Article

Research on Intelligent Analysis and Recognition System of Medical Data Based on Deep Learning

Xinzhe Yuan [1,*]

[1] Northeastern University, San Jose, California, United States
* Correspondence: Xinzhe Yuan, Northeastern University, San Jose, California, United States

Abstract: With the explosive growth in the amount of medical data, traditional data analysis methods can hardly meet the demand, especially in complex medical tasks such as disease diagnosis, patient monitoring and personalized treatment. A deep learning-based system for intelligent analysis and recognition of medical data has emerged, which is capable of automatically extracting features from massive data and efficiently learning through multi-layer neural networks, thus significantly improving diagnostic accuracy and medical efficiency. The system not only covers a variety of deep learning models, such as recurrent neural networks, long and short-term memory networks, gated recurrent units, attention mechanisms, and graph convolutional neural networks, but also combines pre-training models and autoencoders to achieve more accurate data analysis and recognition. Through the combined application of these technologies, it can help doctors make quick and accurate decisions to improve the treatment outcome and quality of life of patients.

Keywords: deep learning; medical data; intelligent analytics; recognition systems research

Published: 23 January 2025

1. Introduction

In recent years, the development of healthcare informatization has led to a dramatic increase in the amount of medical data, which contains a wealth of information and urgently needs to be efficiently utilized. However, traditional data analysis methods are not capable of handling large-scale, high-dimensional and complex time-series medical data, and it is difficult to extract useful information in a short time. Based on this background, deep learning techniques show unique advantages, especially their powerful ability in automatic feature extraction and model self-learning. Deep learning models, such as Recurrent Neural Networks (RNN), Long Short-Term Memory Networks (LSTM), and Gated Recurrent Units (GRU), are able to effectively process temporal data, while Attention Mechanisms and Graph Convolutional Neural Networks (GCN) excel in information fusion and relationship modeling. Pre-trained models, such as Transformer and BERT, further enhance the generalization ability and performance of the models. The fusion application of these technologies not only provides new solutions for intelligent analysis and recognition of medical data, but also shows great potential in clinical diagnosis and patient management. Through the systematic research and innovative application of these technologies, the data processing capability and intelligence level in the medical field will be comprehensively improved, laying a solid foundation for the realization of precision medicine [1].

2. Research Based on Deep Learning Algorithms

2.1. Deep Learning Models

2.1.1. Recurrent Neural Networks

Recurrent Neural Network (RNN) hidden layers pass historical information between them through connections and are able to capture and integrate historical data from the output of the previous moment as part of the input of the current moment. This structure allows the RNN to handle sequence data of arbitrary length while better capturing long-term dependencies in the sequence. This feature of RNNs is particularly important in intelligent analysis and recognition systems for medical data, because many medical data are time-series data, such as electrocardiograms and blood pressure monitoring data. With the RNN model, the system can more accurately identify changes in the patient's state and predict potential health risks. the working principle of RNN is shown in Figure 1. the value x_t of the input vector x at each time step t is passed to the hidden layer through the weight matrix U. The value h_t of the hidden layer at each time step depends not only on the current input x_t, but also on the value of the hidden layer at the previous time step h_{t-1}. the hidden layer's output ht is passed to the output layer through the weight matrix V, which finally generates the output o_t. This mechanism enables RNNs to effectively utilize historical information when processing medical data, improving the accuracy and timeliness of diagnosis [2].

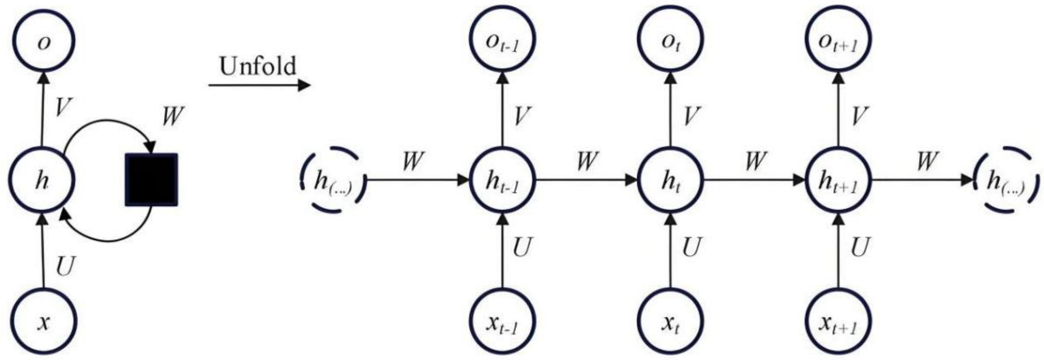

Figure 1. structure of the RNN model.

Although RNN performs well in certain tasks, it still suffers from gradient vanishing and gradient explosion when dealing with very long sequences, which limit its effectiveness in certain scenarios. Therefore, further optimization of RNN models or introduction of other improvement methods, such as Long Short-Term Memory Networks (LSTM) and Gated Recurrent Units (GRUs), becomes a key direction to enhance the performance of intelligent analysis and recognition systems for medical data.

2.1.2. Long- and Short-Term Memory Networks

Long Short-Term Memory (LSTM) is an important variant of recurrent neural networks, which can effectively overcome the problems of gradient vanishing and gradient explosion encountered by traditional RNNs when dealing with long sequential data by introducing forgetting gates, input gates, and output gates. The application of LSTM model is particularly important in the intelligent analysis and recognition system of medical data, because many medical data have time-series characteristics, such as patient's physiological monitoring data, medical records, etc. The structural design of LSTM enables it to better preserve and utilize the historical information, which improves the accuracy of diagnosis and the reliability of prediction [3]. Specifically, the forgetting gate ft of LSTM decides which information should be forgotten from the memory cell ct-1 by calculating the hidden state ht-1 of the current input xt and the previous time step, which is shown in Equation (1):

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f \quad (1)$$

Input gate i_t controls the information related to the current input and adds it to the memory cell, which is calculated as shown in Equation (2):

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i \quad (2)$$

The output gate o_t is responsible for deciding which information is output from the memory cell to the hidden state h_t, which is output to the next layer of the network or directly used as the prediction result through techniques such as embedded selectivity, the calculation of which is shown in Equation (3):

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o \quad (3)$$

The LSTM model under this mechanism is able to not only effectively utilize the historical information but also dynamically adjust the preservation and output of the information to better capture the complex temporal dependencies when processing medical data. In addition, improved versions of LSTM such as dual-directional LSTM and stacked LSTM further enhance the performance of the model in processing multidimensional and high-complexity medical data, which makes the intelligent analysis and recognition system of medical data more efficient and accurate [4].

2.1.3. Gated Circulation Units

The gated recurrent unit (GRU), as a highly optimized recurrent neural network model, shows significant advantages in processing sequence data. The model was proposed by Cho et al. in 2014, and is mainly used to solve the common gradient vanishing and gradient explosion problems in traditional RNNs.GRU is cleverly simplified in its structure, as shown in Figure 2, by fusing the forgetting gate and the input gate in the LSTM into a single Update Gate, and at the same time, a Reset Gate is introduced for real-time control of the flow of information in real time [5]. The role of the Update Gate is to decide which information should be retained in the memory state of the previous time step, while the Reset Gate controls which information will be forgotten, thus avoiding the competition between multiple gating units in the LSTM. This design not only simplifies the complexity of the model, but also significantly reduces the training time and improves the efficiency and accuracy of the model when dealing with long sequential data.

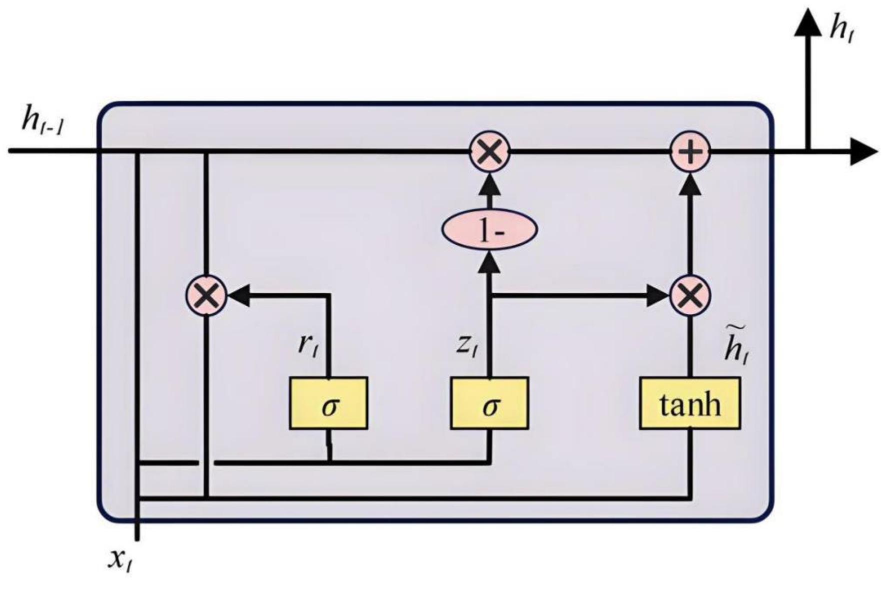

Figure 2. Gating cycle unit structure.

These features of GRUs are especially critical in intelligent analysis and recognition systems for medical data, which often have complex temporal characteristics and long period dependencies. For example, a patient's physiological monitoring data and medical records may contain information at multiple points in time, and GRU is able to capture the temporal dependency of this information more effectively, thus improving the accuracy of diagnosis and prediction. Not only that, GRUs also excel in a variety of areas such as natural language processing and speech recognition, especially in tasks that require long periods of memory, such as text generation and machine translation.

2.1.4. Attention Mechanisms

Attention Mechanism (AM) enhances the comprehension and expressiveness of the model by dynamically assigning weights to focus on key parts of the input sequence. By inputting the input sequence $H=[h_1,h_2, h_3,..,h_n]$ into the attention mechanism computation, the correlation between the query vector q and each input h_i can be computed using the scoring function s to derive a score. Subsequently, these scores are normalized by the softmax function to obtain the attention distribution $a=[a_1,a_2,a_3,\ldots,a_n]$ of the query vector q on different inputs h_i, where each of the values corresponds to the original inputs H one by one. The specific calculation is as in Equation (4):

$$a_i = soft\,max(\,s(h_i, q)) = \frac{exp(s(h_i,q))}{\sum_{j=1}^{n} exp(s(h_j,q))} \tag{4}$$

This "soft" attention distribution allows the model to selectively extract and weight input information to better understand the semantics of the input data, as illustrated in Figure 3. In the intelligent analysis of healthcare data, patient physiological monitoring data and medical records may contain information from multiple points in time, and the attention mechanism is able to identify which data points are the most critical for a particular task (e.g., disease diagnosis). By calculating the attention distribution, the model can assign more weights to these critical data points, thus improving the accuracy of prediction and classification [6]. In addition, by weighting and summing the input information according to the attention distribution, the resulting context vector can more accurately reflect the information that the model needs to pay attention to at the moment, which is calculated as in Equation (5):

$$\text{context} = \sum_{i=1}^{n} a_i \cdot h_i \tag{5}$$

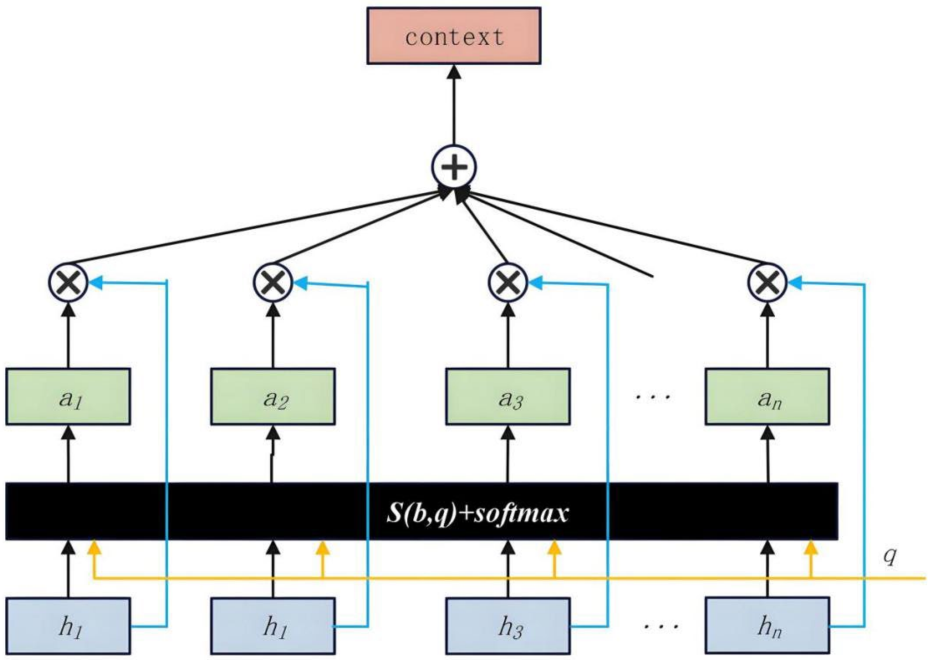

Figure 3. Attention mechanism calculation process.

This approach not only enhances the model's ability to understand the input data, but also improves the efficiency and accuracy of the model in processing long sequential data. Attention mechanisms are also widely used in natural language processing, speech recognition, and other fields, especially in tasks that require understanding of context and long dependencies, such as text generation and machine translation. By dynamically adjusting the weights, the attention mechanism can provide more reliable and efficient performance in a variety of complex tasks.

2.2. Pre-Training Models

2.2.1. Transformer Model

The Transformer model, as an innovative deep learning architecture, has made significant breakthroughs in the field of Natural Language Processing (NLP) since its introduction. The model abandons the time dependency of traditional recurrent neural networks (RNNs) and long short-term memory networks (LSTMs), and introduces the Self-Attention Mechanism to parallelize the processing of all positions in the input sequence, thus dramatically improving the efficiency and performance of the model.The encoder in the Transformer Model -Decoder architecture further enhances its performance capabilities in complex tasks by stacking multiple Self-Attention and Feedforward Neural Network layers. For example, in multimodal data fusion, Transformer can simultaneously process multiple types of data, such as text, image, and speech, and effectively correlate and fuse information from different modalities through the self-attention mechanism [7].

2.2.2. BERT Pre-Training Model

The emergence of BERT (Bidirectional Encoder Representations from Transformers) pre-training model has greatly promoted the development of natural language processing technology, and its bi-directional encoding mechanism and the design of Transformer architecture have enabled it to show excellent performance in medical text analysis. The BERT model is shown in Figure 4.

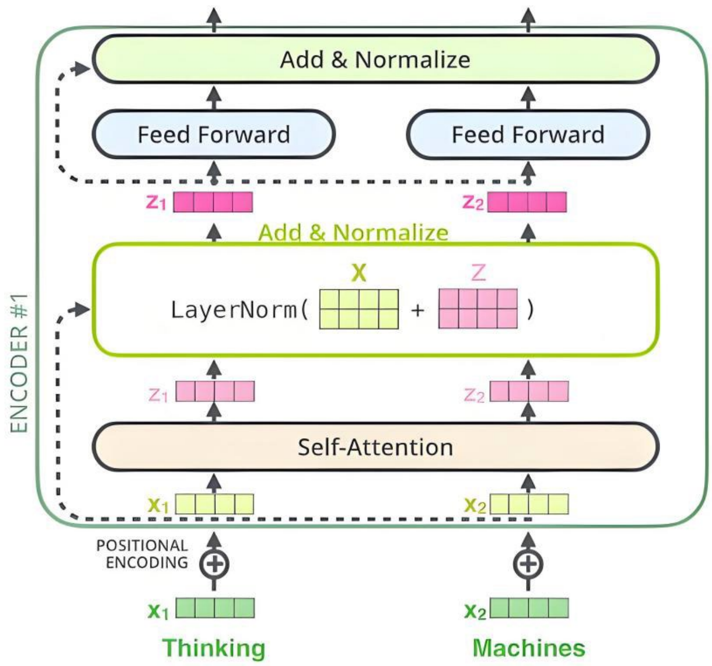

Figure 4. BERT model.

While the diversity and complexity of medical data often overwhelm traditional methods for processing such data, BERT is able to capture deep semantic information in text through large-scale unsupervised pre-training, providing a robust feature representation for subsequent fine-grained tasks such as named entity recognition, relationship extraction, and text categorization.At the heart of BERT lies the bi-directional Transformer encoder, which is capable of taking contextual information into account during the pre-training phase to generate richer word vectors. The encoder is capable of generating richer word vectors by simultaneously considering contextual information in the pre-training phase. In addition, BERT employs two pre-training tasks, Masked Language Model (MLM) and Next Sentence Prediction (NSP), where MLM improves the model's sensitivity to context by randomly masking a portion of the words in the input text and predicting those words; NSP enhances the model's sensitivity to the context of a sentence by predicting the coherence of the two sentences. coherence, which enhances the model's understanding of inter-sentence relations. In practical applications, BERT not only improves the accuracy of diagnosis, but also assists doctors in medical record management and patient monitoring, which greatly improves the efficiency and intelligence of medical work.

2.2.3. WoBERT Re-Trained Model

WoBERT (Weighted BERT) pre-training model plays a key role in the intelligent analysis and recognition system of medical data, which optimizes the performance of BERT model on domain-specific data by introducing a weight adjustment mechanism. As a Transformer-based pre-training model, WoBERT inherits the bi-directional encoding and multi-layer Transformer architecture of BERT, and at the same time introduces a vocabulary weighting strategy in the pre-training phase, which enables the model to pay more attention to words and phrases that are of higher importance in the medical domain. This weighting mechanism not only improves the feature extraction ability of the model in medical text, but also shows significant performance advantages in tasks such as named entity recognition and relationship extraction.The pre-training data of WoBERT contains a large number of professional medical literature and clinical records, which are finely cleaned and annotated to ensure the effectiveness of the model in the medical domain. In practical applications, WoBERT is able to better capture key information in medical texts, such as symptom descriptions, disease names, and treatment options, through fine-grained lexical weighting, thus providing doctors with more accurate assisted diagnosis and decision support [8].

2.3. Autoencoder Feature Extraction Method

Autoencoder, as an unsupervised learning method, shows unique advantages in the intelligent analysis and recognition system of medical data, which can effectively extract the deep features of the data through the process of self-encoding and decoding, and provide high-quality inputs for the subsequent classification, clustering, and prediction tasks, and the Autoencoder model is shown in Figure 5.

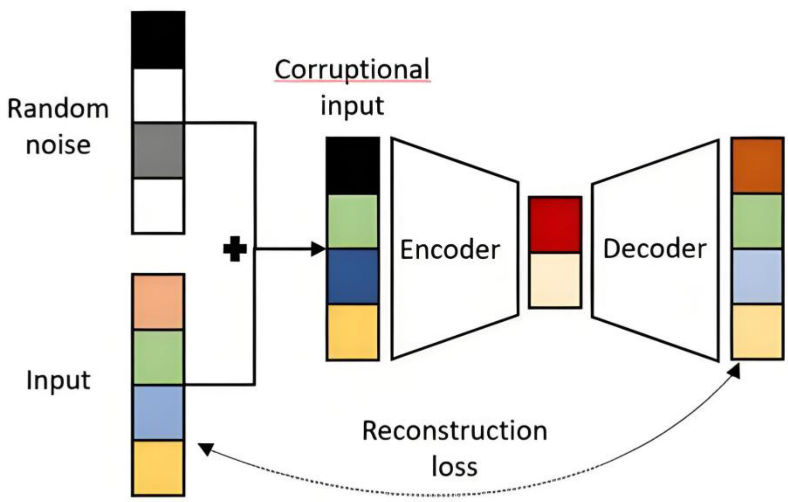

Figure 5. Autoencoder model.

In medical applications, Autoencoder significantly improves the interpretability of data and the generalization ability of models by providing a low-dimensional representation of multimodal data such as medical records, imaging data and physiological signals. Especially when dealing with high-dimensional and complex medical data, Autoencoder is able to compress the original data into a more compact form through the encoder while retaining the key information, and then restore the data through the decoder to ensure the completeness and accuracy of the information. This bi-directional structure not only reduces noise and redundancy in the data during the pre-processing stage, but also captures implicit patterns in the data during feature extraction, such as disease trends and changes in patient health status. In practice, Autoencoder can be combined with deep learning models to form a hybrid architecture, which further enhances the performance of the system.

3. Design of Intelligent Analysis and Recognition System for Medical Data

3.1. System Functional Structure

This system is mainly composed of four core functional modules: data processing subsystem, prior violation management subsystem, big data analysis subsystem and configuration management subsystem, as shown in Figure 6. The data processing subsystem is responsible for the collection, cleaning, standardization and preprocessing of medical data to ensure the quality and consistency of the data and provide a reliable foundation for the subsequent analysis and identification tasks. The prior violation management subsystem, on the other hand, focuses on privacy and security protection of medical data, ensuring that the system complies with relevant laws, regulations and ethical requirements during data processing by identifying and managing potential data security and privacy risks. The big data analysis subsystem is the core of the system, utilizing advanced machine learning and deep learning algorithms to intelligently analyze massive medical data, extract valuable features and patterns, and support applications such as medical diagnosis, disease prediction, and patient management. The configuration management subsystem is responsible for system configuration and optimization, including algorithm selection, parameter tuning, and system monitoring, to ensure that the system can operate efficiently under different environments and demands. Although each of these four subsystems takes on different functions, they access and collaborate with each other in data flow and task execution, forming a unified and efficient system framework that fully supports intelligent analysis and identification of medical data [9].

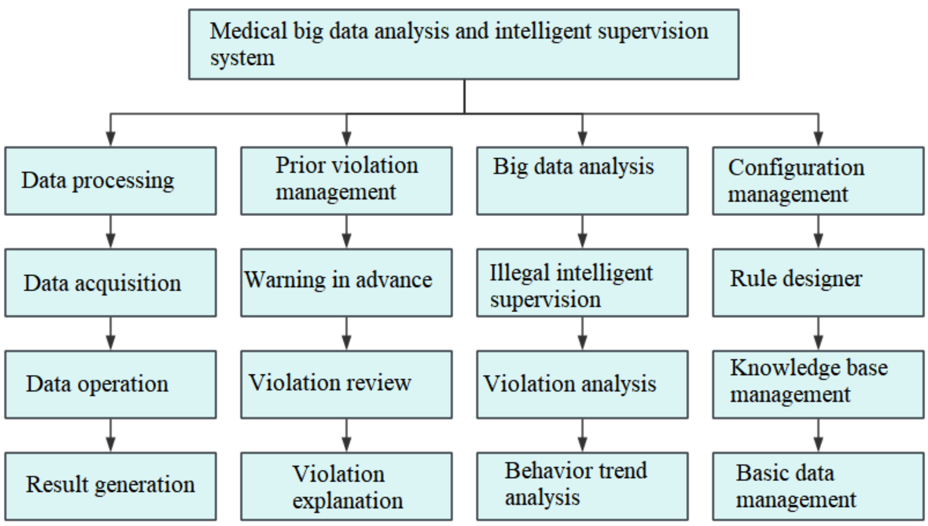

Figure 6. System functional modules.

3.2. System Network Topology

The network topology of the medical data intelligent analysis and identification system is designed to achieve efficient data circulation and functional synergy, ensuring that the subsystems can operate robustly and work together in a complex environment. The system consists of a data processing subsystem, a prior violation management subsystem, a big data analysis subsystem, and a configuration management subsystem, and the subsystems are connected through a high-bandwidth, low-latency network to form an organic whole. The data processing subsystem is responsible for receiving and pre-processing raw data from a wide range of medical devices and systems to ensure data accuracy and consistency. The prior breach management subsystem is tightly coupled with the data processing subsystem to monitor the data flow in real time and deal with potential privacy and security issues to guarantee data compliance. The big data analysis subsystem receives pre-processed data through an efficient network transmission channel and intelligently analyzes it using advanced machine learning and deep learning algorithms to extract valuable features and patterns.

4. Experimental Results and Analysis

4.1. Experimental Data Set

In this regard, the newly-released CCKS 2019 dataset is chosen as the experimental dataset; it is widely used in medical named entity recognition research, containing 1379 Chinese medical text data, of which 1000 are used for training and 379 for testing. To enhance the evaluation of model performance, training data are further divided into a 70%-training set and a 30%-validation set. The dataset has been labeled in six entity types: disease and diagnosis; surgery; drugs; anatomical sites; imaging; and laboratory tests. The data of each entity type is reasonably well distributed in the training set, verification set, and test set, which ensures the reliability and ability of the result in further generalization through experiments. It can be seen, through detailed statistical analysis, that the distributions of data relative to different entity types are shown in Table 1, which provides an important reference in designing the following experiments. For instance, data for diseases and diagnoses constituted the largest proportion in the database, followed by those on drugs and anatomical sites. Such a distribution of data not only reflects the actual characteristics in the clinical records but also provides rich information to facilitate model training and tuning. Through the application of the intelligent analysis and recognition system of medical data based on the deep learning network, features of such large datasets also can be analyzed, thus allowing the structure and parameters of the model to be optimized to raise the overall recognition accuracy [10].

Table 1. Entity types and number of ccks 2019 dataset.

Entity type	Training set	Validation set	Test set
Disease and diagnosis	3645	567	1808
operation	908	121	162
medicine	1593	229	485
Anatomical site	7158	1268	3094
Imaging examination	888	81	348
Laboratory test	991	204	590

4.2. Analysis of Experimental Results

The analysis of the experimental results indicates excellent performance of the proposed method on the CCKS 2019 dataset, with the comparison against the current mainstream methods showing significant advantages of this study's proposed method in multiple indicators. Table 2 provides detailed experimental data, which not only validates the method with which this research was performed but also gives an in-depth technical interpretation. The name entity recognition tasks provide a significant advantage to models that include CRF (Conditional Random Fields) when compared with those that only use CNNS (Convolutional Neural Networks), indicating that the CRF can model the dependencies existing between the labels significantly well while improving the performance of the model. Besides, the accuracy associated with the present model coupled with the BERT pre-training model is extremely high compared to the previous two models, indicating a stronger capability of BERT on semantic understanding and context awareness concerning medical texts. Enjoyment by the entity recognition model with GCN(Since this greatly improves the accuracy of named entity recognition by constructing graphical relationships among entities). The method included bidirectional GCN coding technology and incorporated the attention weight distribution mechanism. Experimental data also show that the proposed method achieves the best metrics concerning accuracy of P, recall R, and F, thus highlighting that the bidirectional GCN coding and attention weight distribution have significant optimization effect on the entity-oriented tasks. The experimental findings thus provide strong support to further refine the intelligent analysis and recognition system of medical data based on deep learning for different uses.

Table 2. Network parameters.

Method model	P/%	R/%	F1/%
CNN	67.98	70.12	69.03
CNN+CRF	74.94	74.72	74.83
BERT+BiLSTM+CRF	84.56	84.19	85.13
BERT+BiLSTM+GCN +CRF	87.37	85.80	86.57
Textual method	87.95	88.18	88.06

5. Conclusion

In summary, the deep learning-based medical data intelligent analysis and recognition system shows great potential in medical data processing through the comprehensive application of multiple models and methods. It is not only capable of automatically extracting features from a large amount of complex data to improve diagnostic accuracy, but also effectively assists doctors in decision-making and improves medical efficiency and patient satisfaction. In the future, with the continuous progress of technology and the continuous accumulation of medical data, the system will be further optimized to provide more powerful support for precision medicine. It is hoped that the research in this paper can inspire more scholars and practitioners to explore and innovate in this field, and jointly promote the development of medical informatization and intelligence, which will ultimately benefit the majority of patients.

References

1. H. Sun, Z. Liu, G. Wang, W. Lian and J. Ma, "Intelligent Analysis of Medical Big Data Based on Deep Learning," in *IEEE Access*, vol. 7, pp. 142022-142037, 2019, doi: 10.1109/ACCESS.2019.2942937.
2. W. -J. Chang, L. -B. Chen, C. -H. Hsu, C. -P. Lin and T. -C. Yang, "A Deep Learning-Based Intelligent Medicine Recognition System for Chronic Patients," in *IEEE Access*, vol. 7, pp. 44441-44458, 2019, doi: 10.1109/ACCESS.2019.2908843.
3. W. Liu *et al.*, "Research on Medical Data Feature Extraction and Intelligent Recognition Technology Based on Convolutional Neural Network," in *IEEE Access*, vol. 7, pp. 150157-150167, 2019, doi: 10.1109/ACCESS.2019.2943927.
4. J. Xu, Z. Hu, J. Zou and A. Bi, "Intelligent Emotion Detection Method Based on Deep Learning in Medical and Health Data," in *IEEE Access*, vol. 8, pp. 3802-3811, 2020, doi: 10.1109/ACCESS.2019.2961139.
5. D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *CoRR*, vol. abs/1409.0473, pp. 1–15, 2014, doi: 10.48550/arXiv.1409.0473.
6. X. Li, D. Li, Y. Deng, *et al.*, "Intelligent mining algorithm for complex medical data based on deep learning," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 2, pp. 1667–1678, 2021, doi: 10.1007/s12652-020-02239-w.

7. S. Yang, F. Zhu, X. Ling, *et al.*, "Intelligent health care: Applications of deep learning in computational medicine," *Front. Genet.*, vol. 12, p. 607471, 2021, doi: 10.3389/fgene.2021.607471.
8. M. Masud, G. Muhammad, H. Alhumyani, *et al.*, "Deep learning-based intelligent face recognition in IoT-cloud environment," *Comput. Commun.*, vol. 152, pp. 215–222, 2020, doi: 10.1016/j.comcom.2020.01.050.
9. X. Li, H. Wang, H. He, *et al.*, "Intelligent diagnosis with Chinese electronic medical records based on convolutional neural networks," *BMC Bioinform.*, vol. 20, pp. 1–12, 2019, doi: 10.1186/s12859-019-2617-8.
10. W. Li, Y. Chai, F. Khan, *et al.*, "A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system," *Mob. Netw. Appl.*, vol. 26, pp. 234–252, 2021, doi: 10.1007/s11036-020-01700-6.