Efficient Processing and Intelligent Diagnosis Algorithm for Internet of Things Medical Data Based on Deep Learning

Wang Liyun
Information Department of Linyi Central Hospital, Linyi City, Shandong Province, 276000, China

***Abstract*—Electronic Medical Record (EMR) is a commonly used tool in medical diagnosis, which has static recording, difficulty in combining and analyzing different forms of data, and insufficient diagnostic efficiency and accuracy. This article proposes a CNN (Convolutional Neural Network)-LSTM (Long Short-Term Memory) algorithm for efficient processing and intelligent diagnosis of Internet of Things (IoT) medical data. The Word2Vec model is applied to clinical text data and its ability is utilized to capture semantic relationships between words. Medical image data is feature extracted using CNN, while physiological signal data is dynamically processed using LSTM to identify trends and anomalies in the data. An attention mechanism is applied to dynamically adjust the model's attention weights for different types of data. By analyzing the samples of health, cardiovascular disease, diabetes, chronic obstructive pulmonary disease, hypertension, and chronic kidney disease, the CNN-LSTM in this article can accurately classify a variety of diseases, and the accuracy rate of healthy individuals has reached 97.8%. By combining CNN-LSTM with multimodal data, the accuracy and efficiency of medical diagnosis have been effectively improved.**

Keywords*—Intelligent diagnosis; Internet of Things medical; electronic medical records; long short-term memory; convolutional neural network

I. Introduction

With the rapid development of information technology and the healthcare industry, electronic medical record (EMR) has become one of the core tools for medical data management. EMR [1-2] stores various data such as patients' medical history, examination reports, imaging data, and medication use, providing important basis for clinical diagnosis and treatment. The use of traditional electronic medical records faces many challenges. Most EMR records patients' conditions in a static manner, lacking real-time monitoring of their health status and unable to reflect dynamic changes in their condition in a timely manner. EMR data comes in various forms, including text, images, structured data, and unstructured data, which makes data integration and analysis complex. The lack of unified data standards between different hospitals or medical institutions increases the difficulty of information sharing and data integration, which in turn affects the efficiency and accuracy of diagnosis. With the rapid increase in the volume of medical data, traditional manual analysis and processing methods have become difficult to cope with. How to efficiently and accurately analyze and process these massive amounts of data has become an important issue in medical data research. The development of Internet of Things (IoT) technology has brought new opportunities to the medical field, especially in the areas of medical data collection, transmission, and processing. Through intelligent sensors, wearable devices, and implantable devices, the Internet of Things can collect real-time physiological data of patients, including heart rate, blood pressure, blood sugar, etc., providing dynamic data sources for electronic medical records and filling the gap of traditional EMR static recording. The massive data generated by the Internet of Things has also brought new challenges, and how to efficiently process and analyze these multimodal and heterogeneous medical data has become a focus of current research. Deep learning techniques [3-4] have emerged, among which convolutional neural networks and long short-term memory networks have shown outstanding performance in processing complex data and automatically extracting features. Combining IoT technology, deep learning algorithms can automatically analyze medical data from different data sources, extract high-value information, and make accurate diagnoses and predictions, thereby improving the processing efficiency of medical data and the intelligence level of diagnosis. This fusion technology provides doctors with diagnosis and treatment advice, promoting the development of personalized medicine.

This study proposes a multimodal diagnostic model based on CNN (Convolutional Neural Network)-LSTM (Long Short-Term Memory), providing a new solution for early diagnosis and monitoring of chronic diseases. This model integrates medical imaging data, temporal physiological data, and clinical text data, significantly improving the accuracy and efficiency of disease classification. The model performs excellently in the classification tasks of healthy individuals and various chronic diseases, surpassing traditional diagnostic methods. This contribution not only provides more accurate decision support for clinical practice, but also provides empirical basis for research in related fields, promoting the development of intelligent healthcare.

The innovation of this study is reflected in multiple aspects. The combination of deep learning algorithms CNN and LSTM fully utilizes the advantages of convolutional neural networks in image feature extraction and the powerful capabilities of long short-term memory networks in temporal data processing, thereby achieving comprehensive analysis of multimodal data. By applying attention mechanism, it can automatically identify and focus on key features in the input data, further improving the diagnostic accuracy. The use of Word2Vec technology to extract key disease descriptions, symptoms, and diagnostic information from clinical text data provides richer contextual information for the model and promotes effective fusion of multimodal data. These innovations have laid the theoretical and practical foundation for future intelligent diagnostic systems.

This article has a clear organizational structure and clear hierarchy. The introduction section clarifies the research background and significance, points out the limitations and urgent needs of traditional diagnostic methods, and then introduces the main objectives and research methods of this study. The methods section provides a detailed explanation of data collection, preprocessing, model construction, and training processes, offering readers comprehensive technical details. In the results section, the performance of the model is visually demonstrated through charts and data analysis, including evaluation indicators such as accuracy and Kappa coefficient, ensuring the transparency and reliability of the research results. The conclusion section summarizes the research findings, analyzes their practical significance and limitations, and provides prospects for future research directions. This clear structure not only facilitates readers' understanding, but also enhances the academic value of the article.

II. Related Works

Medical diagnosis is an important component of the medical field. With the advancement of technology, the methods of medical diagnosis are constantly evolving, gradually shifting from traditional doctor experience judgment to scientific diagnostic methods based on data analysis. Early medical diagnosis [5] mostly relies on the clinical experience and limited laboratory data of doctors, and the accuracy of diagnosis largely depends on the professional knowledge and experience accumulation of doctors. With the development of imaging technology [6] and molecular diagnostic technology, medical diagnosis has entered a data-driven stage. The widespread application of imaging technology has made medical diagnosis more dependent on digital imaging data, providing technical support for early detection and accurate diagnosis of diseases. Molecular diagnostic technology [7], through in-depth analysis of the genome, proteome, and metabolomics, can identify the molecular characteristics of diseases at the microscopic level, especially playing an important role in cancer diagnosis and personalized treatment. The amount of medical data is huge and complex, and how to effectively extract useful information from it remains a huge challenge. Many studies have begun to explore how to improve the accuracy and efficiency of medical diagnosis through intelligent algorithms and big data analysis technologies. Tian Miao's research [8] showed that by combining advanced artificial intelligence and machine learning algorithms, medical diagnosis can be automated and intelligent, especially in image analysis, pathological analysis, and disease prediction, where significant progress has been made. This data-driven diagnostic approach not only improves the accuracy of diagnosis, but also reduces the workload of doctors and promotes the intelligent transformation of the medical field.

Electronic medical records, as the main carrier of medical data, have been widely used in the global healthcare system. It records the entire process data of patients from initial visit to subsequent treatment, including medical history, examination reports, diagnostic conclusions, and medication use, becoming an important basis for clinical diagnosis. With the development of big data and cloud computing technology, researchers have begun to explore how to utilize these rich electronic medical record data to provide support for medical diagnosis. Early EMR diagnosis [9] mainly relies on the structuring and normalization of data for statistical analysis and decision support by doctors and researchers. Due to the heterogeneity and diversity of EMR data, unstructured data such as text, images, and audio are widely present, and traditional diagnostic methods have low efficiency in processing these data. The application of IoT technology [10-11] has brought new opportunities for the diagnosis of electronic medical records. Through wearable devices and implanted sensors, the Internet of Things can monitor patients' physiological data in real-time, including heart rate, blood pressure, blood sugar, etc., and seamlessly integrate these data into electronic medical record systems to achieve dynamic tracking and real-time diagnosis of patients' conditions. The EMR system combined with the Internet of Things can achieve remote monitoring and diagnosis through remote medical devices, providing an effective means for the continuous treatment of chronic disease patients. Diabetes patients can monitor the blood glucose level in real-time through the Internet of Things device, and upload the data to the EMR system. Doctors can adjust the treatment plan through the intelligent analysis results provided by the system. This electronic medical record diagnosis system, which combines IoT technology, is gradually improving the traditional medical diagnosis mode, enhancing diagnostic efficiency and accuracy. The Deep Multi-Scale Fusion Neural Network (DMFNN), as presented by Dinesh Kumar Reddy Basani et al. (2024), was designed for fault detection in IoT systems using data integration. Leveraging their fusion strategy, our framework processes diverse medical IoT datasets by extracting layered information and handling noise which improve diagnostic precision and operational reliability [12]. Naresh Kumar Reddy Panga (2022) utilized Discrete Wavelet Transform (DWT) for analyzing ECG signals in IoT-based health monitoring platforms. Drawing from their methodolody, their DWT approach is employed in our research to isolate features and diminish interference. This enables improved signal quality and reducing computational load, supports to achieve greater accuracy and timely analysis [13]. A structural model combining IoT, fog, and cloud computing was developed by Thirusubramanian Ganesan, (2021) enables continuous ECG surveillance using machine learning. This layered architecture is incorporated in our proposed scheme to manage medical IoT data streams and processing stages, which enhance scalability, and diagnostic accuracy [14]. Rajababu Budda (2021) developed a framework blending Artificial Intelligence and Big Data analytics tailored for IoT healthcare, concentrating on optimized performance and patient-focused services. Building on this foundation, our research narrows the focus of their conceptual framework with an emphasis on deep learning in our work to enable proficient medical data handling and insightful diagnosis, facilitating the creation of scalable, accurate, and real-time monitoring solutions while advancing diagnostic reliability and sustainable care delivery [15]. Sri Harsha Grandhi (2021) proposed an adaptive wavelet transform method combined with wearable IoT devices for effective pediatric health monitoring. Our system embeds this adaptive wavelet transform method to refine raw medical data before deep learning analysis. This ensures cleaner signals and accurate feature extraction through wavelet denoising, promotes stability and efficient real-time observation [16]. In recent years, significant progress has been made in the application of deep learning technology in medical diagnosis, especially in the field of intelligent diagnostic algorithms. Deep learning is a branch of machine learning that automatically extracts features from massive amounts of data through multi-layer neural networks, and then performs classification, prediction, and decision-making. In the medical field, deep learning [17] is widely used in disease diagnosis, image analysis, genomics, and other fields, significantly improving the accuracy and automation of diagnosis. Early research mainly focuses on the use of convolutional neural networks in image diagnosis. Through automated analysis of medical images such as X-rays, deep learning algorithms can effectively identify pathological features such as tumors and lesions. The performance of the lung cancer image recognition system based on CNN [18] in tumor detection has approached or even exceeded the diagnostic level of human radiologists. Over time, the application of deep learning in processing unstructured data has also been widely studied. Recurrent neural networks and long short-term memory networks [19] are widely used to analyze electronic medical record text data, automatically extract key information from medical records, and dynamically predict the patient's condition. Attention mechanisms [20] and new deep learning architectures such as autoencoders have also been applied to the fusion and processing of medical data, enhancing the ability to analyze complex and multimodal data. Combining IoT technology, deep learning algorithms can process real-time medical data from different data sources, achieving personalized and accurate intelligent diagnosis. This intelligent diagnostic system not only improves the efficiency of medical resource utilization, but also provides strong support for personalized and remote healthcare.

In recent years, significant progress has been made in the application of deep learning technology in medical diagnosis, especially in the field of intelligent diagnostic algorithms. Deep learning is a branch of machine learning that automatically extracts features from massive amounts of data through multi-layer neural networks, and then performs classification, prediction, and decision-making. In the medical field, deep learning [17] is widely used in disease diagnosis, image analysis, genomics, and other fields, significantly improving the accuracy and automation of diagnosis. Early research mainly focuses on the use of convolutional neural networks in image diagnosis. Through automated analysis of medical images such as X-rays, deep learning algorithms can effectively identify pathological features such as tumors and lesions. The performance of the lung cancer image recognition system based on CNN [18] in tumor detection has approached or even exceeded the diagnostic level of human radiologists. Over time, the application of deep learning in processing unstructured data has also been widely studied. Recurrent neural networks and long short-term memory networks [19] are widely used to analyze electronic medical record text data, automatically extract key information from medical records, and dynamically predict the patient's condition. Attention mechanisms [20] and new deep learning architectures such as autoencoders have also been applied to the fusion and processing of medical data, enhancing the ability to analyze complex and multimodal data. Combining IoT technology, deep learning algorithms can process real-time medical data from different data sources, achieving personalized and accurate intelligent diagnosis. This intelligent diagnostic system not only improves the efficiency of medical resource utilization, but also provides strong support for personalized and remote healthcare.

III. Methods

A. IoT Medical Data Collection and Preprocessing

1) Device deployment: In the integration of IoT technology and the medical field, selecting sensors and wearable devices that are suitable for the target disease and patient health status is the key to achieving personalized medicine. Real-time monitoring of various physiological parameters through sensor devices helps medical staff obtain comprehensive and dynamic health data. Heart rate sensors are used to monitor heart health, especially for patients with heart disease. Through implantable devices such as pacemakers, sensors can precisely measure the electrical activity of the heart, avoiding delays and errors in traditional methods. This type of device, when combined with external devices, can transmit heart rate data in real-time, providing strong support for remote diagnosis and emergency treatment. Changes in heart rate can reveal early heart problems, and timely intervention can greatly reduce the risk of sudden heart disease.

Blood pressure sensors are also important monitoring tools, especially for patients with hypertension, which can help doctors track blood pressure fluctuations in real-time. Traditional blood pressure monitoring methods require patients to manually measure blood pressure at regular intervals, and most of them are discrete data. Through wearable blood pressure monitoring devices, dynamic changes in blood pressure data can be continuously obtained. By installing on the arms, wrists, and other parts, based on IoT transmission, real-time data can be sent to the cloud for doctors to analyze.

Body temperature sensors are used for patients with fever, infections, and other diseases that require temperature monitoring. Body temperature is automatically monitored and continuous data streams are generated through non-contact or contact sensors placed on the forehead, ears, or wrist. Combined with the Internet of Things transmission network, data is uploaded in real-time to the hospital system, allowing doctors to remotely analyze the trend of temperature changes and predict the deterioration of the condition in advance. For patients with a long-term medical history or weak immune system, temperature fluctuations may be an early signal of infection or other potential problems. With the help of IoT devices, rapid detection and measures can be taken to reduce the probability of disease deterioration.

Blood glucose monitoring is a vital health management link for patients with diabetes. The blood glucose level is continuously detected through subcutaneous sensors, and the data is transmitted to intelligent devices in real-time to facilitate patient self-management. Doctors can automatically adjust medication doses or dietary plans based on historical data. This non-invasive and continuous monitoring method can improve patients' quality of life and greatly reduce the risk of acute complications.

The blood oxygen saturation sensor continuously monitors the oxygen content in the blood through optical sensors installed on fingers, earlobes, and other parts. The fluctuation of blood oxygen levels is a key indicator for judging respiratory distress or abnormal lung function. Through real-time uploaded blood oxygen data from IoT devices, doctors can promptly determine whether patients need oxygen therapy or hospitalization. By combining multiple data sources such as heart rate and blood pressure, the IoT platform can conduct comprehensive analysis and generate personalized treatment plans through algorithms, helping patients achieve self-monitoring in a home environment and reducing hospitalization needs. The IoT data collection network is shown in Fig. 1.

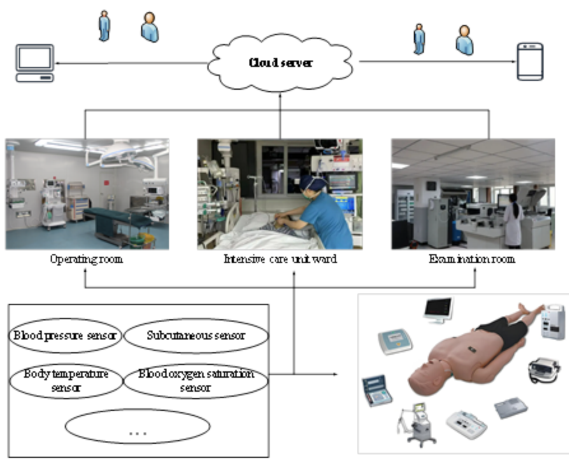

Fig. 1. IoT data collection network.

2) Data collection and transmission: Various physiological and non physiological data are collected through different types of sensors, including physiological signal data, medical imaging data, and clinical text data. Real-time physiological data collected by sensors is transmitted to a medical data management platform through wireless networks, and the data can be uploaded to the cloud in real-time, ensuring that all monitoring data can be continuously stored, updated, and analyzed without interruption. IoT devices can also interface with the hospital's electronic medical record system, ensuring that these real-time data can be seamlessly integrated with the patient's historical medical records.

This article is conducted in a tertiary comprehensive hospital, recruiting a total of 300 participants, all of whom are outpatient or inpatient patients. Among the subjects, 100 are healthy individuals for the control group, and another 200 are patients with different types of diseases, including five types of diseases: cardiovascular disease patients (60), diabetes patients (50), chronic obstructive pulmonary disease patients (40), hypertension patients (30), and chronic kidney disease patients (20). All participants sign informed consent forms before participation, and data is collected in real-time through IoT devices, covering physiological parameters such as heart rate, blood pressure, blood glucose, and blood oxygen saturation, aiming to evaluate the application effect and accuracy of the intelligent diagnostic system in different disease scenarios.

The data collection period is from June 2022 to December 2022, and the collected physiological data is shown in Table I.

In Table I, part of physiological data of a healthy individual is presented, with systolic and diastolic blood pressure data included in the blood pressure. Through IoT sensors, physiological data information of subjects can be accurately and continuously collected. The collected image data [21-22] are shown in Fig. 2.

TABLE I. DISPLAY OF PART OF PHYSIOLOGICAL DATA

Date	Blood pressure (mmHg)	Body temperature (°C)	Blood glucose (mg/dL)	Blood oxygen saturation (%)
2022-6-1	120/80	36.6	90	98
2022-6-2	118/79	36.7	88	99
2022-6-3	121/80	36.5	91	98
2022-6-4	119/78	36.6	89	99
2022-6-5	122/81	36.6	92	98
2022-6-6	120/79	36.7	90	99
2022-6-7	119/77	36.5	89	98
2022-6-8	121/80	36.6	91	99
2022-6-9	120/78	36.6	90	98
2022-6-10	119/79	36.5	88	99

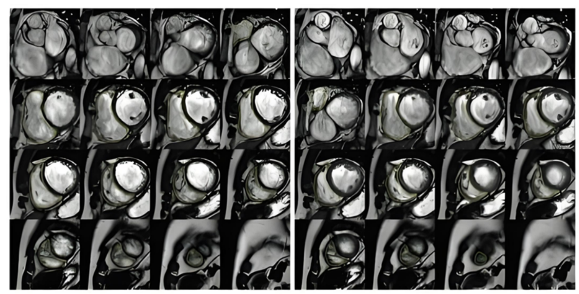

Fig. 2. Collected image data.

3) Data preprocessing: The data collected by IoT devices is noisy due to environmental interference, sensor accuracy, or other factors. To remove noise, Gaussian filters are used to smooth the data and reduce the interference of random noise. The kernel function formula for Gaussian filtering is:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

σ is the standard deviation, which controls the smoothness of the filter. In order to avoid the impact of missing data on the analysis results, data filling must be carried out. The formula for mean interpolation method is:

$$x(t) = \frac{x(\mathrm{t}-1) + x(\mathrm{t}+1)}{2} \tag{2}$$

$x(t)$ is the missing data at time point t , and $x(t-1)$ and $x(t+1)$ represent adjacent observations before and after.

Physiological signals such as heart rate, blood oxygen saturation, and blood pressure have different numerical ranges. Through standardization, is is ensured that data from each dimension falls within the same range. The standardized formula for Z -score is:

$$x' = \frac{x-\mu}{\sigma} \tag{3}$$

Due to the fact that IoT data may come from multiple devices, there may be inconsistencies in the data collected by each device at the same time or in the same scenario. Therefore, consistency checks and corrections are necessary before data fusion. By using time alignment and device calibration techniques, it is ensured that data from different sources accurately reflect the patient's status at the same time.

B. Multimodal Medical Data Fusion

In IoT healthcare systems, combining multiple data sources to form multimodal datasets provides more comprehensive diagnostic evidence. The collected physiological signal data, medical imaging data, and clinical text data are combined to form a multimodal dataset. After data integration is completed, feature extraction is performed for different types of data. The Word2Vec is used to automatically extract key disease descriptions, symptoms, and diagnostic information from clinical text data. For medical image data, CNN is used for automated feature extraction. LSTM is used to process dynamic changes in physiological data and extract key trend information. Word2Vec [23-24] is a deep learning model that automatically extracts disease descriptions, symptoms, and diagnostic information by capturing semantic relationships between words. Through training, Word2Vec is able to generate vector representations for each word, making words with similar meanings closer together in the vector space. The formula for calculating word vectors is:

$$v(w)=\frac{1}{T}\sum_{t=1}^{T}\log P\left(w_t \mid w_{t-n},\ldots,w_{t+n}\right) \tag{4}$$

CNN can effectively extract high-level features of input data by stacking multiple convolutional and pooling layers. Each convolutional layer performs feature mapping on the input data by applying convolutional kernels, as follows:

$$Z_{i,j}=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}X_{i+m,j+n}\cdot K_{m,n}+b \tag{5}$$

$Z_{i,j}$ is the output feature map, and K represents the convolution kernel. Maximum pooling is used to reduce the dimensionality and computational complexity of feature maps, and the formula is:

$$Z_{i,j}=\max_{m,n}X_{i+m,j+n} \tag{6}$$

By combining multi-layer convolution and pooling, CNN can automatically extract useful features from complex input data, enhancing the model's classification performance. LSTM can effectively capture long-term dependencies and solve the gradient vanishing problem by applying gating mechanisms. When processing physiological data, LSTM can analyze physiological parameters at different time steps in real-time, identify their dynamic trends, and provide support for intelligent diagnosis. This ability makes LSTM a powerful tool for processing complex time series data.

C. Design and Training of Deep Learning Models

1) Model architecture design: The model designed in this article integrates convolutional neural networks and long short-term memory networks [25-26], aiming to improve the intelligent diagnostic capabilities of medical image analysis and temporal physiological data processing. CNN is responsible for processing medical imaging data and effectively extracting lesion features from images through multi-layer convolution and pooling operations. Convolutional layers can capture local features and identify potential lesion areas in images, while pooling layers help reduce feature dimensions and enhance the model's focus on important features. This process enables the model to accurately identify pathological features in complex imaging data, improving diagnostic accuracy.

The model structure designed in this article is shown in Fig. 3.

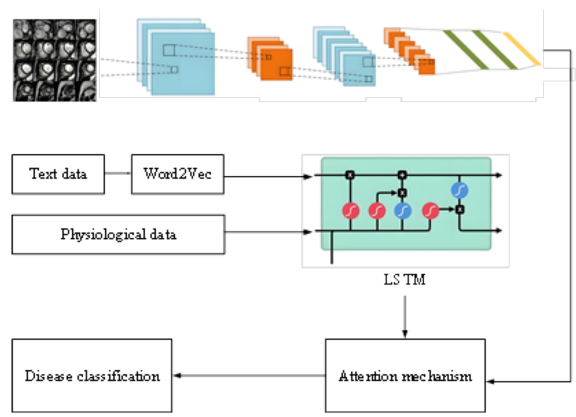

Fig. 3. Model structure.

2) Data annotation and model training: The annotation process involves associating the doctor's diagnostic results with input data, including medical imaging, time-series physiological data, and textual data. Doctors determine the diagnostic label and specific disease category for each sample based on imaging analysis and clinical evaluation results. Professional medical personnel are assisted to ensure the accuracy and reliability of the labels. Using multiple doctors for independent annotation and resolving annotation differences through collective discussion can further improve the consistency and objectivity of annotation results.

After annotation is completed, the dataset is divided into a training set and a test set, with 80% as the training set and 20% as the test set. Cross-validation method is used to further prevent overfitting of the model. By further dividing the training set into multiple subsets, the model undergoes multiple rounds of training and validation on different subsets, effectively reducing its dependence on specific data and ensuring the robustness and reliability of the model.

The model is trained using annotated multimodal data and the model parameters are optimized using the Adam optimizer. The Adam optimizer combines the advantages of momentum and adaptive learning rate, making the training process more efficient and stable. The cross entropy loss function is used to measure the performance of the model in classification tasks. The cross entropy loss function can be expressed as:

$$L(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{c} y_{ij}\log\left(\hat{y}_{ij}\right)(7) \quad (7)$$

D. Diagnostic Accuracy Performance

Using CNN-LSTM for intelligent diagnosis, the confusion matrix results for different disease diagnoses are shown in Fig. 4.

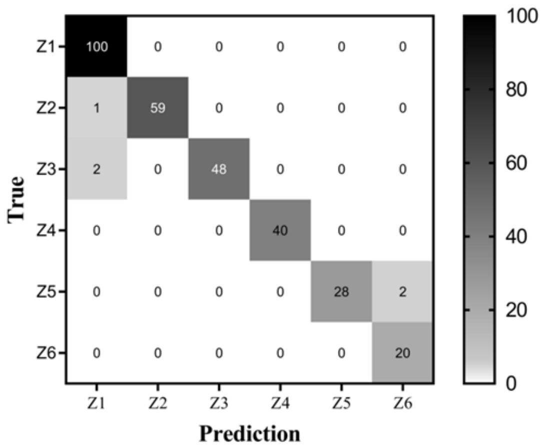

Fig. 4. Confusion matrix.

In Fig. 4, health, cardiovascular disease, diabetes, chronic obstructive pulmonary disease, hypertension, and chronic kidney disease are represented by Z1, Z2, Z3, Z4, Z5, and Z6 respectively. The CNN-LSTM model shows good classification ability in the classification tasks of six types of health conditions and diseases. All 100 healthy individual samples are correctly classified as Z1, indicating that the model has very high robustness in identifying disease-free individuals. For cardiovascular diseases, the model also shows excellent classification performance. Although the model correctly classifies 59 samples, one sample is misclassified as a healthy individual, indicating that the model may have a small margin of error in distinguishing healthy individuals from patients with mild cardiovascular symptoms. There is still high reliability in identifying patients with cardiovascular disease. In the classification of diabetes, the model correctly classifies 48 samples, and 2 samples are wrongly classified as healthy individuals. Some characteristics of diabetes patients may be similar to those of healthy individuals, leading to slight confusion of models. The classification performance of chronic obstructive pulmonary disease and chronic kidney disease is relatively excellent, with the model correctly classifying 40 and 20 samples, respectively, without any misclassification, demonstrating its reliable classification performance in these two types of diseases. Two samples of hypertension are misclassified as chronic kidney disease, which may be due to certain similarities in physiological signals or symptoms between the two types of diseases, resulting in confusion in the model. The CNN-LSTM model has high classification accuracy for different categories of diseases, but there are a few misclassifications between healthy individuals and certain chronic diseases.

The comparison results of accuracy for different disease categories are shown in Fig. 5.

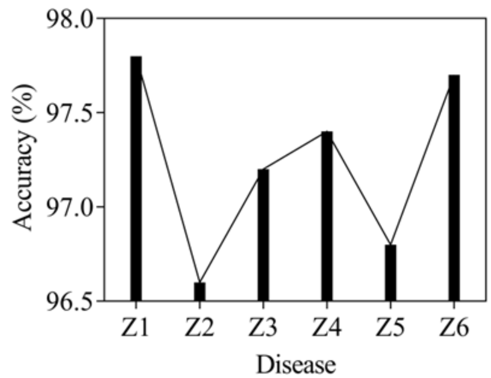

Fig. 5. Accuracy of different disease categories.

In the accuracy results of the model for different disease categories in this article, the CNN-LSTM model shows relatively stable and efficient accuracy in classifying various diseases. The accuracy rate of healthy individuals is 97.8%, indicating that the model can accurately identify disease-free individuals, which reflects the model's good ability to distinguish healthy samples. The accuracy of chronic kidney disease also reaches 97.7%, second only to healthy individuals, indicating that the model has high sensitivity and reliability in distinguishing kidney diseases. The accuracy rates of diabetes, chronic obstructive pulmonary disease and cardiovascular disease are 97.2%, 97.4% and 96.6% respectively, which means that the model can still maintain a high classification accuracy when dealing with these common chronic diseases. Although the accuracy of cardiovascular disease is slightly lower, it still maintains a high level of over 96%, proving that the model also has a certain degree of robustness in identifying cardiovascular disease. The classification accuracy of hypertension diseases is 96.8%, slightly lower than other categories, but the difference is not significant, indicating that the model also has a good recognition effect on blood pressure fluctuation diseases. Overall, the classification accuracy of the model is higher than 96% in all disease categories, indicating its excellent performance in multimodal data processing and feature extraction, with strong generalization ability and application potential.

E. Comparison with Baseline Model

In order to comprehensively analyze the intelligent diagnostic performance of the model in this article, it is compared with other models, and their performance are analyzed through AUC values. The AUC (Area Under the Curve) data is shown in Table II.

By evaluating the performance of different models on multimodal medical datasets using AUC metrics, the CNN-LSTM model significantly outperforms other models, exhibiting the highest AUC values (between 0.95 and 0.98). The advantage of CNN-LSTM lies in its effective integration of the characteristics of convolutional neural networks and long short-term memory networks: CNN excels at extracting spatial

TABLE II. AUC DATA TABLE

Disease	CNN-LSTM	CNN	LSTM	GRU (Gated Recurrent Unit)	RF (Random Forest)	SVM (Support Vector Machine)	Transformer
Z1	0.98	0.95	0.92	0.93	0.85	0.83	0.94
Z2	0.97	0.94	0.91	0.92	0.84	0.82	0.93
Z3	0.96	0.93	0.9	0.91	0.83	0.8	0.92
Z4	0.97	0.94	0.92	0.93	0.86	0.84	0.94
Z5	0.95	0.92	0.89	0.9	0.81	0.79	0.91
Z6	0.98	0.96	0.93	0.94	0.88	0.85	0.95

features from medical images, while LSTM can capture time-dependent changes in physiological signals provided by IoT devices, such as fluctuating trends in heart rate and blood pressure. Through this combination, the CNN-LSTM model can not only capture subtle lesion features in images, but also identify long-term change patterns in physiological data, greatly improving the diagnostic accuracy of diseases.

The separate CNN and LSTM models perform well in processing single modal data, but the AUC value is slightly lower due to the lack of processing capacity for another modal data. CNN performs well in image processing, with AUC values ranging from 0.92 to 0.96, while LSTM performs well in processing time series data, but with AUC values only between 0.89 and 0.93 in the absence of image data. Traditional machine learning models, random forests, and support vector machines perform the worst, with AUC values ranging from 0.79 to 0.88, mainly because they rely on manual feature extraction and cannot fully exploit complex features in multimodal data.

The analysis results of Kappa coefficient and Matthews correlation coefficient are shown in Fig. 6.

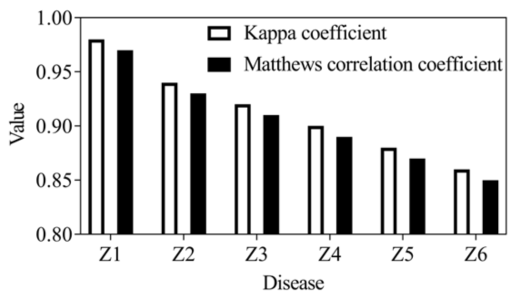

Fig. 6. Analysis results of Kappa coefficient and matthews correlation coefficient.

The CNN-LSTM model performs well in the diagnosis of multiple types of diseases, especially in healthy individuals and cardiovascular diseases. The Kappa coefficient and Matthews correlation coefficient are both close to 1.0, indicating that the model has very high classification accuracy and consistency for these two categories. A Kappa coefficient close to 1 means that the consistency between the model's predictions and the true labels is very good, avoiding the influence of random classification; MCC is a more comprehensive evaluation indicator that takes into account the balance between true positives, false positives, true negatives, and false negatives. With the increase of disease complexity, Kappa and MCC slightly decrease in chronic obstructive pulmonary disease, hypertension, and chronic kidney disease, but still remain above 0.85, demonstrating the robustness of the model in the diagnosis of complex diseases. CNN-LSTM can effectively capture features in imaging and physiological data, but the overall performance of the model may be affected by the imbalance of some datasets or the ambiguity of certain features. The CNN-LSTM model exhibits strong generalization ability and consistency when processing multimodal data, and has high diagnostic accuracy.

F. Ablation Experiment

The macro-average precision can measure the performance of multi-class classification, and the ablation experiment results are shown in Table III.

TABLE III. RESULTS OF ABLATION EXPERIMENTS

Fold number	CNN-LSTM (%)	CNN (%)	LSTM (%)
1	96	93	92
2	95	92	91
3	97	94	92
4	96	93	91
5	95	92	91
6	96	93	92
7	97	94	92
8	95	92	91
9	96	93	92
10	97	94	92

The macro-average precision of the CNN-LSTM model performs the best in 10 folds, maintaining between 95.0% and 97.0%, demonstrating its robustness and superior performance in multimodal medical data classification tasks. The macro-average precision of the CNN model ranges from 92.0% to 94.0%, slightly lower than that of the CNN-LSTM, indicating that the CNN model performs well in simple image processing but cannot fully utilize the information of temporal data. The macro-average precision of the LSTM model ranges from 91.0% to 92.0%, mainly due to its emphasis on temporal feature extraction but lack of CNN's ability to process image features. The CNN-LSTM model significantly improves the performance of multimodal data classification by combining the image feature extraction capability of CNN and the temporal feature processing advantage of LSTM, making it suitable for application in complex medical diagnosis scenarios.

G. Diagnosis Time

The diagnostic method in this article is compared with the traditional electronic medical record diagnostic method, and the comparison of diagnostic time is shown in Fig. 7.

For healthy individuals, this method only takes 15 seconds, while traditional diagnostic methods require 200 seconds, with a significant difference. In the diagnosis of cardiovascular diseases, this method takes 18 seconds, while the traditional method takes 211 seconds, showing a significant improvement in efficiency. Diabetes and chronic obstructive pulmonary disease take 16 seconds and 14 seconds respectively, which show faster response time compared with the traditional 156 seconds and 145 seconds. For hypertension and chronic kidney disease, the diagnostic method in this article is also more efficient, completing diagnosis in 19 seconds and 17 seconds

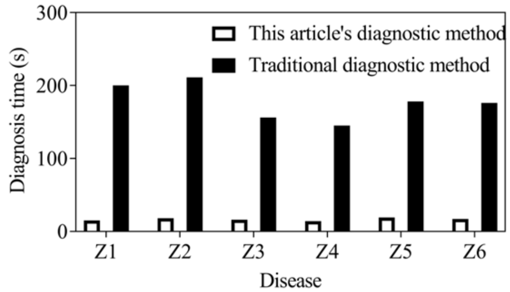

Fig. 7. Diagnosis time.

respectively, while traditional methods require 178 seconds and 176 seconds. This indicates that models based on CNN-LSTM can quickly and efficiently process multimodal data, especially in the context of the Internet of Things, greatly reducing diagnostic time and helping to monitor patients' health status in real-time and provide timely personalized treatment plans.

IV. Conclusions

This article proposes a multimodal diagnostic model based on CNN-LSTM, which significantly improves the accuracy and efficiency of chronic disease diagnosis by combining medical imaging data, temporal physiological data, and clinical text data. This model has achieved high accuracy in the classification tasks of healthy individuals and cardiovascular diseases, diabetes, chronic obstructive pulmonary disease, hypertension and chronic kidney disease, and is superior to traditional diagnostic methods. This achievement not only provides more precise diagnostic tools for clinical medicine, but also provides patients with faster health monitoring methods, which has important practical significance. This article combines multimodal data fusion with deep learning algorithms to promote the development of intelligent healthcare. Despite achieving a series of positive results, the research still has limitations, such as a relatively small sample size, which may affect the model's generalization ability. In addition, the performance of models in handling specific diseases may also be limited by the quality and diversity of input data. Future research can focus on expanding the sample size, enhancing the adaptability and robustness of the model, and exploring the combination of other deep learning architectures with traditional methods to further enhance the application potential of the model in complex clinical scenarios. Through continuous optimization and improvement, this study has laid the foundation for achieving more intelligent and personalized medical services.

Funding

There is no specific funding to support this research

Conflicts of Interests

Authors do not have any conflicts.

Data Availability Statement

No datasets were generated or analyzed during the current study.

Code Availability

Not applicable

Authors' Contributions

Wang Liyun is responsible for designing the framework, analyzing the performance, validating the results, and writing the article.

References

[1] N.E. Lee, M.M. Parker, and J.Q. Concepcion. An electronic medical record (EMR) prompt improves screening rates for metabolic conditions among children with obesity. *Obesity*, 31:1376–1382, 2023. https://doi.org/10.1002/oby.23257

[2] D. Setyadi and M. Nadjib. The Effect of Electronic Medical Records on Service Quality and Patient Satisfaction: A Literature Review. *J. Res. Soc. Sci. Econ. Manag.*, 2:2780–2791, 2023.

[3] A. Kebaili and S. Ruan. Deep learning approaches for data augmentation in medical imaging: a review. *J. Imaging*, 9:81, 2023. https://doi.org/10.3390/jimaging9040081

[4] T. Dhar, N. Dey, S. Borra, and R.S. Sherratt. Challenges of deep learning in medical image analysis—improving explainability and trust. *IEEE Trans. Technol. Soc.*, 4:68–75, 2023. https://doi.org/10.1109/TTS.2022.3170386

[5] S.H. Park. Methods for clinical evaluation of artificial intelligence algorithms for medical diagnosis. *Radiology*, 306:20–31, 2023. https://doi.org/10.1148/radiol.2023222525

[6] X. Zhao. Clinical applications of deep learning in breast MRI. *Biochim. Biophys. Acta Rev. Cancer*, 1878:188864, 2023. https://doi.org/10.1016/j.bbcan.2023.188864

[7] Q. Liu. Advances in the application of molecular diagnostic techniques for the detection of infectious disease pathogens. *Mol. Med. Rep.*, 27:1–14, 2023. https://doi.org/10.3892/mmr.2023.12862

[8] M. Tian, Z. Shen, X. Wu, K. Wei, and Y. Liu. The Application of Artificial Intelligence in Medical Diagnostics: A New Frontier. *Acad. J. Sci. Technol.*, 8:57–61, 2023.

[9] A. Adegoke. Patients' reaction to online access to their electronic medical records: the case of diabetic patients in the US. *Int. J. Appl. Sci. Current Future Res. Trends*, 19:105–115, 2023.

[10] S. Alam. An overview of blockchain and IoT integration for secure and reliable health records monitoring. *Sustainability*, 15:5660, 2023. https://doi.org/10.3390/su15075660

[11] M.Q. Alsudani. Blockchain-based e-medical record and data security service management based on IoMT resource. *Int. J. Pattern Recognit. Artif. Intell.*, 37:2357001, 2023. https://doi.org/10.1142/S0218001423570010

[12] D.K.R. Basani, B.R. Gudivaka, R.L. Gudivaka, and R.K. Gudivaka. Enhanced Fault Diagnosis in IoT: Uniting Data Fusion with Deep Multi-Scale Fusion Neural Network. *Internet of Things*, 101361, 2024. https://doi.org/10.1016/j.iot.2024.101361

[13] N.K.R. Panga. Applying Discrete Wavelet Transform for ECG Signal Analysis in IoT Health Monitoring Systems. *Int. J. Inf. Technol. Comput. Eng.*, 10(4):157–175, 2022.

[14] T. Ganesan and M.V. Devarajan. Integrating IoT, Fog, and Cloud Computing for Real-Time ECG Monitoring and Scalable Healthcare Systems Using Machine Learning-Driven Signal Processing Techniques. *[Journal Name]*, [Volume(Issue)]:[Pages], [Year]. [DOI or URL if available]

[15] R. Budda. Integrating Artificial Intelligence and Big Data Mining for IoT Healthcare Applications: A Comprehensive Framework for Performance Optimization, Patient-Centric Care, and Sustainable Medical Strategies. *Int. J. Manag. Res. Rev.*, 11(1):86–97, 2021.

[16] S.H. Grandhi. Enhancing Children's Health Monitoring: Adaptive Wavelet Transform in Wearable Sensor IoT Integration. *Curr. Sci. Humanit.*, 10(4):15–27, 2022.

[17] A. Padhi. Transforming clinical virology with AI, machine learning and deep learning: a comprehensive review and outlook. *VirusDisease*, 34:345–355, 2023. https://doi.org/10.1007/s13337-023-00780-0

[18] L. Gai, M. Xing, W. Chen, Y. Zhang, and X. Qiao. Comparing CNN-based and transformer-based models for identifying lung cancer: which is more effective? *Multimed. Tools Appl.*, 83:59253–59269, 2024. https://doi.org/10.1007/s11042-024-02963-x

[19] D.C. Edara. Sentiment analysis and text categorization of cancer medical records with LSTM. *J. Ambient Intell. Humaniz. Comput.*, 14:5309–5325, 2023. https://doi.org/10.1007/s12652-022-04101-0

[20] S. Liu. New onset delirium prediction using machine learning and long short-term memory (LSTM) in electronic health record. *J. Am. Med. Inform. Assoc.*, 30:120–131, 2023. https://doi.org/10.1093/jamia/ocac194

[21] Z.C. Wang. Deep learning for discrimination of hypertrophic cardiomyopathy and hypertensive heart disease on MRI native T1 maps. *J. Magn. Reson. Imaging*, 59:837–848, 2024. https://doi.org/10.1002/jmri.30078

[22] A.M. Salih, E. Ruiz Pujadas, and V.M. Campello. Image-Based Biological Heart Age Estimation Reveals Differential Aging Patterns Across Cardiac Chambers. *J. Magn. Reson. Imaging*, 58:1797–1812, 2023. https://doi.org/10.1002/jmri.30201

[23] S.S. Johnson, M.R. Murty, and I. Navakanth. A detailed review on word embedding techniques with emphasis on word2vec. *Multimed. Tools Appl.*, 83:37979–38007, 2024. https://doi.org/10.1007/s11042-023-14004-1

[24] G. Curto. Are AI systems biased against the poor? A machine learning analysis using Word2Vec and GloVe embeddings. *AI Soc.*, 39:617–632, 2024. https://doi.org/10.1007/s10209-023-00911-2

[25] S. Khorram and N. Jehbez. A hybrid CNN-LSTM approach for monthly reservoir inflow forecasting. *Water Resour. Manag.*, 37:4097–4121, 2023. https://doi.org/10.1007/s11269-023-03057-7

[26] Z. Alshingiti. A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. *Electronics*, 12:232, 2023. https://doi.org/10.3390/electronics12010232