European Journal of AI, Computing & Informatics

Vol. 1 No. 2 2025

Article Open Access

Discussion on Using RNN Model to Optimize the Accuracy and Efficiency of Medical Image Recognition

Minkang Zhang [1,*]

[1] Clinical Diagnostics Group (CDG) Software Team, Bio-Rad Laboratories, Hercules, CA, 94547, USA
[*] Correspondence: Minkang Zhang, Clinical Diagnostics Group (CDG) Software Team, Bio-Rad Laboratories, Hercules, CA, 94547, USA

Received: 22 May 2025
Revised: 03 June 2025
Accepted: 16 June 2025
Published: 20 June 2025

Abstract: With the continuous development of artificial intelligence technology, especially represented by deep learning, recurrent neural networks have made revolutionary breakthroughs in medical image recognition. This paper first introduces the concept of RNN pattern and its application in medical image recognition. By analyzing various applications of RNN in medical image classification, we explore how to improve the accuracy and computational efficiency of medical image recognition by optimizing the RNN model. Specifically, the gating mechanism, convolutional neural network (CNN) construction, lightweight technology and other optimization strategies, multimodal learning and attention mechanism input are discussed. Finally, the prospects and challenges of the RNN model in medical image recognition are summarized, and the future research directions in this field are also discussed.

Keywords: medical image recognition; precision; efficiency

1. Introduction

With the continuous increase of medical image data, the pressure of manual image analysis in the face of massive image information is rising. Digital image analysis based on deep learning has become an important method for medical image analysis. At present, the most common ones are convolutional neural network and recurrent neural network. Among them, recurrent neural networks can improve the accuracy and efficiency of diagnosis due to their strong capability in processing time series data. Although recurrent neural network can describe the processing of time series data skillfully, there are still many problems in the recognition of medical images, such as diverse image types, high complexity and high computing cost.

2. Theoretical Basis of RNN Model

2.1. Recursive Structure and Time Series Data Processing

An RNN is a neural network designed for time series data that can retain prior information and incorporate it into the current computational step. This allows the RNN to use the data from the previous frame in each image frame to extract the features of that frame, thus capturing the timing dependencies between image sequences. Compared with general neural networks, RNN can better process time series medical images through recursive structure, and can accurately grasp the dynamic changes of the pathological region between different sequences to enhance the accuracy of image interpretation. However, due to the phenomenon of gradient disappearance or gradient explosion in general RNNS, long-term dependence is difficult to maintain for long sequence images. To solve this problem, gated cyclic unit GRU and recurrent neural network LSTM put forward the idea of gating, which can selectively control the information flowing through and enhance long-term dependence. Therefore, RNN can be better applied in the processing of time series medical images, especially in the analysis of disease development and the diagnosis of a series of images.[1]

2.2. Information Transfer and Memory Mechanism

The mechanism of information transfer and memory is the core feature of recurrent neural network (RNN). The RNN uses a recursive structure where each step merges the new input with the previous hidden state and updates the hidden state, thereby preserving past information for future use (see Figure 1). This mechanism makes it possible for RNN to be effectively applied to the processing of time series data, and it has achieved good results in medical image analysis. For example, in CT or MRI image series, the development of lesions is often an evolutionary process, and RNNS can well describe the progression of the disease by combining the lesions in previous stages.

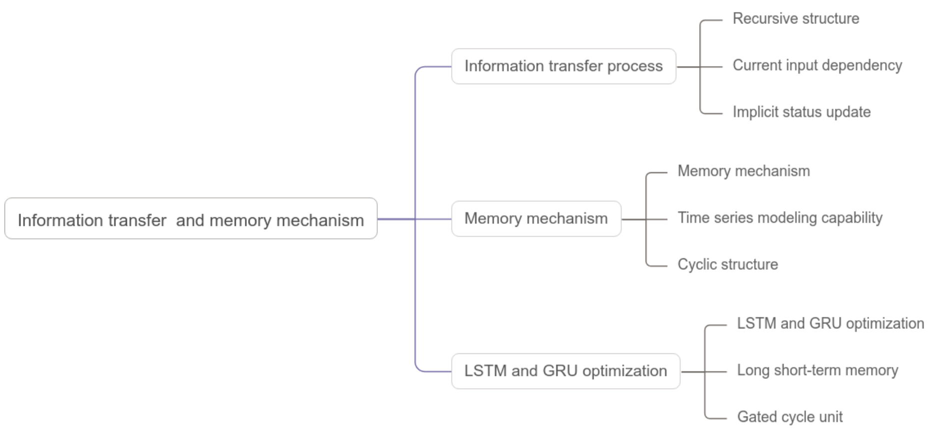

Figure 1. Basic framework of information transfer and memory mechanism.

However, the traditional RNN processing of long sequences will appear the phenomenon of gradient disappearance or explosion. For this reason, LSTM and GRU were born, in which the gating mechanism limits the output of information well, retains long-term relevant information, and improves the ability of RNN to handle long sequences. Through these constructions, RNNS perform better in medical image sequences, so the accuracy of early diagnosis of diseases is higher.[2]

3. Application of Rnn Model in Medical Image Recognition Accuracy and Efficiency

3.1. Lesion Detection in Dynamic Medical Image Sequences

Medical images that capture early signs of disease often appear in dynamic imaging modalities such as CT scans, MRI videos, and ultrasound sequences. These image sequences inherently contain time-series data, revealing the evolution of pathological tissues over time. Sequential data arranged in chronological order is well-suited for learning and analysis by RNN models, which can effectively extract temporal features from image sequences. Therefore, the LSTM and GRU models in RNN can integrate the information of past images while analyzing the current images, so as to locate the lesion more accurately. For example, lesions in dynamic CT scan image sequences will show different shape features with different image sequences due to the existence of sequence information, but the RNN model can find the site of the lesion by virtue of its recognition ability of time series, and timely give appropriate diagnostic results according to the changes in the state of the lesion. In addition, the time series characteristics of RNN help it distinguish the growth rate, growth frequency and spread direction of the lesion, which has a crucial impact on the early diagnosis of the disease and personalized treatment plan design. Through the optimization of RNN structure and training method, the accuracy and efficiency of lesion detection can be further improved.

3.2. Auxiliary Generation of Medical Image Report

The increasing volume of medical imaging data has placed a heavy burden on radiologists to interpret these images. However, the traditional production of image reports depends on the professional ability and practical experience of doctors, which is a time-consuming and subjective process influenced by human factors. In order to reduce the burden of doctors and improve the efficiency and accuracy of making image reports, recurrent neural networks are used to assist in generating image reports. Recurrent neural networks can correlate medical images with their corresponding textual reports. By training the image and matching text reports on a large scale, the recurrent neural network can learn to associate the attributes of the image with the text of the image of the described lesion. By learning the results, the recurrent neural network can receive a new medical image and automatically **generate** a matching image report. For example, when receiving a chest and lung X-ray image, the recurrent neural network model can publish a report with the analysis of the image results, including the suspected lesion site, type, and possible diagnostic recommendations.[3]

3.3. Classification and Identification of Fine-Grained Lesions

In terms of medical imaging, fine-grained pathologic classification has important research significance, mainly for the fine discrimination and classification of relatively small diseases, such as the early diagnosis and recognition of breast cancer, lung cancer, brain tumor and skin cancer. For efficient and accurate path classification, the application of deep learning is crucial, including convolutional neural network (CNN), RNN, SVM, reinforcement learning, etc. Through efficient feature extraction, CNNs can effectively identify core features of subtle diseases in images, while RNNs are capable of modeling disease progression and temporal changes using sequence data. However, problems with fine-grained pathologic classification can lead to misjudgments, and poor generalization may limit the model's performance when applied to large and diverse datasets. To solve the above problems, multimodal learning, attention mechanism, transfer learning and other methods can improve the classification accuracy, addressing sample imbalance and improving model generalization can lead to significant performance improvements. The details are shown in Table 1.

Table 1. Classification and recognition of fine-grained lesions.

Classification and recognition of fine-grained lesions	content
Lesion type	Breast cancer, lung cancer, brain tumor, skin cancer and other minor lesions
Common technique	Convolutional neural networks (CNN), recurrent neural networks (RNN), support vector machines (SVM), reinforcement learning
advantage	It can accurately detect subtle lesions, improve the accuracy of early diagnosis and reduce the misdiagnosis rate.

Challenges faced	The image features of small lesions are difficult to distinguish, the noise interference is serious, and the generalization across data sets is a problem.
Latest progress	Multimodal learning, attention mechanism, transfer learning and other techniques are introduced to improve the accuracy and solve the problem of sample imbalance.

It can be concluded from the above table that fine-grained lesion classification is mainly used for the detection of small lesions such as breast cancer and lung cancer. Common techniques include CNN, RNN, SVM, etc. Fine-grained lesion classification can improve the accuracy of early diagnosis but faces challenges such as difficulty in distinguishing image features and susceptibility to noise interference. Recent advances have introduced multimodal learning and attention mechanisms.

3.4. Joint Modeling with CNN to Improve Efficiency and Accuracy

Although RNN can process time series data, it cannot extract individual features as effectively as CNN. In recent years, CNN and RNN have been mixed in medical image analysis to avoid this problem. This approach ensures that advanced image features (such as edges, textures, and shapes) are fully extracted, thereby enhancing the accuracy and efficiency of medical image analysis. In the process of medical image recognition, CNN mainly extracts advanced features from the image input, and then RNN makes in-depth analysis of features through time sequence information to find the change information of the lesion site. Therefore, the hybrid model can process static image data and sequential data simultaneously to achieve improved image classification and disease detection outcomes.

4. Strategies for Optimizing the Accuracy and Efficiency of Medical Image Recognition Using RNN Models

4.1. Introducing Gating Mechanism to Optimize Model Performance

Introducing the gating mechanism to optimize model performance is a key step in enhancing the effectiveness of recurrent neural networks in complex tasks. RNN uses recursion to transmit information. When it is used to analyze long sequence data, it will encounter gradient disappearance or gradient explosion in the learning process, resulting in deviations in the understanding of long-term correlation of the data. In order to solve this problem, LSTM and GRU introduce gating mechanism to control the process of transferring information and forgetting information, so as to improve the learning ability of RNN. LSTM precisely controls information through forget gates, input gates, and output gates, retaining long-term information and preventing degradation problems associated with traditional RNNs. GRU provides an efficient solution by simplifying the architecture of LSTM, combining the forget and input gates into a single update gate. This method enables RNNs to retain valuable information across long time series and overcome the limitations of traditional RNNs in modeling long-term dependencies. For the analysis of medical dynamic image data, the gating mechanism is also used in the time history positioning of lesions, which improves the early accuracy of the model. In contrast, the gated systems of LSTM and GRU can not only alleviate the problem of gradient disappearance, but also improve the modeling ability of the model's long-range dependent structure, which can be better applied to complex data environments, as shown in Table 2.

Table 2. Strategies and applications of introducing gating mechanism to optimize RNN model performance.

Gating mechanism is introduced to optimize model performance	content
The role of the gating mechanism	The gating system enhances the learning capability of the RNN model by regulating information transmission and forgetting, thereby mitigating the issues of gradient vanishing and exploding.
Optimization of LSTM and GRU	LSTM accurately regulates the information flow through it by adding forgetting gates, input gates and output gates to retain long-term dependent information, while GRU simplifies the structure of LSTM and improves its computing performance by integrating forgetting gates and input gates.
Improve the ability to learn long sequences	Through the gating mechanism, RNNS can effectively capture meaningful features when learning long time series data, which solves the difficult problem of traditional RNNS learning long time sequential data. In medical image data detection, RNN can effectively capture the time series evolution of lesions.
advantage	It can solve the gradient disappearance in conventional RNN processing, and improve the modeling and prediction effect of the model for long-distance dependence. It is mainly suitable for processing complex medical image data, especially dynamic medical image data.
Application field	It is mainly used in tasks that require long-term dependencies, such as dynamic medical image analysis, speech recognition, and natural language processing

According to the contents of the above table, it can be drawn that the introduction of gating mechanisms (such as LSTM and GRU) to optimize the RNN model can solve the problem of gradient disappearance by accurately controlling information transmission and forgetting, and improve the performance of the model in long sequence learning, especially in dynamic medical image analysis.[4]

4.2. Fusion Convolutional Neural Network for Feature Pre-extraction

CNN can effectively extract the local features of the image, and RNN can process the continuous time series data well. The combination of the two can solve the temporal and spatial attributes of the medical image data. In the comprehensive construction, CNN extracts key static image features such as edges, shapes, and textures, providing a solid foundation for the subsequent RNN to model temporal dynamics. In this way, the model demonstrates strong performance in both static image recognition and temporal sequence analysis, and improves the recognition accuracy and efficiency. It is more suitable for the analysis of medical images, especially dynamic images. CNN rapidly identifies diagnostic features from images, while RNN tracks lesion evolution across image sequences based on temporal information. Take the tumor as an example, such as the detection of lung cancer, CNN extracts the tumor location in lung X-ray or CT images and other image data, and RNN tracks the change position of the tumor at various stages. This method can not only improve the accuracy of original disease diagnosis, but also improve the overall processing speed, especially when faced with massive medical image data, which significantly enhances the efficiency of clinical diagnosis.

4.3. Lightweight and Model Pruning Technologies Improve Deployment Efficiency

In the process of medical image recognition, RNN model often deals with massive image data and high-dimensional features, which makes it difficult to apply the RNN model with very heavy computation. Reducing model size and computational complexity is the most basic and effective RNN optimization method to improve model efficiency in moving edge devices or some memory-limited environments. Among them, lightweight techniques are crucial for reducing memory consumption and computational time and computing time by reducing the number of model parameters and computational complexity, and retaining a high enough accuracy. The commonly used lightweight methods include quantization, shared weights, low-rank decomposition, etc., which can greatly reduce the model size and speed up the reasoning process. In addition, another lightweight approach to model pruning is to reduce computational time by removing unnecessary neurons and their connections, thereby improving computational efficiency and enabling faster deployment. This compression method can not only reduce the computational amount of the model, but also be suitable for real-time medical image processing and device use. In the recognition of medical images, real-time and accurate localization are both essential, particularly for emergency diagnostics and large-scale screenings where speed is critical.[5]

4.4. Introduction of Multimodal Learning and Attention Mechanism

Medical images are usually composed of a variety of modal data, such as CT, MRI, X-ray and ultrasound, which not only vary in image quality, resolution and contrast, but also have their own characteristics in the role of each modal image in medical diagnosis. In order to effectively integrate these different types of medical image data, the combination of RNN and multimodal learning has become a key strategy to optimize medical image recognition. Multimodal learning helps models capture more comprehensive image features by simultaneously utilizing data from different modes. For example, in cancer diagnosis, CT images can provide three-dimensional structural information about a tumor, while MRI can provide soft tissue details of a tumor. By feeding these different image modes into the same multimodal learning framework, RNNS can fuse this information to provide more accurate diagnostic results. The introduction of attention mechanism further improves the performance of the model. The attention mechanism enables RNNS to automatically focus on the most critical areas of the image when processing images of different modes. By giving the model "attention" on key areas, it is able to identify focal areas more precisely and reduce the impact of distracting information when making disease diagnoses. The additive attention mechanism can be expressed as:

$$a_i = \frac{exp(score(F_{query}, F_{key}))}{\sum_j exp(score(F_{query}, F_{key}))} \tag{1}$$

among, F_{query} is the query feature (usually from the decoder or hidden state at the current moment), F_{key} is the key feature (usually from the encoder or input feature), a_i is the i th attention weight, score is a function used to compute the similarity between the query and the key, usually a dot product or additive function. By weighting input features, the model can focus on the most important parts of multiple inputs, improving learning.

5. Conclusion

With the continuous development of deep learning technology, the application potential of the RNN model in medical image recognition is becoming increasingly prominent. RNN variants incorporating gating mechanisms, hybrid models combining CNN and RNN, lightweight technologies, and methods based on multimodal learning and attention mechanisms have significantly optimized the accuracy and efficiency of RNNs in medical image recognition, effectively enhancing their adaptability to complex and dynamic medical image patterns. In particular, it has extremely favorable performance in disease detection, image report analysis, classification of small lesions, etc. It is anticipated that with the continuous improvement of computational power and algorithms, more RNN models will be applied to medical diagnosis to assist doctors in analyzing disease conditions and improving the efficiency and accuracy of diagnosis.

References

1. K. R. Singh and S. Dash, “Detection of arrhythmia from ECG signal using bat algorithm-based deep neural network,” in *Proc. Int. Conf. Adv. Comput. Intell. Eng.*, Singapore: Springer Nature Singapore, 2022. doi: 10.1007/978-981-99-5015-7_8.
2. X.-L. Wang, W.-X. Xie, and L.-Q. Li, “Structure identification of recursive TSK particle filtering via type-2 intuitionistic fuzzy decision,” *Int. J. Fuzzy Syst.*, vol. 23, no. 5, pp. 1294–1312, 2021, doi: 10.1007/s40815-020-01021-6.
3. T. Gao, L. Huang, S. Gao and K. Wang, "SSA: A Uniformly Recursive Bidirection-Sequence Systolic Sorter Array," *IEEE Trans. Parallel Distrib. Syst.*, vol. 35, no. 10, pp. 1721-1734, 2024, doi: 10.1109/TPDS.2024.3434332.
4. M. Cobbinah and A. Alnaggar, “An attention encoder-decoder RNN model with teacher forcing for predicting consumer price index,” *J. Data Inf. Manag.*, vol. 6, no. 1, pp. 65–83, 2024, doi: 10.1007/s42488-024-00114-3.
5. R. K. Behera et al., “Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data,” *Inf. Process. Manage.*, vol. 58, no. 1, p. 102435, 2021, doi: 10.1016/j.ipm.2020.102435.