

Research Article

Duan Yuanchuan, Diao Hang, Li Shi*, Liu Kailin, and Feng Yijie

Auxiliary diagnosis study of integrated electronic medical record text and CT images

<https://doi.org/10.1515/jisys-2022-0040>

received December 06, 2021; accepted February 14, 2022

Abstract: At present, most of the research in the field of medical-assisted diagnosis is carried out based on image or electronic medical records. Although there is some research foundation, they lack the comprehensive consideration of comprehensive image and text modes. Based on this situation, this article proposes a fusion classification auxiliary diagnosis model based on GoogleNet model and Bi-LSTM model, uses GoogleNet to process brain computed tomographic (CT) images of ischemic stroke patients and extract CT image features, uses Bi-LSTM model to extract the electronic medical record text, integrates the two features using the full connection layer network and Softmax classifier, and obtains a method that can assist the diagnosis from two modes. Experiments show that the proposed scheme on average improves 3.05% in accuracy compared to individual image or text modes, and the best performing GoogleNet + Bi-LSTM model achieves 96.61% accuracy; although slightly less in recall, it performs better on *F1* values, and has provided feasible new ideas and new methods for research in the field of multi-model medical-assisted diagnosis.

Keywords: GoogleNet model, Bi-LSTM model, fusion features, medical-assisted diagnosis

1 Introduction

At present, the application of artificial intelligence in medical data is becoming more and more extensive, and has made many achievements, mainly focusing on the prediction of disease, auxiliary diagnosis, prognosis evaluation of disease, new drug research and development, health management, medical image recognition and other aspects [1]. The major goal of AI applications in health care is to investigate the relationship between preventative or therapeutic options and health experience. For the diagnosis and therapy of illness conditions, Intelligent systems can save cash and effort, enabling public healthcare increasingly efficient and productive. Among them, medical-assisted diagnosis is the most popular direction. The diagnostic framework comprises components that each supply descriptive statistics toward the network separately from a patient's medical examinations as well as numerous diagnostic procedures. It provides doctors with diagnostic reference decisions by analyzing patient records and sociohistorical data. An important part of doctors' learning and growth is to constantly observe experiences and lessons on the basis of old cases, so as to improve the accuracy of existing cases. The computer auxiliary diagnosis (CAD) technology, which is centred on medical imaging, aids in speedy identification and gives doctors legitimate "feedback." The goal of this study is to classify and evaluate the CAD approach for identifying ischemic stroke using imaging techniques. Therefore, a doctor often has to pass many years of learning practice, attend a lot of cases to know the disease, but the uneven distribution of medical resources limits the growth

* **Corresponding author: Li Shi**, College of Information and Computer Engineering, Northeast Forestry University, 150040, Harbin City, Heilongjiang Province, China, e-mail: bill20202222@163.com

Duan Yuanchuan, Diao Hang, Liu Kailin, Feng Yijie: College of Information and Computer Engineering, Northeast Forestry University, 150040, Harbin City, Heilongjiang Province, China

of grassroots and young doctors. Medical auxiliary diagnosis system through simulating the process of doctor learning, using the method of artificial intelligence, makes them learn a lot of cases and provide medical professional knowledge, so that can provide an auxiliary reference for doctors [2].

With the development of artificial intelligence (AI) technology and its application in computer vision, imaging, and natural language processing (NLP), AI has been widely used in medical-assisted diagnosis, has been proved to provide fast and accurate detection and diagnosis, mostly based on manually extracted features, automatic learning, and achieved excellent results in disease classification, lesion segmentation and disease prediction [3]. NLP tools that can comprehend and categorize clinical documents are a prominent utilization of AI technology in medicine. NLP algorithms can evaluate unorganized clinical information on individuals, providing invaluable information on accuracy, technique improvement, and improved health outcomes. A scanner is a type of all the data included in a supplier's patient's chart: health records, diagnosis, prescriptions, vaccination schedules, sensitivities, test findings, and physician's comments. There is also a lot of research on electronic medical record-assisted diagnosis based on an AI system, through the records of patients from admission diagnosis to discharge, electronic medical records-based auxiliary diagnosis, label classification and intelligent prescribing [4]. Socialized medicines through increasing all elements of treating patients, particularly security, efficacy, patient-centeredness, collaboration, learning, speed, productivity, as well as equality, are all advantages of electronic health records. Limitations – Holding confidentiality of patient information in the cloud, since many EMRs do, exposes the information to hacking lacking adequate safety precautions.

At present, the auxiliary intelligent diagnosis of electronic medical record text and computed tomographic (CT) imaging has developed relatively mature, but not all diseases can rely only on a single mode for diagnosis. Therefore, the patient history should be taken into account. The disease-assisted diagnosis with multi-model is making little progress due to technical limitations, data separation, and other reasons. Multimodal learning is a useful approach for representing the combined interpretations of several dimensions. The variety of learning paradigms is made up of two deep Boltzmann networks, one for every sensory. To create the combined depiction, an extra hidden level is inserted on top of the two Boltzmann Machines. In recent years, due to the improvement of deep learning and computer computing ability, the multi-modal learning method of image and text fusion have been developed rapidly. Whenever a lot of our senses – visual, aural, and kinaesthetic – are activated throughout the acquisition, we comprehend and retain longer, according to multimedia learning theory. Teachers experienced knowledge in a number of ways when such methods are combined, resulting in a diversified educational experience. Its benefits are involving all active learning, developing learners' skills in all mediums, enhancing the effectiveness of learning, illustrating the complexity of direct conversation, and so on. Its application in the field of medical diagnosis can broaden the application scope of medically assisted diagnosis, improve the diagnosis effect of the current system, reduce the burden of doctors, and provide new ideas for other researchers [5]. Stochastic Gradient Descent with Momentum is used to develop the Convolutional Neural Network (CNN). To classify images, the system comprises a neural network, three recurrent as well as averaging convolution layers, and a softmax densely integrated output neuron.

This article with ischemic stroke as the research object shows that it is one of the first cause of illness and disability, in 2013. The national and section epidemiological investigation shows that of more than 3 million cases, about 2 million cases died of stroke, including ischemic stroke accounted for about 70% of all stroke types; because of its complex diagnosis; rapid evaluation, diagnosis, and intervention are important to improve the outcome of ischemic stroke patients, so improving the quality of ischemic stroke care is of great significance.

The main contributions of this article are summarized as follows:

This article proposes a neural network model integrating GoogleNet model and bidirectional long-short term memory (Bi-LSTM) model. The GoogleNet model can extract the image feature information about the studied disease in the CT image. The Bi-LSTM model can extract the relevant semantic features in the medical record text. The two extracted features are then fused at the feature level, Classification to assist diagnosis.

This article comprehensively analyzes the characteristics of related diseases from the perspective of multi-modal mode, selects the medical record text-related items to associate with the corresponding patient CT image, establishes the ischemic stroke data set, and classifies the research model on the patient data set. Experimental results show that the proposed model achieves good results and outperforms the baseline model.

2 Related work

In hospitals, a large amount of patient information is entered into the medical information system every day. Medical information is an important data source for the hospital, and the auxiliary diagnosis problem is based on data and makes use of machine learning to predict results of an undiagnosed case. Interior documentation is kept in a standardized template for three years after therapy begins. A patient's or approved operator's demand for medical information must be recognized as well as documentation supplied within 72 h. As early as 2005, Abe et al. [6] proposed a time-series-based medical data mining system, which integrates relevant pattern recognition methods, rule introduction methods, and rule evaluation methods, and provides visual operation. Predicting medication, customer services, fraudulent and misuse identification, medical administration, and analyzing the success of various therapies are all domains where business intelligence has shown to be useful in medicine. At least once, an analyst must examine the entire or portion of the incoming parameters in order to acquire this statistical model. A learning algorithm constructs the rule evaluation model upon collecting the experimental original data. A practitioner provides forecasts for proposed regulations depending on the principles of the target indicators during the classification stage. At present, thanks to the improvement of computer computing ability and the development of deep learning methods, the performance of the auxiliary diagnosis system has also been greatly improved. We created the rule evaluation integrates additional using rule evaluation models, which are produced using quantitative metrics of mining classifier as well as human expert ratings for every rule, to lower the expenses of rule assessment operations.

In the field of medical images, great results have been achieved in the classification and lesion identification of pathological images using deep learning methods. In terms of pathological image classification, Talo et al. [7] use neural networks with a pre-trained ResNet34 architecture to multi-classify MR images in the brain. Migration learning enables it to train the model with a relatively small amount of data. They are pre-trained with the ImageNet database. Using a total of 613 brain MR images from the Harvard Medical School's website for training and verification, the accuracy rate reached 97.38%; Mashmoud et al. [8] sent a technology based on brain tumor classification by CNN, hold-out testing "evaluation on motion-corrected hold-out testing" images, high accuracy is achieved, the accuracy is 97.8%, the specificity is 99.2%, and the sensitivity is 97.32%.

In terms of pathological image segmentation, Zhou et al. [9] overfit the dynamic changes of actual patients measuring data at different points in time, greatly alleviate the data scarcity problem, and propose a fully automatic, fast, accurate, and machine agnostic method that can divide and quantify infected areas, dice reached 90.3%, and recall reached 89.8%. Huang et al. [10] investigate a novel diagnostic method based on deep transfer Convolutional neural networks (DTCNN) and limit learning machine (ELM) that incorporates the synergy of the two algorithms to handle benign and malignant nodule classification, first using the optimal DTCNN to extract advanced features of pulmonary nodules, which have been trained in advance through the ImageNet data set. Because there is a lack of large data sets in clinical categorization as well as developing CNN is time-consuming, infrastructure deep transfer learning is employed to maximize the productivity of learning CNN models with a constrained quantity of labeled data. Unlike ANN, ELM starts the hidden layer settings at randomness and finds the production characteristics using the minimal norm least-squares algorithm. The ELM model ultimately finds the best variables and has the fewest learning mistakes. Thereafter, the ELM classifier has been further developed to classify both benign and malignant pulmonary nodules. The accuracy is 94.57%, sensitivity is 93.69%, specificity is 95.15%, receiver operator curve area is 94.94%, and test time is 0.5 ms per nodule.

These studies are able to effectively learn the characteristics of images such as CT or MRI and conduct the diagnosis; however, these studies are limited to patient image-based images and cannot effectively use the medical record information obtained during their consultation and treatment during diagnosis.

In the medical text field, currently mainly using machine learning algorithm and knowledge graph reasoning [11] on knowledge mapping. Among them, Xiu *et al.* [12] focus on knowledge graph patterns and semantic relations, represents knowledge in Chinese electronic medical records with fine-grained and semantics. A knowledge graph schema containing 7 classes and 16 semantic relationships is constructed. Then, the tumor knowledge map of the digestive system is completed by knowledge extraction, named entity links, and drawing knowledge mapping. In “rationality of schema structure,” “scalability,” and “readability of results,” etc., they scored 4.72, 4.67, and 4.69, well above average.

However, due to the complex logic of medical knowledge, it is difficult to build the knowledge atlas, both the disease coverage rate and the diagnosis accuracy are not ideal, and the data resources and energy consumption are great. Unlike knowledge mapping applications, research using machine learning algorithms mainly focuses on case classification, condition prediction, readmission prediction, etc.

In terms of case classification, Ananthakrishnan *et al.* [13] extracted the disease and drug terms, limited properties based on electronic medical record (EMR) coding data and clinical annotation information, and constructed four different models to diagnose Crohn’s disease and ulcerative colitis, where the combined model in different data sets AUC (area under the curve) is 0.95 and 0.86, respectively.

The formation of an atherosclerotic plaque on an artery feeding the cardiovascular system is the most significant source of cardiomyopathy. Inscriptions can become unsustainable, burst, and stimulate the development of a thrombus that plugs the channel; it really can happen in a matter of minutes. In terms of condition prediction, Zhang *et al.* [14] constructed randomized forest models using 14 clinical variables in electronic medical records and used synthetic few over-sampling techniques for preprocessing algorithms to predict 1-month rates of acute myocardial infarction (AMI) and one-month mortality in $<$. Preprediction accuracy between random forest, logical regression, support vector clustering, and K nearest neighbor models is compared, where a randomized forest model was used to predict full-cause mortality of AMI $<$ (AUC) 1 month in patients with chest pain and 0.999 in emergency departments.

In terms of admission prediction, Ashfaq *et al.* [15] investigated a deep learning framework to predict 30-day unplanned readmissions in patients with congestive heart failure, using expert features and context embeddings of clinical concepts to propose cost-sensitive formulas for LSTM neural networks and solve class imbalance problems with AUC as 0.77 and F as 10.51.

These results are mainly based on electronic medical record text mining, processing large amounts of data and providing doctors with valuable information, but most of these studies are limited to patient electronic medical record text records, and many diseases are greatly associated with patient images, and these research methods can only be studied through the information provided by medical record text.

3 Research methods

The method is shown in Figure 1: First, the electronic medical record text and CT image of ischemic stroke patients are processed, where the CT image is a group of a gray map, and the text information is the corpus including the patient’s main complaint, current medical history, previous history, personal history, and admission. Taking the analysis of the patient medical records, by extracting the text features and image features, and automatically learning the transformation matrix with semantic maintained feature mapping through the fusion feature model, realizing the auxiliary diagnosis of the CT image, and comparing with the classical learning model and the existing fusion model in experiments, the model improves the accuracy without increasing the training time and proves that the proposed model is feasible and effective.

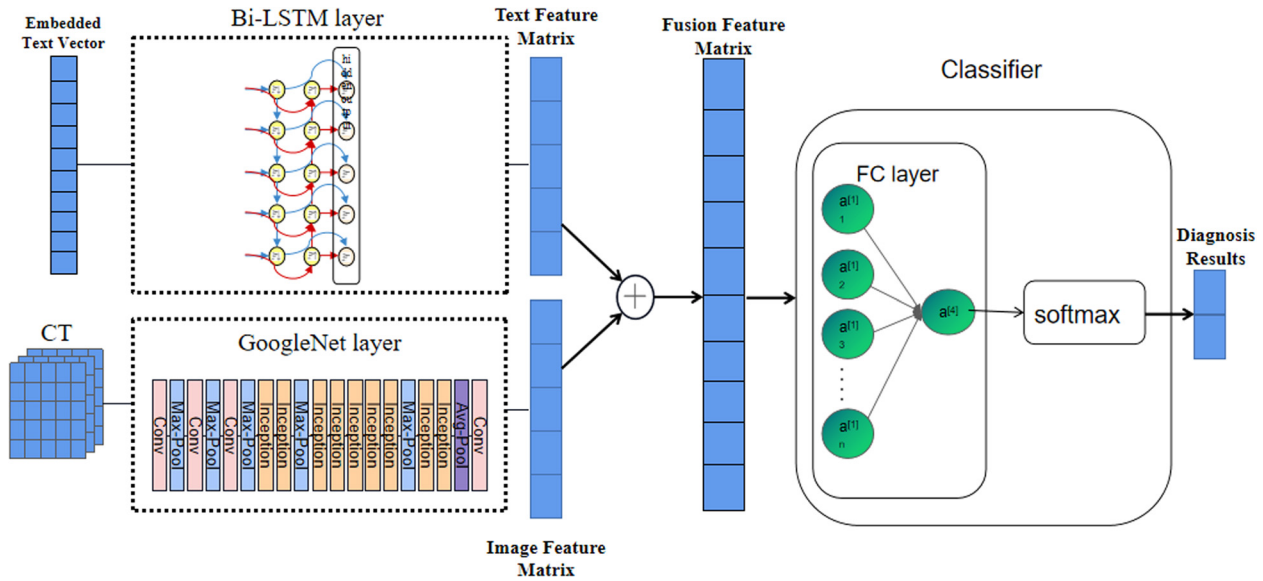


Figure 1: Project implementation schematic diagram.

3.1 Characteristic extraction based on electronic medical record text

There are five inputs into the feature extraction model based on electronic medical records: patient complaint, present history, previous history, personal history, and admission. Due to its high degree of formatting, the features of each part are clear. Most of the information contained is complementary. The existing studies contrast the performance of structured multiple inputs and tail splicing input and found that structured multiple input methods only increased 0.01. So the method of splicing the tail is adopted, forming a character string. Then by the text vectorization method, it is converted to a word vector input matrix. The Cause-and-Effect Matrix connects Various Stages to Processes Inputs (X 's) and Process Documentation to Input information. Acceptance Criteria (or Y 's) are rated in terms of precedence toward the client in a C&E Matrix. The Correlation between the Input variables (X 's) and Outcomes is graded. The vector extracts the features through the Bi-LSTM model and generates the feature matrix [16]. Bi-LSTM is the technique of allowing any neural network to store sequencing including both reverse (prospective to present) and forward (present to years ahead) orientations (past to future). Our input consists of two main ways in a bidirectional LSTM, which distinguishes it from a conventional LSTM. The goal is to see how useful layers of complexity of the classification process would be in fine-tuning the variables in question. The study revealed that Bi-LSTM-based modeling, which incorporates extra data acquisition, gives improved forecasts than those of normal LSTM-based systems.

Model input is the text vector $W = [w_1, w_2, w_3, \dots, w_n]$, w_n is the embedding of the n element. After the W input into the Bi-LSTM model combining the forward LSTM with the backward LSTM, the hidden state output to the LSTM before the t moment is \vec{h}_t . The backward LSTM hidden state output is \overleftarrow{h}_t :

$$\vec{h}_t = \overrightarrow{\text{LSTM}}(\omega_t, \vec{h}_{t-1}), \quad (1)$$

$$\overleftarrow{h}_t = \overleftarrow{\text{LSTM}}(\omega_t, \overleftarrow{h}_{t+1}). \quad (2)$$

The hidden state output of Bi-LSTM at the t moment is made of \vec{h}_t and \overleftarrow{h}_t spliced; that is, $h_t = [\vec{h}_t, \overleftarrow{h}_t]$ gets the hidden state set $H = [h_1, h_2, \dots, h_n]$ after sequence training, assuming u is the number of LSTM hidden nodes, and $uHn \times 2u$ is the dimension.

Softmax is a numerical measure that turns a column of values into a matrix of likelihood, with the probability of every outcome proportionate toward the element's comparable basis. Every number in the

softmax function's return is taken as the likelihood of belonging to every category. The softmax function reduces a column of K real values to a matrix of K real values that add up to 1. The softmax is particularly beneficial in this situation since it turns the results into a normalized probability distribution that may be shown to users or utilized as information for certain other technologies. After the Bi-LSTM model obtains the eigenvector H of the text, it is input into the full connection layer and the Softmax classifier to calculate the probability distribution (Figure 2).

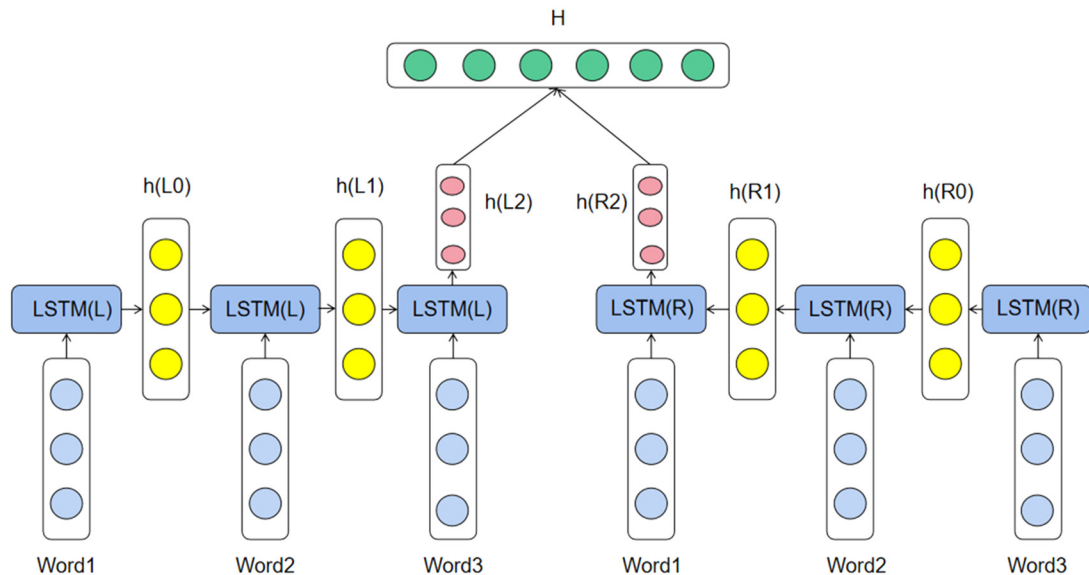


Figure 2: Schematic diagram of Bi-LSTM feature training and fusion.

3.2 Features extraction based on CT images

There are many mature network models in the field of image classification. This article uses the improved AlexNet, respectively GoogleNet, and ResNet, to conduct the research [17]. Where Google Net draws from the NIN (Network in Network) structure, to reduce dims, NIN which is based on CNN uses a 1×1 convolution kernel and uses global average pooling instead of full connection layer reduction parameters. ResNet also uses global average pooling to reduce parameters and prevent overfitting. However, for patient brain slice CT images of $1,608 \times 2,010$ pixels are mainly used in this article, while the three full connection layers used by earlier original AlexNet in this task make the number of parameters unacceptable 109782861440.

In previous studies, many studies have referred the NIN method, improved the accuracy, and reduced the model parameters. Therefore, this article adopts the improved AlexNet model of Yang Zhirui *et al.* [18], replacing the full connection layer with the global mean pooling layer to prevent the overfitting generated by the full connection and improve the generalization ability of the network model. Medical image classification is among the most pressing issues in the field of computer vision, with the goal of categorizing medical pictures into distinct groups to aid clinicians in illness detection as well as study. AlexNet is an eight-layer CNN. Users may import a training dataset variation of the system through Imagenet data set, which has been learned on over a million photos. The technology can categorize photos into 1000 different classification tasks, such as keyboards, mice, and pencils, besides a variety of species. A residual neural network (ResNet) is a kind of ANN that is dependent on prismatic cell architectures in the frontal lobe. Skip connectors, else customizations, are utilized through residual neural network models to hop past certain stages. The effect of improving AlexNet, GoogleNet, and ResNet in auxiliary diagnosis based on CT images and feature extraction were compared in experiments, and finally, the GoogleNet best performed in both aspects was finally used for feature extraction (Figure 3).

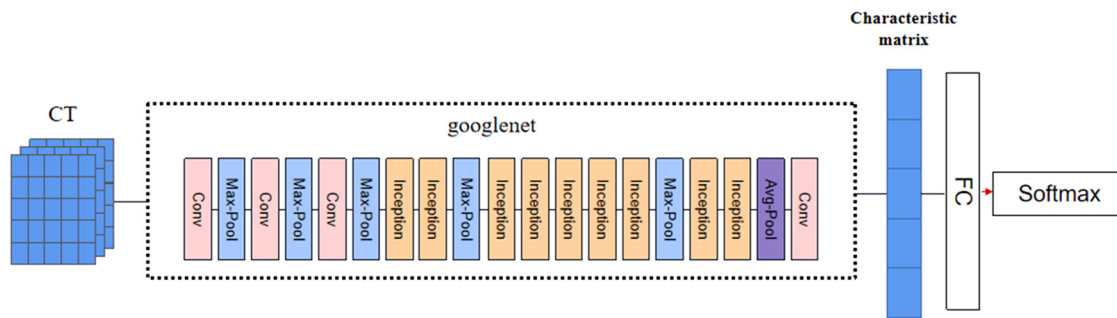


Figure 3: The GoogleNet model structure.

The typical GoogleNet model with 22 layers not only deepens in the depth, but also introduces the Inception structure to expand the width of the network, avoid the gradient and overfitting problem, and solve the gradient and overfitting problems caused by the network expansion. One of the ways GoogLeNet improves performance is by reducing the input sequence while maintaining key visual relationships. The major goal of this layer is to compress the image representation quickly while preserving geospatial data by using high filtration strengths. In addition, the network finally uses the global average pooling layer to replace the full connection layer, which also makes the convergence faster and reduces the overfitting (Figures 4 and 5).

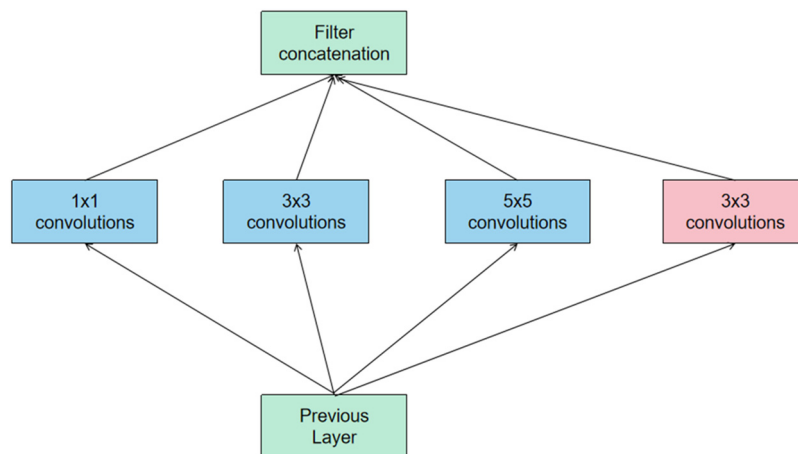


Figure 4: Original inception structure.

A sparse neural network is one in which just a fraction of the available connections are present. Consider a completely linked layer that is lacking certain interconnections. The inception structure hopes to express the probability distribution of the data set through a large and sparse neural network, in which you can cluster the highly relevant units to the previous layer and connect each cluster (Cluster) of the cluster together to build the network layer by layer. This idea is finally realized by a series of scale convolutions to extract the characteristics of different scales of the image, using 1×1 convolutional core in each branch. While improving the convolutional core feeling, it also performs dimension reduction to accelerate the network computation and reduce the computational complexity, thus increasing a layer of feature transformation and nonlinearization with a very small linearization. The feature cascade layer connects the convolution results of 1×1 , 3×3 , and 5×5 , preventing the demand for computational resources caused by the increased number of layers, and expanding both the width and depth of the network.

In this article, the detailed parameters of three-channel CT images into Google N et adopt the GoogleNet parameters proposed by Szegedy et al. [19]. GoogLeNet is a 22-layer deep CNN. Users can import a system that

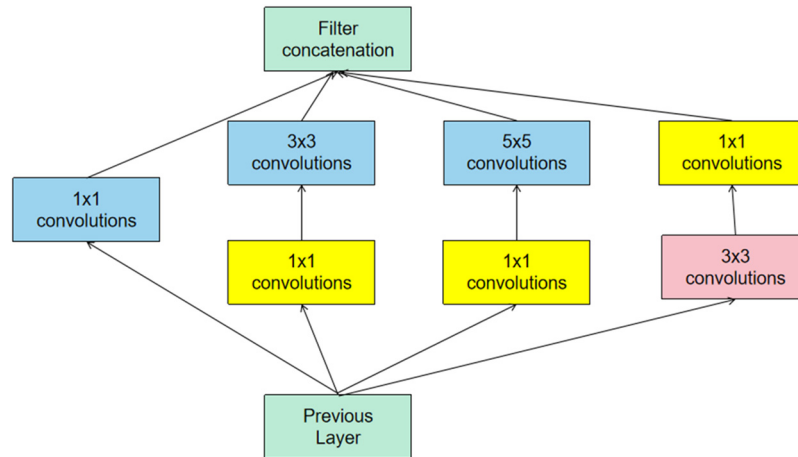


Figure 5: Inception module with dimensionality reduction.

has already been developed on the ImageNet else Places365 data sources. The ImageNet-trained system separates photos across 1,000 different classifiers, including keyboards, mice, pencils, and a variety of creatures. After a series of feature extraction processes including convolution and pooling shown in the figure, the resulting features with a size of $1 \times 1 \times 1,024$ to a one-dimensional vector as the feature matrix extracted by GoogleNet, input into the full connection layer and Softmax to obtain the diagnostic classification results.

3.3 Auxiliary diagnosis of integrating electronic medical record text and CT image

At present, there has been extensive research in the classification field of graphic and text information fusion and fully mining analysis by using image and text. At present, the information fusion method is mainly divided into feature-level fusion, decision-level fusion, and mixed fusion methods.

Decision-level fusion is the decision-making and training of images and text, respectively, and then integrating the obtained decision results through appropriate rules. Chen [20] proposes a cross-modal emotion classification method of TAI-CNN graph and text based on the attention mechanism. TAI-CNN includes TCNN which is text sentiment classification model and ICNN which is image sentiment classification model, and the maximum graph decision level fusion method is adopted to construct the cross-modal emotion classification model, and its accuracy is 0.8446, recall rate is 0.8421, and *F1*-value is 0.8463.

Hybrid fusion mainly integrates the consistency between pictures and texts and extracts features by machine learning method. Borth et al. [21] constructs the middle property SentiBank, based on low-level features like picture color histograms, Gist descriptors, and local binary mode (local binary pattern) and then represents the picture as adjective noun pairs with emotion values, partly addressing the semantic gap between text and the picture.

Feature-level fusion is a fusion of extracting text and image features and predicting after the fusion of features. You et al. [22] should jointly analyze images and text in a structured way, establish a semantic tree structure based on syntactic analysis to map words and picture areas in text, pre-train on the weak labeled Twitter data set, then fine-tune the manually labeled image data set, and finally integrate image features with text features to realize the classification of tweets based on graph and text fusion.

The above graphic fusion research methods have achieved excellent results, proving the complementary role of image and text information in the classification research, further improving the classification effect, and also explaining the application prospects of this method in the medical field based on image or text analysis (Figures 6 and 7).

The feature lengths extracted herein are 1,024 and 512, respectively, and want to learn the correlation characteristics of medical text and CT images. The data set are smaller, unable to effectively learn the

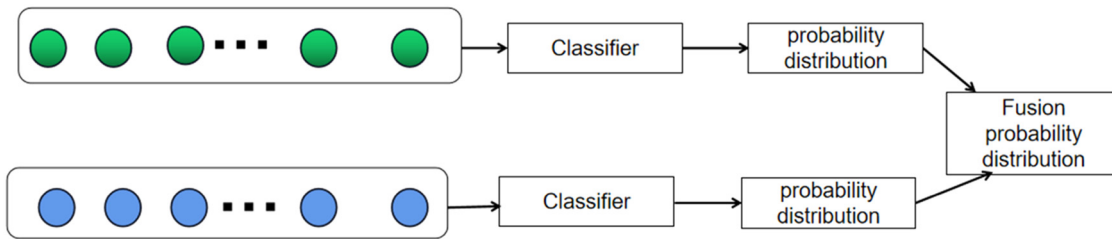


Figure 6: Decision-level fusion.

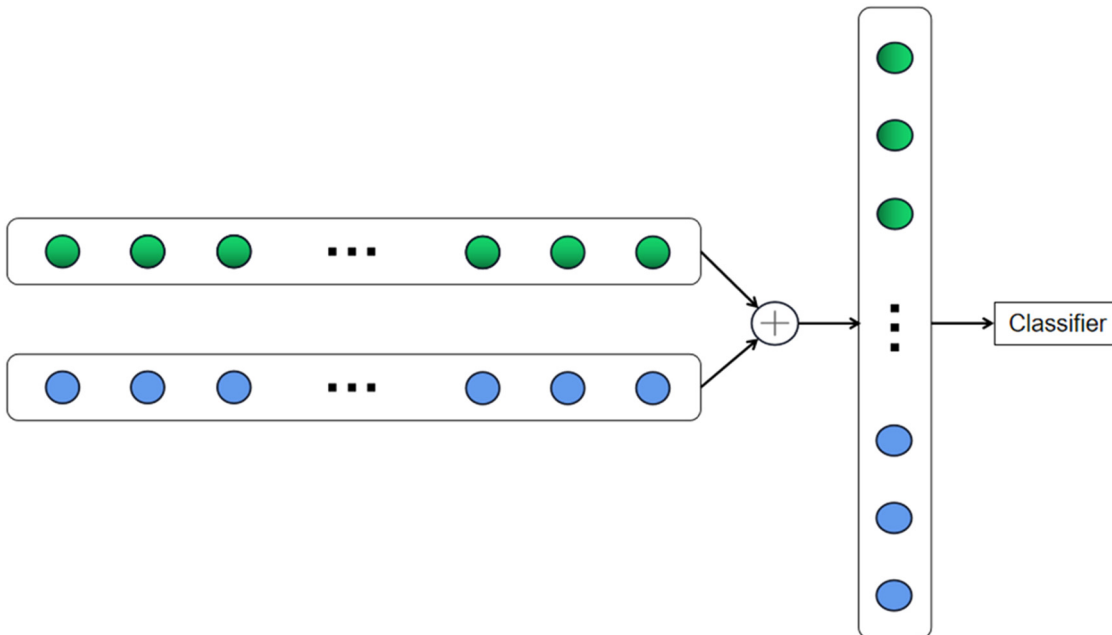


Figure 7: Feature-level fusion.

decision method when CT images are extremely inconsistent with the medical record text performance and perform mixed-fusion studies. I hope that it has a strong generalization ability, robustness, and better performance in other medical tasks. After detailed research and comparison, this article uses the method of one-dimensional characteristic vector splicing and a fully connected layer as FC (Fully Connected) layer and Softmax classifier. As shown in Figure 8, the extracted text feature matrix and the CT feature matrix are combined, respectively. The final probability distribution is then obtained via the FC layer with the Softmax. This preserves the acquired features as much as possible and is judged through the full connection layer (Figure 8).

4 Experimental analysis

4.1 Data set

Since most of the currently disclosed medical data are isolated using CT images or medical record text, this article establishes a case data set associated with the text, associated with the same medical record number, name, and date in the CT cloud database, and desensitizing the name, address, and disease number is

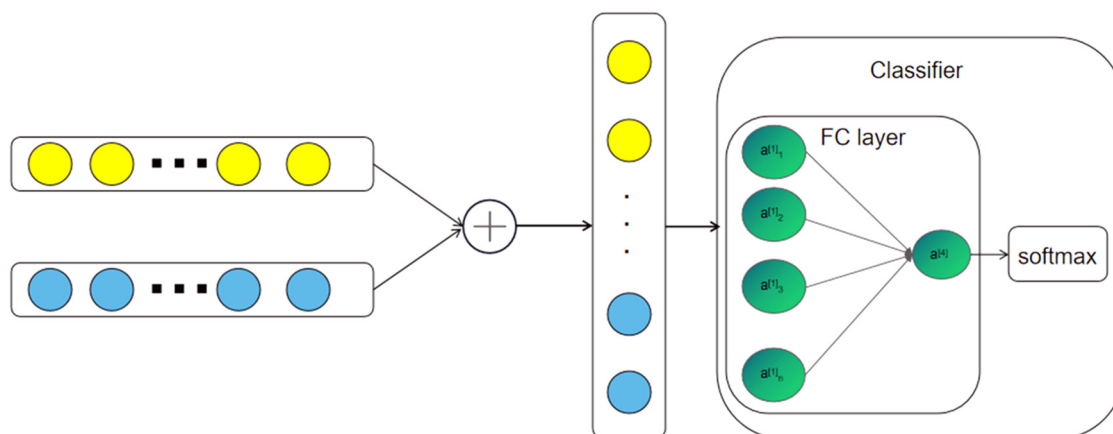


Figure 8: Figure the fusion method.

encrypted or deleted. A case data set of image text associations was finally established as shown in the table. Non-ischemic stroke samples include hemorrhagic stroke, concussion, brain tumor, and brain tissue injury. Whenever a weak blood artery breaks as well as spills into the nervous system, it is known as a hemorrhagic stroke. In accordance with the responsibilities of stroke symptoms, those who have this kind of hemorrhage will probably have a quick arrival migraine or skull discomfort, which is a cautious indication that may not appear throughout an ischemic stroke. In addition, due to the small data set, after dividing the training set and the test set, the data set is expanded four times in the way of 90° rotating images, and the total number of pictures is 1,120.

As shown in the data set, the type CT image is a set of 20 grayscale maps of 402×402 pixels, and the text information is a structured corpus with an average total length of 462 characters, including the patient main complaint, present medical history, previous history, personal history, and admission situation. This model can still be applied to other forms of application scenarios.

The experimental environment is Intel Xeon processor GPU RTX3090 win10 64-bit system, and the deep learning framework is Pytorch 1.6 (Figures 9 and 10; Table 1).

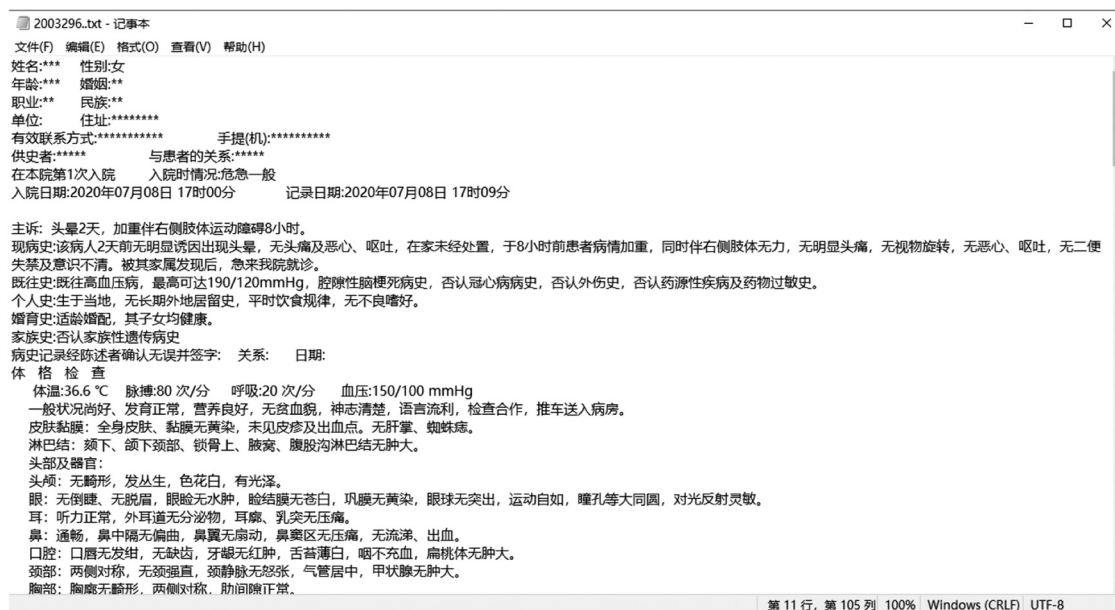


Figure 9: Medical record text sample.

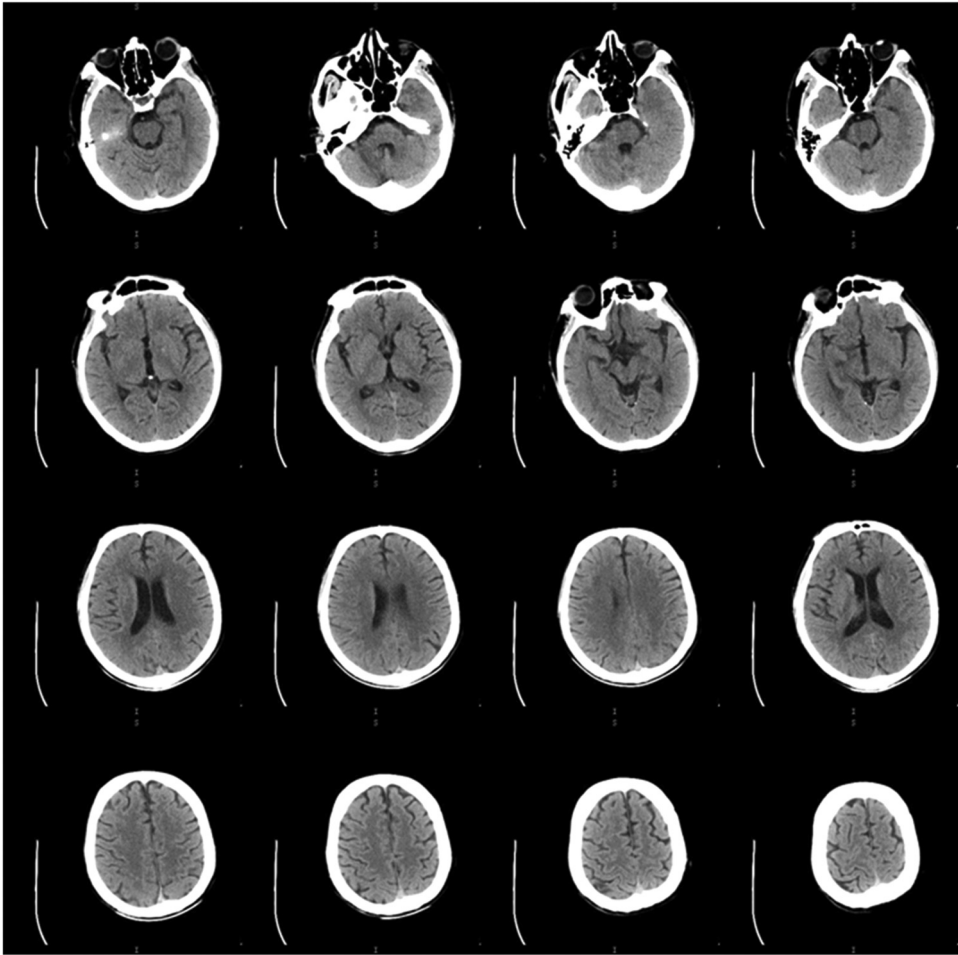


Figure 10: A CT image sample.

Table 1: Number of raw data set parameters

Type of disease	Number of illustrations
Ischemic stroke	151
Non-ischemic stroke	129
Total	280

4.2 Evaluation methods and standards

Due to the particularity of disease detection, in order to fully evaluate the method, this article uses accuracy, and recall rate to evaluate the effect of the *F1* value.

4.3 Experimental setting and result analysis

4.3.1 Single feature-assisted diagnostic experiment

This article classifies the CT images and the Bi-LSTM model, and the results are shown in the table:

The experiment showed that GoogleNet performed better in accuracy and other rates. In terms of auxiliary diagnosis of electronic medical record text, although Bi-LSTM's comprehensive evaluation is weaker than GoogleNet, it has a higher overall examination rate (recall), which is one of the bases for better performance after fusion.

4.3.2 Image text feature fusion-assisted diagnostic experiment

Computer-aided detection else computer-aided diagnosis is a computer-based technology that assists clinicians in making quick judgments in the context of healthcare. Medical devices are concerned with photographic images that health professionals, as well as clinicians, must review or interpret abnormalities in a brief span of time. Xie Hao *et al.* [23] select the fusion algorithm of multi-layer semantic information emotion classification as the comparison method, using the same data set. Sensor fusion techniques integrate sensory input to assist reduce machine perception variability at the time of suitably manufactured. They are tasked with the amalgamation of information from various sensors in which every with its own set of benefits as well as limitations to detect the greatest accurate assignments of belongings. The experimental results are shown in the table:

According to Tables 2 and 3 and Figure 11, during the feature fusion, the features extracted by GoogleNet and Bi-LSTM were greatly improved compared with the separate diagnosis. Although the checking rate of ResNet and Bi-LSTM is better than GoogleNet, the relative sensitivity improvement causes a decrease in accuracy, and the fusion method of GoogleNet and Bi-LSTM is adopted after the comprehensive evaluation.

Table 2: Comparison of the test results of the single-modal model

	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Alexnet	89.69	91.28	90.67	90.97
GoogleNet	94.65	96.58	94.00	95.27
ResNet34	91.60	93.84	91.33	92.57
Bi-LSTM	94.01	93.02	96.15	94.56

Table 3: Comparison of the combined results of different models

	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Xie-Baseline	95.87	94.26	97.46	95.83
GB-Fusion	96.61	100.00	93.22	96.49
RB-Fusion	95.76	93.55	98.31	95.87
AB-Fusion	94.07	93.33	94.92	94.12

Although the comprehensive evaluation of the integration of AlexNet and Bi-LSTM is worse than that of Bi-LSTM alone, the accuracy and registration rate are somewhat improved compared to the Bi-LSTM model, and the integration of GoogleNet and ResNet and Bi-LSTM is significantly improved, which can prove that the integration of CT image and electronic medical record text is feasible.

In addition, the early fusion model of Xie Hao is applied to this field. Although the recall rate of this model is reduced, the accuracy and registration rate are significantly better, and the comprehensive evaluation of this model is improved to a certain extent, which is more suitable for the medical field.

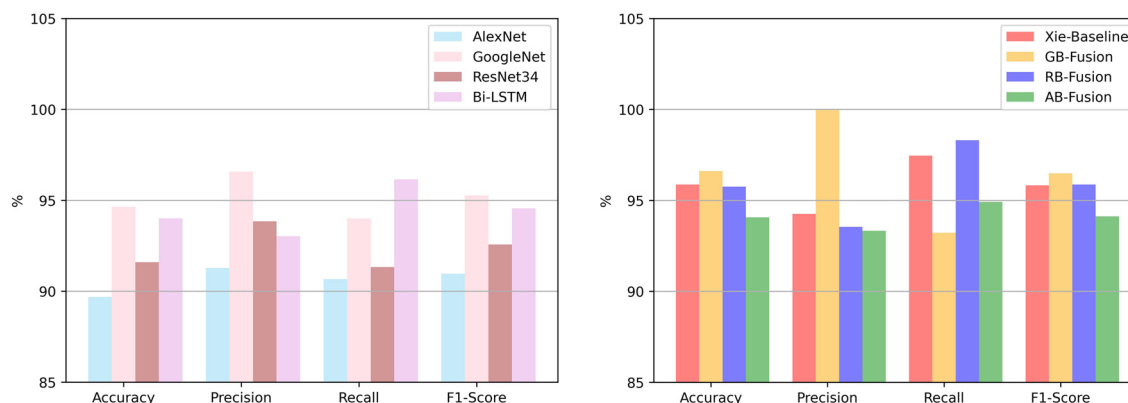


Figure 11: Comparison between unimodal and multimodal model results.

5 Conclusion

In this article, the data set between separate electronic medical record text and CT images proposes a method for medical data. The method is simple, easy to implement, and highly accurate, which uses GoogleNet and Bi-LSTM for features to extract CT images and electronic medical record text and then uses GoogleNet and Bi-LSTM for auxiliary diagnosis of ischemic stroke. The test results show that the method has a good diagnostic effect and provides new ideas for medical-assisted diagnosis. This article also has some defects. For example, due to the small data set, without the mixed fusion method, we will increase the amount of data and focus on collecting cases where all kinds of medical records information conflict with CT images and use the mixed fusion method to improve model accuracy and confidence.

Funding information: This article is the research result of The National College Students' Innovation and Entrepreneurship Training Program Project in 2020, the "Auxiliary Diagnostic Research Based on Electronic Medical Record Image and Text" (20202010225081).

Conflict of interest: Authors state no conflict of interest.

References

- [1] Liz H, Sánchez-Montañés M, Tagarro A, Domínguez-Rodríguez S, Dagan R, Camacho D. Ensembles of convolutional neural network models for pediatric pneumonia diagnosis. *Future Gener Comput Syst.* 2021;122:220–33.
- [2] Kumar PM, Lokesh S, Varatharajan R, Chandra Babu G, Parthasarathy P. Cloud and IoT based disease prediction and diagnosis system for healthcare using fuzzy neural classifier. *Future Gener Comput Syst.* 2018;86:527–34.
- [3] Jialin P, Liver JP. CT image segmentation based on inter-sequence prior constraints and multi-perspective information fusion. *Electron Inform.* 2018;40(4):971–8.
- [4] Weibing. L. Research on multi-label classification algorithm for obstetric assisted auxiliary diagnosis. Zhengzhou University, Zhengzhou; 2019.
- [5] Piedrahita-Gonzalez J, Cubillos-Calvachi J, Gutiérrez-Ardila C, Montenegro-Marin C, Gaona-García P. IOT system for self-diagnosis of heart diseases using mathematical evaluation of cardiac dynamics based on probability theory. *Information systems and technologies to support learning smart innovation. Syst Technol.* 2018;111:433–41.
- [6] Abe H, Ohsaki M, Yokoi H, Yamaguchi T. Implementing an integrated time-series data mining environment based on temporal pattern extraction methods: a case study of an interferon therapy risk mining for chronic hepatitis. *Annual Conference of the Japanese Society for Artificial Intelligence.* Berlin, Heidelberg: Springer; 2005. p. 425–35.
- [7] Talo M, Baloglu UB, Yıldırım Ö, Rajendra Acharya U. Application of deep transfer learning for automated brain abnormality classification using MR images. *Cognit Syst Res.* 2019;54(MAY):176–88.

- [8] Men Ga ShHA, Mahmoud HAH. Brain cancer tumor classification from motion-corrected MRI images using convolutional neural network. *Comput Mater Contin.* 2021;68(2):1551–63.
- [9] Zhou L, Li Z, Zhou J, Li H, Chen Y, Huang Y, et al. A rapid, accurate and machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis. *IEEE Trans Med Imaging.* 2020;99:1.
- [10] Wang Y, Ke Z, He Z, Chen X, Zhang Y, Xie P, et al. Deep transfer convolutional neural network and extreme learning machine for lung nodule diagnosis on CT images. *Knowl Syst.* 2020;46:1829–38.
- [11] XiaoZheng L. Research on EMR Diagnosis Model Based on Deep Learning Integrated with Lexical Semantic. Huaqiao University. 2020;01.
- [12] Xiu X, Qian Q, Wu S. Construction of a digestive system tumor knowledge graph based on chinese electronic medical records: development and usability study. *JMIR Med Informatics;* 2020;8(10):e18287.
- [13] Ananthakrishnan AN, Cai T, Savova G, Cheng SC, Chen P, Perez RG, et al. Improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. *Inflamm Bowel Diseases.* 2013;19(7):1411–20.
- [14] Zhang PI, Hsu CC, Kao Y, Chen CJ, Kuo YW, Hsu SL, et al. Real-time AI prediction for major adverse cardiac events in emergency department patients with chest pain. *SCJ Trauma Resusc Emerg Med.* 2020;28(1):93.
- [15] Ashfaq A, Sant'anna A, Lingman M, Nowaczyk S. Readmission prediction using deep learning on electronic health records. *J Biomed Inform.* 2019;97:103256.
- [16] Baochen D. Research and implementation of text mining based on medical record data. Beijing University of Posts and Telecommunications. 2019;08.
- [17] Nguyen GN, Viet NHL, Elhoseny M, Shankar K, Gupta BB, El-Latif AAA. Secure blockchain enabled cyber–physical systems in healthcare using deep belief network with ResNet model. *J Parallel Distrib Comput.* 2021;153:150–60.
- [18] Zhirui Y, Hong Z, Zhongyuan G, Xiaohang X. Defect detection of jujube based on NIN convolutional neural network. *Food Machinery.* 2020;36(2):140–5, 181.
- [19] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D. Going deeper with convolutions. *IEEE Comput Soc.* 2015. p. 1–9.
- [20] Xiaomin C, Xiaohu X, Dikai F, Junsheng X. TAI-CNN cross-modal emotion classification method based on attention mechanism. *Comput Appl Softw.* 2021;5:31.
- [21] Borth D, Ji R, Chen T, Breuel T, Chang S. Large-scale visual sentiment ontology and detectors using adjective noun pairs. *Proceedings of the 21st ACM International Conference on Multimedia;* 2013. p. 223–32.
- [22] You Q, Luo J, Jin H, Yang J. Robust image sentiment analysis using progressively trained and domain transferred deep networks. *Twenty-Ninth AAAI Conference on Artificial Intelligence;* 2015. p. 381–8.
- [23] Hao X, Jin M, Gang L. Sentiment classification of image-text information with multi-layer semantic fusion. *Data Anal Knowl Discov.* 2021;5(6):103–14.