am J Med* (1992) 93:363–9. doi: 10.1016/0002-9343(92)90164-7

28. AIUM practice parameter for the performance of a thyroid and parathyroid ultrasound examination. *J Ultrasound Med* (2016) 35:1–11. doi: 10.7863/ultra.35.9.1-c

29. Patel NU, McKinney K, Kreidler SM, Bieker TM, Russ P, Roberts K, et al. Ultrasound-based clinical prediction rule model for detecting papillary thyroid cancer in cervical lymph nodes: A pilot study. *J Clin Ultrasound* (2016) 44(3):143–51. doi: 10.1002/jcu.22309

30. Snozek CL, Chambers EP, Reading CC, Sebo TJ, Sistrunk JW, Singh RJ, et al. Serum thyroglobulin, high-resolution ultrasound, and lymph node thyroglobulin in diagnosis of differentiated thyroid carcinoma nodal metastases. *J Clin Endocrinol Metab* (2007) 92(11):4278–81. doi: 10.1210/jc.2007-1075

31. Sohn YM, Kwak JY, Kim EK, Moon HJ, Kim SJ, Kim MJ. Diagnostic approach for evaluation of lymph node metastasis from thyroid cancer using ultrasound and fine-needle aspiration biopsy. *AJR Am J Roentgenol* (2010) 194 (1):38–43. doi: 10.2214/AJR.09.3128

32. Langer JE, Mandel SJ. Sonographic imaging of cervical lymph nodes in patients with thyroid cancer. *Neuroimaging Clin N Am* (2008) 18(3):479–489, vii-viii. doi: 10.1016/j.nic.2008.03.007

33. Kamaya A, Gross M, Akatsu H, Jeffrey RB. Recurrence in the thyroidectomy bed: sonographic findings. *AJR Am J Roentgenol* (2011) 196(1):66–70. doi: 10.2214/AJR.10.4474

34. Papini E, Guglielmi R, Bianchini A, Crescenzi A, Taccogna S, Nardi F, et al. Risk of malignancy in nonpalpable thyroid nodules: Predictive value of ultrasound and color-Doppler features. *J Clin EndocrinolMetab* (2002) 87:941–1946. doi: 10.1210/jcem.87.5.8504

35. Kim EK, Park CS, Chung WY, Oh KK, Kim DI, Lee JT, et al. New sonographic criteria for recommending fine-needle aspiration biopsy of nonpalpable solid nodules of the thyroid. *AJR Am J Roentgenol* (2002) 178 (3):687–91. doi: 10.2214/ajr.178.3.1780687

36. Hoang JK, Lee WK, Lee M, Johnson D, Farrell S. US Features of thyroid malignancy: pearls and pitfalls. *Radiographics* (2007) 27:847–60. doi: 10.1148/rg.273065038

37. Moon WJ, Jung SL, Lee JH, Na DG, Baek JH, Lee YH, et al. Thyroid study group, Korean society of neuro- and head and neck radiology. benign and malignant thyroid nodules: US differentiation–multicenter retrospective study. *Radiology* (2008) 247(3):762–70. doi: 10.1148/radiol.2473070944

38. Cappelli C, Castellano M, Pirola I, Gandossi E, De Martino E, Cumetti D, et al. Thyroid nodule shape suggests malignancy. *Eur J Endocrinol* (2006) 155 (1):27–31. doi: 10.1530/eje.1.02177

39. Alexander EK, Marqusee E, Orcutt J, Benson CB, Frates MC, Doubilet PM, et al. Thyroid nodule shape and prediction of malignancy. *Thyroid* (2004) 14 (11):953–8. doi: 10.1089/thy.2004.14.953

40. Na DG, Baek JH, Sung JY, Kim JH, Kim JK, Choi YJ, et al. Thyroid imaging reporting and data system risk stratification of thyroid nodules: Categorization based on solidity and echogenicity. *Thyroid* (2016) 26(4):562–72. doi: 10.1089/thy.2015.0460

41. Arpaci D, Ozdemir D, Cuhaci N, Dirikoc A, Kilicyazgan A, Guler G, et al. Evaluation of cytopathological findings in thyroid nodules with macrocalcification: macrocalcification is not innocent as it seems. *Arq Bras Endocrinol Metabol* (2014) 58(9):939–45. doi: 10.1590/0004-2730000003602

42. Na DG, Kim DS, Kim SJ, Ryoo JW, Jung SL. Thyroid nodules with isolated macrocalcification: malignancy risk and diagnostic efficacy of fine-needle aspiration and core needle biopsy. *Ultrasonography* (2016) 35(3):212–9. doi: 10.14366/usg.15074

43. Kim MJ, Kim EK, Kwak JY, Park CS, Chung WY, Nam KH, et al. Differentiation of thyroid nodules with macrocalcifications: role of suspicious sonographic findings. *J Ultrasound Med* (2008) 27(8):1179–84. doi: 10.7863/jum.2008.27.8.1179

44. Bonavita JA, Mayo J, Babb J, Bennett G, Oweity T, Macari M, et al. Pattern recognition of benign nodules at ultrasound of the thyroid: which nodules can be left alone? *AJR Am J Roentgenol* (2009) 193(1):207–13. doi: 10.2214/AJR.08.1820

45. Chan BK, Desser TS, McDougall IR, Weigel RJ, Jeffrey RBJr. Common and uncommon sonographic features of papillary thyroid carcinoma. *J Ultrasound Med* (2003) 22:1083–90. doi: 10.7863/jum.2003.22.10.1083

46. Russ G. Risk stratification of thyroid nodules on ultrasonography with the French TI-RADS: description and reflections. *Ultrasonography* (2016) 35(1):25–38. doi: 10.14366/usg.15027

47. Kim DW, Park JS, In HS, Choo HJ, Ryu JH, Jung SJ. Ultrasound-based diagnostic classification for solid and partially cystic thyroid nodules. *AJNR Am J Neuroradiol* (2012) 33(6):1144–9. doi: 10.3174/ajnr.A2923

48. Persichetti A, Di Stasio E, Coccaro C, Graziano F, Bianchini A, Di Donna V, et al. Inter- and intraobserver agreement in the assessment of thyroid nodule ultrasound features and classification systems: A blinded multicenter study. *Thyroid* (2020) 30(2):237–42. doi: 10.1089/thy.2019.0360

49. Zhao CK, Ren TT, Yin YF, Shi H, Wang HX, Zhou BY, et al. A comparative analysis of two machine learning-based diagnostic patterns with thyroid imaging reporting and data system for thyroid nodules: diagnostic performance and unnecessary biopsy rate. *Thyroid* (2021) 31(3):470–81. doi: 10.1089/thy.2020.0305

50. Zhang B, Tian J, Pei S, Chen Y, He X, Dong Y, et al. Machine learning–assisted system for thyroid nodule diagnosis. *Thyroid* (2019) 29(6):858–67. doi: 10.1089/thy.2018.0380

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact