IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 28 November 2023, accepted 18 December 2023, date of publication 25 December 2023, date of current version 5 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3347029

RESEARCH ARTICLE

Empowering Medical Data Analysis: An Advanced Deep Fusion Model for Sorting Medicine Document

GUAN BO[1,4], WANG SHANSHAN[1], ZHANG QING[1], PANG BO[2], AND ZUO YAN[3]
[1]Tianjin Hospital, Tianjin University, Tianjin 300211, China
[2]School of Medicine, Tianjin University, Tianjin 300072, China
[3]Legal Supervision Office, Tianjin Nankai Hospital, Tianjin 300100, China
[4]Graduate School of International Studies, Dong-A University, Busan 49236, South Korea

Corresponding author: Guan Bo (gbpxxrmn725898@yahoo.com)

This work was supported in part by the Key Program of Society for Degree and Postgraduate Education of China Research on the Path of Cultivating the Practical Ability of Postgraduate Medical Students in the Condocument of Smart Medicine under Grant 2020ZDB50, and in part by the Key Program of Institute of Educational Sciences of Academy of Management Sciences of China under Grant KJCX20782.

ABSTRACT To enhance the accuracy of medical document classification, we propose an advanced deep fusion model for sorting medicine document. Specifically, we enhance text representation using the bidirectional encoder representation from transformers (BERT). BERT is a bidirectional model that considers context information in input sequences. This capability is particularly valuable for medical document, as medical information often requires understanding in a global context, such as diagnoses, medical history, and treatment plans. Furthermore, BERT can learn the semantics of words and phrases, comprehending the different meanings of the same word in distinct contexts, which is crucial for representing medical document. For example, In the context of cardiology, *stroke*often refers to a cerebrovascular accident, which is a condition where blood flow to the brain is disrupted, leading to neurological impairment. This type of stroke is related to the brain and is a significant concern in the field of cardiology due to its impact on the circulatory system. In dermatology, *stroke*might be used to refer to a type of skin condition, such as *stroking the skin*. However, this context is less common and not related to the cerebrovascular meaning. Subsequently, we employ both Convolutional Neural Network (ConvNet) and Bidirectional Long Short Term Memory (Bi-LSTM) to extract local features and global long-term dependencies, respectively. Their outputs are then fused to extract useful document features at multiple levels, effectively capturing the documental structure. The proposed deep fusion model leverages the complementary strengths of these components, enhancing the model's generalization ability and mitigating the risk of over-fitting. Ultimately, by comparing our approach with state-of-the-art methods in medical document classification, we demonstrate the effectiveness of the proposed methodology.

INDEX TERMS Medicine data analysis, word embedding, deep learning, deep fusion model.

I. INTRODUCTION

As the information era rapidly unfolds, online medical consultation platforms are becoming increasingly popular [1]. Patients can conveniently and promptly consult doctors by describing their health conditions online, gaining information on medications, treatment plans, and more. This online platform enhances the accessibility and convenience of medical consultation, facilitating better dissemination of medical knowledge and health information. Additionally, online medical consultation can alleviate the burden on traditional healthcare institutions, especially at the grassroots level [2]. It helps redirect some patients, allowing hospitals to concentrate on those who require physical visits, thus improving the allocation efficiency of medical resources.

In recent years, artificial intelligence (AI) has emerged as a significant driving force behind a new wave of technological revolution and industrial transformation, greatly enhancing

The associate editor coordinating the review of this manuscript and approving it for publication was Yu-Da Lin.

the efficiency, quality, and innovation of intelligent healthcare [3]. First and foremost, AI has significantly impacted medical image diagnostics. Deep learning technologies, such as Convolutional Neural Networks (ConvNet) and Long Short-Term Memory (LSTM), enable computers to analyze medical images, identifying and locating disease markers like tumors and plaques [4]. This not only reduces the burden on doctors but also enhances diagnostic accuracy and speed, facilitating early detection and treatment. Furthermore, AI plays a crucial role in medical data analysis and personalized treatment [5]. By analyzing extensive healthcare data through big data analytics and machine learning, medical institutions can extract valuable information, providing support for clinical decision-making. In terms of personalized treatment, AI can design more accurate and efficient treatment plans for each patient based on factors such as genetic information and medical history, improving treatment outcomes. Additionally, AI has driven innovation in the healthcare sector, particularly in virtual health assistants, remote monitoring, and automation of surgeries [6]. Smart health assistants offer services like health counseling, medication reminders, and health data monitoring, helping patients better manage their health. Remote monitoring systems enable doctors to monitor patients' physiological data remotely, allowing timely interventions, which is especially crucial for managing chronic diseases and caring for high-risk patients. In terms of automated surgeries, AI's application, such as machine-assisted surgery and precise manipulation robots, enhances surgical precision and safety, opening new possibilities for medical procedures [7].

The significance of medical document classification lies in its ability to help the healthcare sector efficiently manage and utilize vast amounts of medical document information, thus promoting the development of medical research, health management, et al. [8]. The medical field encompasses extensive documental data such as medical records, research reports, and clinical trial data, all of which contain valuable medical knowledge. Through document classification, we can achieve the following objectives:

(1) Medical document classification aids in the rapid and accurate retrieval and organization of medical information [9]. Doctors and researchers can efficiently access literature specific to their fields, searching for disease cases, treatment plans, medical guidelines, and more, saving time and effort.

(2) Medical document classification supports clinical decision-making [10]. Doctors can better comprehend patients' conditions and formulate reasonable diagnoses and treatment plans based on the results of document classification, thereby improving the quality of healthcare.

(3) Medical document classification is of great significance for medical research [11]. Researchers can systematically categorize and analyze relevant studies, discovering new disease correlations, treatment methods, and advancing the accumulation of medical knowledge.

(4) Medical document classification contributes to health management and public health [12]. By analyzing extensive health data and epidemiological research literature, we can better understand the distribution and trends of diseases, formulate rational prevention strategies, and enhance public health levels.

In order to better analyze medical data, improve the generalization capability of medical document classification models, and mitigate the risk of over-fitting, we propose an advanced deep fusion model, i.e., BCB, for medical document classification. Specifically, the proposed model encodes the text vectors using the pre-trained BERT model, leveraging the advantages of ConvNet for extracting local features, as well as the high memory capacity of the BiLSTM (Bidirectional Long Short Term Memory) for capturing long-term dependencies. These extracted features are correctly concentrated, ensuring more effective and robust medical document classification. Experimental results demonstrate that the proposed model outperforms other state-of-the-art (SOTA) models. In summary, our contributions are as follows:

(1) To further enhance the accuracy of medical document classification, we propose a deep fusion method combining BERT, ConvNet, and BiLSTM. Collaborating BERT, ConvNet, and BiLSTM allows us to complement their respective strengths, facilitating feature learning at different levels and thereby improving the model's predictive ability for unseen samples.

(2) To provide richer semantic representations, we use the BERT model to train word vectors during the medical text vectorization process. BERT has shown excellent performance in natural language processing, capable of learning rich semantic representations of words. Using BERT as word vectors helps capture more accurate semantic information, contributing to enhanced expressiveness of medical document features.

(3) The model leverages the unique advantages of both ConvNet and BiLSTM. Specifically, ConvNet excels at extracting local features from medical document, while BiLSTM can understand long-term dependencies. Combining these two models allows the extraction of useful features from medical document at multiple levels, better capturing the structure and contextual information. Additionally, the proposed deep fusion model can automatically learn abstract features from the data, aiding in handling complex medical document data. The deep fusion model gradually improves performance, adapting to different medical document classification tasks as data volume increases and model optimization advances.

(4) Extensive comparisons are conducted with other advanced benchmark models, evaluating performance using metrics such as accuracy, precision, recall, and F1-score, validating the effectiveness of proposed method.

The rest of this paper is organized as follows: Section II covers word embedding techniques and related work on document classification in the medical domain. Section III provides a detailed introduction to the proposed method. Section IV conducts experiments to evaluate and compare the proposed method with several other classification models. Finally, the research conclusions are presented in Section V.

II. RELATED WORKS

Document classification is a crucial task in Natural Language Processing (NLP), aiming to categorize documents into one or more predefined classes based on their characteristics. It finds widespread applications in sentiment analysis, spam detection, topic classification, among others, serving as an effective means to assist in information retrieval, filtering, and utilization.

A. WORD EMBEDDING

The fundamental yet essential component of document classification is the transformation of text into numeric vectors, enabling computers to conduct subsequent calculations and processing. This process is known as text representation [13]. Word embedding is a common method for text representation, which embeds words from the text into a space and expresses them in the form of vectors. One-hot, Bag-Of-Words models, TF-IDF, etc., are frequently used text representation methods. However, the aforementioned feature representations disregard the context relationships within the text, treating each word as independent, which fails to capture semantic information. Moreover, these representations can lead to issues of sparsity and dimensionality disaster in the feature matrix. Consequently, the focus of subsequent research has been on constructing low-dimensional, distributed, dense word vectors. Word2Vec [14] is a neural network language model that considers both contextual semantics and avoids dimensionality issues, yielding superior performance compared to earlier models. Additionally, FastText [15], released by Facebook in 2016, is a tool for word vector computation and document classification. In classification tasks, FastText often achieves accuracy comparable to deep neural networks but with faster training times. However, both Word2Vec and FastText are static models since they have a one-to-one correspondence between words and vectors, making them incapable of dynamically adjusting and optimizing for specific tasks, thus unable to address the issue of polysemy.

In contrast, BERT is a pre-trained language model trained on a large-scale corpus, based on the multi-layer Transformer encoder architecture, utilizing attention mechanisms to directly encode word meanings, effectively addressing the polysemy issue based on contextual information [16]. BERT places greater emphasis on pre-training word meanings, allowing downstream NLP tasks to perform fine-tuning operations based on the specific task's requirements.

B. MEDICAL DOCUMENT CLASSIFICATION MODELS

The earliest document classification methods applied in the medical field included rule-based and machine learning-based methods. Rule-based methods require human involvement in determining the rules. Yaqoob et al. [17] proposed an effective disease classification method combining rule-based features and knowledge guidance. Although manually crafted rules may be more accurate, they become problematic when rules change or require updates, leading to the need for manual resummarization and rule-making. Thus, they are costly to maintain, have limited scalability, and handle structured and unstructured data. Among machine learning-based methods, rules can be established through data-driven methods. By using pre-labeled samples as training data, the intrinsic relationships between document segments and their labels can be learned. Di et al. [18] employed the chi-square method for feature word selection, and by using Bayesian networks and naive Bayesian classification methods on different quantities of feature words and classes, they enhanced the effectiveness of machine learning methods in remote healthcare.

In recent years, deep learning has been increasingly applied in document classification and has also been successfully applied to automated processing of medical document. Two representative deep models are ConvNet and Recurrent Neural Network (RNN), both of which have achieved advanced performance in many clinical data mining tasks.

Arbolí et al. [19] developed a neural network method for disease status classification of patients. The model outperformed SVM (Support Vector Machine), random forest, and decision tree models, successfully learning the high-dimensional EHR (Electronic Health Record) data structure of phenotype hierarchy. Jacquin et al. [20] compared ConvNets with traditional rule-based entity extraction systems using cTAKES and n-gram features. They tested ten different phenotype tasks on discharge summaries. ConvNet outperformed other methods on the prediction of ten phenotypes, and they concluded that deep learning-based methods in NLP improved patient phenotype performance compared to other methods. Dai et al. [21] applied ConvNet and RNN to classify semantic relationships among medical concepts in discharge summaries from the i2b2-VA challenge dataset. The results showed that ConvNet and RNN with only word embedding features achieved performance similar to advanced systems through feature engineering. Chen et al. [22] used TextCNN based on pre-trained word vectors to extract approximate information from short documents, and then employed a label-based text representation model to extract special meanings of vocabulary in the field of traditional Chinese medicine. They fused the outputs of both models into a linear network to classify traditional Chinese medical symptom documents, addressing the shortage of training data for specialized domain short document classification. Li et al. [34] focused on addressing the issue of insufficient feature extraction from Chinese medical text and low accuracy when classifying text information consulted by patients and recommending the correct department. They introduced a dual-channel Chinese medical text classification model that comprehensively extracts features at different granularity levels from Chinese medical text, aiming to obtain effective feature information and enhance department recommendations.

III. PROPOSED METHOD

We have designed a deep fusion model, i.e., BCB, for medical document classification based on BERT, ConvNet, and BiLSTM, as illustrated in Figure 1. The framework primarily consists of the BERT layer, one-dimensional convolutional layer, BiLSTM layer, feature fusion, and the Softmax layer. First, in the word embedding layer, we utilize BERT instead of Word2Vec to train word embeddings, enhancing the capability of word vector representations, resulting in semantically rich, low-dimensional word vectors. ConvNet possesses strong capabilities in extracting local features and allows parallel computation, contributing to higher training speeds. However, due to ConvNet's lack of contextual information acquisition, simple RNNs face issues of gradient explosion and vanishing gradients. To address this, we employ BiLSTM to capture long-term dependencies and acquire more accurate semantic information. Subsequently, the features obtained from ConvNet and BiLSTM are combined to form a comprehensive fusion feature input into the fully connected (FC) layer. Finally, the Softmax classifier outputs the classification results.

A. BERT

In BERT, there are three forms of embedding layers: token embeddings, segment embeddings, and position embeddings. The input document is first tokenized, and additional tokens [CLS] and [SEP] are added at the beginning and end of phrases. [CLS] represents classification, typically for downstream tasks. In scenarios like question answering and sentence matching, input sentences TextA and TextB are separated by [SEP] symbol. For document classification tasks, there's only one sentence input, so there's no TextB. Segment embeddings are used to differentiate between two similar vectors in a text. The first vector assigns the value 0 to each token in the first sentence, and the second vector assigns the value 1 to each token in the second sentence. If the input text contains only one sentence, its segment embedding is 0. The last layer is the position embedding, which describes the position of each word in the sentence, allowing the model to capture sequence information. These embeddings are linearly combined to generate the composite embedding representation [23].

B. CONVNET

The matrix E output from the word embedding layer is used as the input for convolution operations. In the convolution layer, many filters W with different window sizes (only varying in height h) slide over the entire row of E, i.e., the width of the filters is typically the same as the width of E. Each filter convolves with E, generating different feature maps C_i:

$$C_i = f(W \otimes E_{i:i+h-1}) + b \tag{1}$$

where b is the bias term used to adjust the output and weighted sum of the neuron input, and f is the ReLU nonlinear activation function, chosen because it reduces the number of iterations required for convergence in deep networks. Then, for elements within the same feature map, a max-pooling operation is applied to extract the most important features.

C. BILSTM

RNNs connect past outputs with current inputs, controlled by the *tanh* activation function, considering sequence states. The derivatives of RNNs at time t propagate to time t-1, t-2, ..., 1, leading to a multiplication factor. Continuous multiplication introduces two issues: exploding and vanishing gradients [24]. Thus, in the forward process, the impact of the input

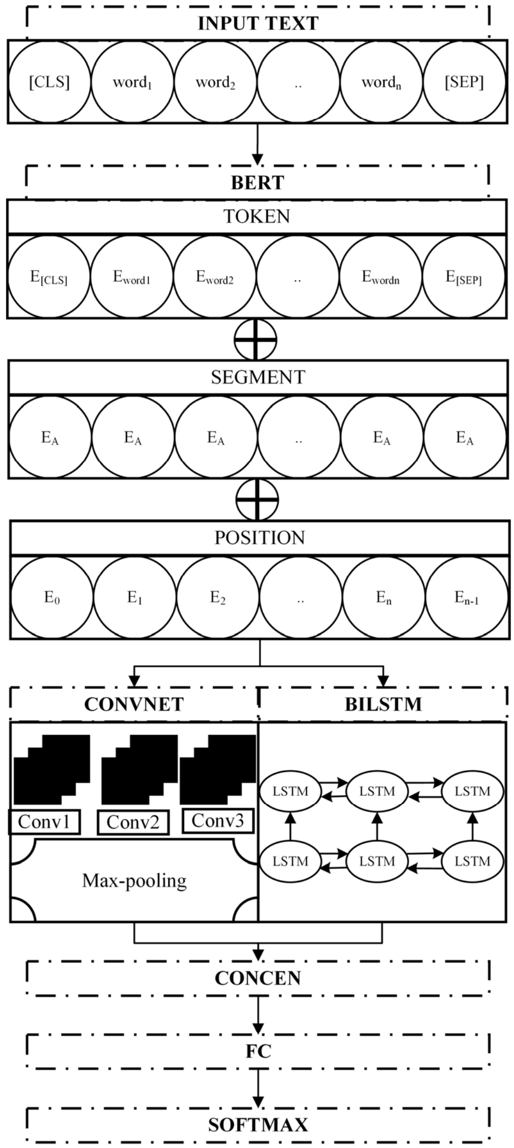

FIGURE 1. Overall architecture.

from the initial sequence on subsequent sequences diminishes, leading to the problem of long-range dependencies. LSTM [25] resolves these problems by introducing multiple gates, effectively capturing contextual information from the input document.

LSTM gates selectively memorize input, retaining essential information while forgetting minor details. It determines what new information to store in the current state. The previous state output h_{t-1} and current input information x_t are input to the sigmoid function, generating a value between 0 and 1, determining how much new information to retain. Through the forget gate and input gate, the complete state c_t for the next time step is obtained, used to generate the hidden layer h_t for the next state, i.e., the output of the current unit. The output gate determines what information to output from the cell state. Similar to the input gate, a sigmoid function generates a value o_t between 0 and 1, determining how much cell state information to output. When the cell state information is multiplied by o_t, it is first activated through the *tanh*layer, obtaining the output information of this LSTM structure h_t.

When using LSTM to extract semantic information, not only the preceding information of words needs to be considered but also the subsequent information. BiLSTM [26] captures information from both directions in the sentence and merges it, making it better suited for this requirement. Consequently, it can better capture bidirectional semantic dependencies. Assuming that at time step t, forward LSTM outputs hidden state $\vec{h}_t$, and backward LSTM outputs hidden state $\overleftarrow{h}_t$, then the output $h_t = [\vec{h}_t, \overleftarrow{h}_t]$.

IV. EXPERIMENT

In this section, the performance of the proposed document classification method, i.e., BCB, is validated using publicly available datasets, and comparisons are made with several SOTA methods. The following subsections describe the experimental datasets, experimental settings, ablation results, and contrast results.

A. DATASETS

To evaluate the effectiveness of the proposed BCB model, three medical document datasets were used in the experiments. The datasets are briefly described below:

(1) MDD dataset: This dataset is sourced from a public dataset on GitHub named *Chinese medical dialogue data* (https://github.com/Toyhom/Chinese-medical-dialogue-data). It comprises over 790,000 medical consultation records from an online platform, including six major departments: urology (90,000 records), internal medicine (220,000 records), obstetrics and gynecology (180,000 records), oncology (70,000 records), pediatrics (100,000 records), and surgery (110,000 records).

(2) MHQ dataset [32]: This dataset consists of 5,000 medical health questions provided by the *Chinese Medical Association*. The questions are categorized into seven major groups: diagnosis, treatment, anatomy/physiology, epidemiology, healthy lifestyle, doctor selection, and others.

(3) KUAKE-QIC dataset [33]: This dataset is from the *Chinese Health Information Processing Conference* and contains over 6,900 samples for medical search query intent classification. The medical questions in this evaluation are divided into 11 types: diagnosis, disease analysis, causative analysis, treatment plan, medical advice, index interpretation, disease description, consequences description, precautions, efficacy description, and medical costs.

All three datasets were divided into training, testing, and validation sets in a 6:2:2 ratio.

B. EXPERIMENTAL SETUP

All experiments were conducted on a computer with an AMD Ryzen 5 4600H CPU, 16.0 GB RAM, and an NVIDIA GEFORCE RTX 1650 graphics card. The experiments were carried out using Python 3.8 and the PyTorch 1.10.1+cuda12.2 deep learning framework.

The pre-trained BERT-Base-Chinese feature vectors were used. The model was configured with a 12-layer bidirectional Transformer encoder, 768 hidden units, 12 multi-head attention mechanisms, and 110 million parameters.

For the hyper-parameters of the proposed deep fusion neural network layer, 256 convolution filters with window sizes of 2, 3, and 4 were used in the ConvNet layer to extract various features. The size of hidden layer in the BiLSTM was set to 128. The FC layer used the ReLU nonlinear activation function, and a dropout rate of 0.5 was set to prevent over-fitting. L2 regularization was applied to the weight parameters. The Adam optimizer was chosen as the gradient descent optimization algorithm, and cross-entropy was used as the loss function. The initial learning rate was set to 0.001, and the number of iterations was set to 2,000, with a batch size of 128.

C. EVALUATION METRICS

The evaluation metrics for the classification models primarily include accuracy, precision, recall, and F1-score based True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) [27]. They are defined as follows:

TP: It refers to the model correctly predicting a positive example as positive. In other words, samples that are actually positive are correctly predicted as positive.

TN: It refers to the model correctly predicting a negative example as negative. In other words, samples that are actually negative are correctly predicted as negative.

FP: It refers to the model incorrectly predicting a negative example as positive. In other words, samples that are actually negative are wrongly predicted as positive.

FN: It refers to the model incorrectly predicting a positive example as negative. In other words, samples that are actually positive are wrongly predicted as negative.

Accuracy: The proportion of correctly classified samples among the total samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Precision: The ratio of TP predictions to the sum of true positive and false positive predictions.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Recall: The ratio of TP predictions to the sum of TP and FN predictions.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

F1-score: The weighted harmonic mean of precision and recall.

$$\begin{aligned} F1 - Score &= \frac{TP}{TP + \frac{1}{2(FP+FN)}} \\ &= \frac{2 \times \Pr ecision \times Recall}{\Pr ecision + Recall} \end{aligned} \tag{5}$$

D. ABLATION EXPERIMENT

We initially compare our method with BERT+ ConvNet and BERT+BiLSTM. The variations in training accuracy and training loss throughout the iterative process are depicted in Figures 2 and 3, respectively. It is evident that proposed deep fusion model, BCB, demonstrates superior performance. However, we have observed an advantage of BERT +BiLSTM over BERT+ConvNet, possibly attributable to two aspects: sequence modeling and semantic representation.

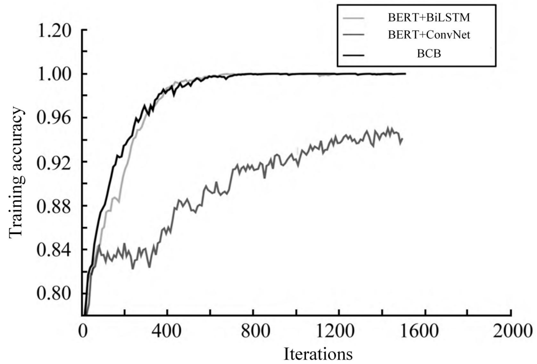

FIGURE 2. The change trend of training accuracy of the global, local and proposed method.

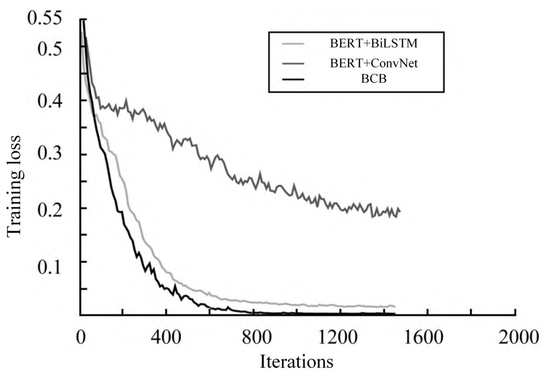

FIGURE 3. The change trend of training loss of the global, local and proposed method.

Firstly, the stronger sequence modeling capacity of BiLSTM in sequence data is advantageous. It captures long-term dependencies within input sequences, exhibiting enhanced aptitude for handling temporal and contextual associations in text data. This is particularly beneficial for tasks in NLP such as document classification, sentiment analysis, and named entity recognition that require careful consideration of textual context.

Secondly, BiLSTM boasts richer semantic representation capabilities. By virtue of bidirectional information flow, BiLSTM can holistically consider a word's contextual information both before and after it, thereby more accurately expressing the word's semantics. This ability proves crucial in constructing comprehensive word embeddings and effectively modeling semantic nuances within sentences.

Thus, it appears that within medical document classification, global contextual information holds greater significance compared to local information. Consequently, we redirect our focus towards examining the ablation outcomes of various long-term relation extractors. To identify the optimal long-term relation extractor, we employ ablation experiments that contrast the proposed method with BERT+ConvNet+BiRNN (Bidirectional Recurrent Neural Network) as well as BERT+ConvNet+BiGRU (Bidirectional Gated Recurrent Unit).

Further elaborating, we provide separate introductions for BiRNN and BiGRU.

(1) BiRNN [28]: a BiRNN is a type of RNN that processes input sequences in both forward and backward directions. It has the ability to capture context information from both past and future elements in the input sequence. This bidirectional nature is particularly useful in tasks where understanding the entire sequence context is crucial, such as NLP and speech recognition.

In a BiRNN, the input sequence is processed by two separate RNNs: one in the forward direction and another in the backward direction. The outputs from both directions are usually combined, allowing the model to consider both preceding and succeeding elements of the input when making predictions. This enhances the model's ability to capture long-term dependencies and context, which can lead to improved performance on tasks that require understanding the sequence as a whole.

(2) BiGRU [29]: a BiGRU is a specific variant of the BiRNN that employs Gated Recurrent Units (GRUs) as the building blocks for processing input sequences [30]. GRUs are a type of gated recurrent unit, which is a variation of the traditional RNN cell that addresses some of the issues related to vanishing gradients and capturing long-term dependencies.

The BiGRU, similar to the BiRNN, processes input sequences in both forward and backward directions. Each GRU unit in the BiGRU has gating mechanisms that allow it to control the flow of information, making it more effective in capturing relevant context and handling sequential data. The bidirectional aspect ensures that the model can leverage information from both past and future elements in the sequence, leading to better representation learning and improved performance on tasks that involve sequential data analysis, such as language modeling, sequence classification, and time series forecasting.

In Figure 4, we illustrate the trends in training accuracy during each iteration for the three methods, revealing minimal significant differences. Figure 5 portrays the trends in training loss during iterations, indicating the distinct advantage of BiLSTM over BiRNN and BiGRU. The reasons for this disparity are as follows:

(1) BiLSTM excels in capturing long-term dependencies due to its stronger memory capacity, proving adept at grasping extensive relationships within documental data. This is of utmost importance when analyzing complex contextual information, semantic correlations, and logical reasoning in document.

(2) BiLSTM effectively mitigates the vanishing gradient problem through mechanisms such as forget gates, input

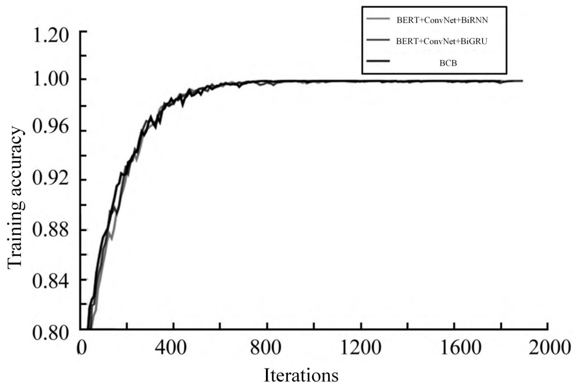

FIGURE 4. The change trend of training accuracy of global method and proposed method.

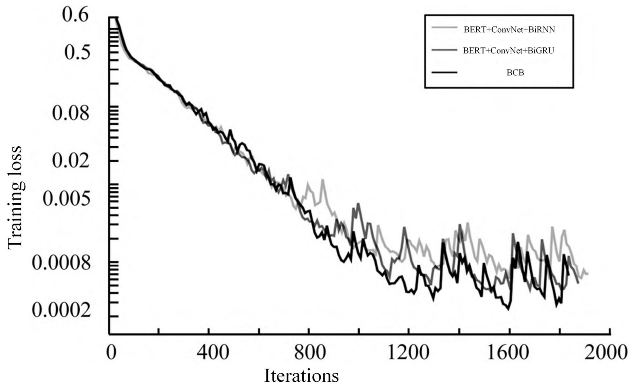

FIGURE 5. The change trend of training loss of global method and proposed method.

gates, and output gates, rendering the network more amenable to training, particularly with long sequence data.

(3) BiLSTM's performance balance remains relatively stable across various sequence modeling tasks. While BiGRU may occasionally converge more rapidly, BiLSTM generally outperforms in terms of generalization and adaptation to complex tasks. The intricate gate mechanism of BiLSTM, in comparison to BiGRU, allows for finer control over information flow and retention.

E. CONTRAST EXPERIMENT

To validate the effectiveness of the proposed BCB model, extensive comparative experiments were conducted against SOTA methods. The benchmark methods include:

TextCNN [22]: A ConvNet-based text classification method with a simple architecture and swift training speed. It leverages multiple differently sized convolutional kernels to extract crucial information from sentences, capturing their local correlations.

TextRNN [31]: An extensively employed NLP model that deals with variable-length textual sequences, effectively learning long-term dependencies within sentences.

Before conducting experimental comparisons with benchmark methods, it's crucial to first acknowledge the theoretical key differences:

(1) Text Representation Capability: Our approach utilizes BERT for text representation, which is a bidirectional model capable of a deeper understanding of context in the text. In contrast, TextCNN uses convolutional kernels for capturing local information, while TextRNN is employed to learn long-term dependencies. BERT offers a more comprehensive grasp of the global context in text, making it particularly suitable for medical documents where global information often holds significant importance.

(2) Multilevel Feature Extraction: Our method integrates ConvNet and Bi-LSTM to extract local features and global long-term dependencies. This means the system comprehends the text at multiple hierarchical levels. In contrast, TextCNN is mainly focused on extracting local features, and TextRNN primarily deals with long-term dependencies. This multilevel feature extraction enhances the comprehensiveness of document understanding.

(3) Feature Fusion: Our approach adopts a multilevel feature fusion strategy by combining the outputs of BERT, ConvNet, and Bi-LSTM. This contributes to a more comprehensive understanding of the document and improves the accuracy of text classification. In comparison, TextCNN and TextRNN do not incorporate BERT or multilevel fusion.

(4) Overfitting Mitigation: Our deep fusion model is designed to integrate the strengths of various components, enhancing the model's generalization ability and reducing the risk of overfitting. This is because different components can complement each other's deficiencies, providing a more robust performance. In contrast, TextCNN and TextRNN do not introduce a similar deep fusion strategy.

Therefore, given the influence of distinct word embeddings on classification outcomes, we executed experiments with different word embeddings, namely Word2Vec and BERT, forming four comparative sets: Word2Vec+TextCNN, Word2Vec+TextRNN, BERT+TextCNN, and BERT+TextRNN. In these experiments, Word2Vec embedding was configured with a vector dimension of 300, while BERT embedding was set at 768. The results are presented in Tables 1, 2, and 3, revealing the following insights:

(1) BERT integration as the word embedding significantly enhances accuracy compared to Word2Vec, with BERT+TextRNN exhibiting a 3.72% accuracy improvement over Word2Vec+TextRNN. This stems from BERT's ability to incorporate positional information into the embedding, utilizing deep bidirectional Transformer training for word vectors. In contrast to Word2Vec, BERT excels at extracting contextual, semantic, and grammatical features, resulting in stronger semantic representation and generalization capabilities. Integrating BERT into the framework substantially enhances classification performance.

(2) The proposed deep fusion neural network model, BCB, surpasses single BiLSTM and ConvNet model variants. Compared to BERT+TextCNN and BERT+TextRNN, the BCB model demonstrates improvements of 1.69% and 3.62%, respectively. This is attributed to BCB capitalizing on the strengths of both ConvNet and BiLSTM networks, effectively

combining local and contextual pathological information features. Additionally, although Table 1 suggests TextRNN superiority over TextCNN, in Tables 2 and 3, TextRNN lags behind TextCNN. This can be attributed to the shorter sequence lengths of the MHQ dataset and KUAKE-QIC dataset relative to the MDD dataset. This underscores the limitation of RNN in extracting remote features when dealing with short text sequences, favoring ConvNet's ability to rapidly converge and extract key local features, rendering it more suitable for short text classification.

(3) Notably, the classification outcomes of all models significantly outperform on the MDD dataset compared to the other two datasets. This can be attributed to the MDD dataset's larger sample size, demonstrating that models benefit from greater training data volume.

In summary, gauging from the values of various evaluation metrics, it is evident that BCB exhibits the best performance in medical document classification tasks, yielding accuracy improvements ranging from 1.8% to 4.15%, 1.69% to 5.15%, and 1.74% to 5.91% across the three datasets. This can be attributed to BERT's effective extraction and learning of word contextual information, the robust local learning capabilities of ConvNet networks, the ability to extract richer information through various filter sizes, and BiLSTM networks' enhancement of semantic information through sentence-level feature extraction. As such, the model synergizes the advantages of BERT, ConvNet, and BiLSTM, resulting in higher classification accuracy than other models.

TABLE 1. Contrast results in MDD dataset.

Methods	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Word2Vec+TextCNN	80.70	77.22	78.58	79.79
Word2Vec+TextRNN	81.64	79.29	80.11	81.30
BERT+TextCNN	81.47	78.63	79.74	80.85
BERT+TextRNN	82.04	80.53	81.17	82.14
BCB	84.45	82.41	83.26	83.94

TABLE 2. Contrast results in MHQ dataset.

Methods	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Word2Vec+TextCNN	69.94	67.24	68.32	71.73
Word2Vec+TextRNN	70.20	65.54	67.12	71.03
BERT+TextCNN	73.27	74.05	73.44	74.49
BERT+TextRNN	71.00	68.35	69.33	72.56
BCB	76.50	72.97	74.58	76.18

F. FURTHER ANALYSIS

In order to provide a more detailed assessment of the proposed method, Figure 6 presents the specific classification accuracy for different diseases and medical departments. It's worth noting that to facilitate easy visualization of the results for each type, the horizontal axis only displays 23 diseases or departments. Based on the evaluation metrics, we can observe that Cardiovascular Surgery has the highest accuracy, reaching up to 95%, Infertility, Diabetes, Gynecological Oncology, et al. attain accuracy levels of 90%.

TABLE 3. Contrast results in kuake-QIC dataset.

Methods	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Word2Vec+TextCNN	77.76	73.42	74.29	73.38
Word2Vec+TextRNN	73.47	70.83	72.07	71.24
BERT+TextCNN	76.85	75.41	75.90	75.41
BERT+TextRNN	75.32	69.56	71.01	74.96
BCB	78.60	71.69	74.07	77.15

In addition, good classification results are also achieved for diseases such as AIDS, Pediatric Medicine, and Breast Reconstruction, et al. In contrast, the performance of the compared SOTA methods, TextRNN & TextCNN, in these categories is not quite satisfactory, demonstrating results that are relatively difficult to recognize. However, our accuracy for Ophthalmology, Family Planning, et al. is not very satisfactory.

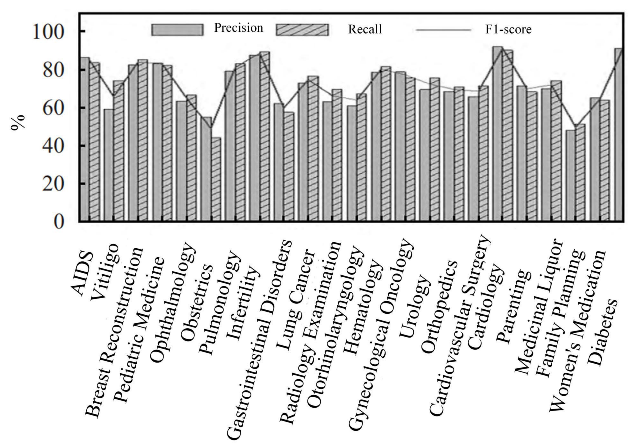

FIGURE 6. The change trend of training loss of global method and proposed method.

Analyzing the discriminative outcomes for disease types based on medical knowledge and documentual descriptions within the dataset, it becomes evident that certain diseases, like AIDS, have more distinct disease characteristics, leading to better identification rates.

V. DISCUSSION

Our work addresses several specific gaps in the existing literature on medical document classification and offers substantial benefits for future research and the field in general. One of the primary gaps our work fills is the need for higher accuracy in classifying complex medical documents, especially in fine-grained disease categories. Existing models, such as TextRNN and TextCNN, although state-of-the-art, have shown limitations in accurately identifying certain medical conditions. Our model, with its advanced deep fusion technique integrating BERT, ConvNet, and BiLSTM, demonstrates a significant improvement in this regard, with accuracies exceeding 90% in challenging categories like cardiovascular surgery, infertility, and diabetes. This advancement is a substantial step forward in the precision of medical document classification. In addition, our model effectively processes a wider range of medical data, including diverse datasets.

In terms of providing insights for future work, our study offers several avenues for expansion and exploration:

1) Advanced Semantic Representations: The use of BERT in our model highlights the importance of deep semantic understanding in medical text classification. Future research can explore the integration of even more advanced language models that are continually emerging in the field of NLP. This can include investigating how newer models like domain-specific language models can further enhance semantic representation and accuracy in medical contexts.
2) Comprehensive Feature Learning: Our work demonstrates the effectiveness of learning both local features and long-term dependencies. Future models could expand on this by exploring additional neural network architectures that might offer improvements in capturing these aspects, such as Transformer-based models or newer forms of recurrent neural networks.
3) Cross-Domain Applications: The fusion technique we developed has potential applications beyond medical text classification. Future research could explore its applicability in other areas of NLP, such as sentiment analysis, text summarization, or even non-NLP fields.
4) Interpretability and Explainability: While our model improves performance, there's an ongoing challenge in the field regarding the interpretability of complex models. Future research could focus on making these models more transparent and understandable, which is particularly important in the medical field for gaining trust and actionable insights.
5) Integration with Healthcare Systems: A practical aspect for future research is the integration of advanced NLP models into existing healthcare systems. This includes not only technical integration but also considerations of usability, accessibility, and compliance with healthcare regulations.

VI. CONCLUSION

We propose an advanced deep fusion method for sorting medicine document named BCB, which is based on BERT for text representation, ConvNet for extracting local features, and Bi-LSTM for capturing long-term dependencies. BERT has demonstrated remarkable performance in the different NLP tasks, enabling the learning of rich semantic representations of words. Using BERT as the word embedding can capture more accurate semantic information, enhancing the expressive power of text features. ConvNet and Bi-LSTM, as deep learning models, can automatically learn abstract features from the data, aiding in handling complex medical text data. Deep learning models gradually improve performance, adapting to different medical text classification tasks as the dataset size increases and models are optimized. The experimental results demonstrate the effectiveness of the proposed BCB method in medical document classification, achieving the highest classification accuracy compared to other benchmark methods.

REFERENCES

[1] H. Tan, X. Zhang, and Y. Yang, "Satisfaction or gratitude? Exploring the disparate effects of physicians' knowledge sharing on patients' service evaluation in online medical consultations," *Inf. Syst. J.*, vol. 33, no. 5, pp. 1186–1211, Sep. 2023.
[2] D. Ygiyeva, L. Pivina, A. Messova, Z. Urazalina, Y. Semenova, A. Dyussupov, A. Dyussupov, T. Belikhina, M. Syzdykbayev, G. Batenova, A. Akhmetova, and A. Adams, "Paramedic to provider consultation challenges in the emergency department in Kazakhstan: An online survey," *Open Access Emergency Med.*, vol. 15, pp. 69–77, Mar. 2023.
[3] V. A. Dang, Q. Vu Khanh, V.-H. Nguyen, T. Nguyen, and D. C. Nguyen, "Intelligent healthcare: Integration of emerging technologies and Internet of Things for humanity," *Sensors*, vol. 23, no. 9, p. 4200, Apr. 2023.
[4] J. Zhou, H. Li, X. Liao, B. Zhang, W. He, Z. Li, L. Zhou, and X. Gao, "Audit to forget: A unified method to revoke patients' private data in intelligent healthcare," *BioRxiv*, vol. 2023, Jan. 2023, Art. no. 529040.
[5] A. Ahad and M. Tahir, "Perspective—6G and IoT for intelligent healthcare: Challenges and future research directions," *ECS Sensors Plus*, vol. 2, no. 1, Mar. 2023, Art. no. 011601.
[6] T. Jabeen, I. Jabeen, H. Ashraf, N. Z. Jhanjhi, A. Yassine, and M. S. Hossain, "An intelligent healthcare system using IoT in wireless sensor network," *Sensors*, vol. 23, no. 11, p. 5055, May 2023.
[7] K. Guo, F. Zhu, X. Zhou, L. Zhang, Y. Wang, and J. Kang, "LesionTalk: Core data extraction and multi-class lesion detection in IoT-based intelligent healthcare," *ACM Trans. Sensor Netw.*, vol. 19, no. 3, pp. 1–20, Aug. 2023.
[8] M. Y. Landolsi, L. Hlaoua, and L. Ben Romdhane, "Information extraction from electronic medical documents: State of the art and future research directions," *Knowl. Inf. Syst.*, vol. 65, no. 2, pp. 463–516, Feb. 2023.
[9] A. Thielmann, C. Weisser, A. Krenz, and B. Säfken, "Unsupervised document classification integrating web scraping, one-class SVM and LDA topic modelling," *J. Appl. Statist.*, vol. 50, no. 3, pp. 574–591, Feb. 2023.
[10] J. Li, Q. Huang, S. Ren, L. Jiang, B. Deng, and Y. Qin, "A novel medical text classification model with Kalman filter for clinical decision making," *Biomed. Signal Process. Control*, vol. 82, Apr. 2023, Art. no. 104503.
[11] G. McLachlan, A. Broomfield, and R. Elliott, "Completeness and accuracy of adverse drug reaction documentation in electronic medical records at a tertiary care hospital in Australia," *Health Inf. Manage. J.*, vol. 52, no. 2, pp. 108–111, May 2023.
[12] M. Botlagunta, M. D. Botlagunta, M. B. Myneni, D. Lakshmi, A. Nayyar, J. S. Gullapalli, and M. A. Shah, "Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms," *Sci. Rep.*, vol. 13, no. 1, Jan. 2023.
[13] F. Masood, J. Masood, H. Zahir, K. Driss, N. Mehmood, and H. Farooq, "Novel approach to evaluate classification algorithms and feature selection filter algorithms using medical data," *J. Comput. Cognit. Eng.*, vol. 2022, pp. 1–13, May 2022.
[14] J.-J. Zhu and Z. J. Ren, "The evolution of research in resources, conservation & recycling revealed by Word2Vec-enhanced data mining," *Resour., Conservation Recycling*, vol. 190, Mar. 2023, Art. no. 106876.
[15] S. Chatterjee, L. Evenss, P. Bhattacharyya, and J. Mondal, "LSJSP at SemEval-2023 task 2: FTBC: A FastText based framework with pretrained BERT for NER," in *Proc. The 17th Int. Workshop Semantic Eval.*, 2023, pp. 1254–1259.
[16] Z. Xinxi, "Single task fine-tune BERT for text classification," in *Proc. 2nd Int. Conf. Comput. Vis., Image, Deep Learn.*, Kunming, China, Oct. 2021, pp. 194–206.
[17] A. Yaqoob, R. M. Aziz, N. K. Verma, P. Lalwani, A. Makrariya, and P. Kumar, "A review on nature-inspired algorithms for cancer disease prediction and classification," *Mathematics*, vol. 11, no. 5, p. 1081, Feb. 2023.
[18] F. Di Martino and F. Delmastro, "Explainable AI for clinical and remote health applications: A survey on tabular and time series data," *Artif. Intell. Rev.*, vol. 56, no. 6, pp. 5261–5315, Jun. 2023.
[19] E. Rodríguez-Arbolí, M. Othus, C. Orvain, L. C. Zarling, B. M. Sandmaier, F. Milano, G. Schoch, C. Davis, H. J. Deeg, F. R. Appelbaum, R. Storb, and R. B. Walter, "Contribution of measurable residual disease status to prediction accuracy of relapse and survival in adults with acute myeloid leukemia undergoing allogeneic hematopoietic cell transplantation," *Haematologica*, vol. 108, no. 1, pp. 273–277, Sep. 2022.

[20] C. Jacquin et al., "1p36 deletion syndrome: Review and mapping with further characterization of the phenotype, a new cohort of 86 patients," *Amer. J. Med. Genet., A*, vol. 191, no. 2, pp. 445–458, Feb. 2023.
[21] H. Dai, L. Zhang, L. Zhao, Z. Wu, Z. Liu, D. Liu, X. Yu, Y. Lyu, C. Li, N. Liu, T. Liu, and D. Zhu, "Hierarchical semantic tree concept whitening for interpretable image classification," 2023, *arXiv:2307.04343*.
[22] M. Chen, C. Yao, X. Li, and L. Shen, "A text classification model based on Gaussian multi-head self attention mechanism for Chinese medical data," in *Proc. IEEE 2nd Int. Conf. Electr. Eng., Big Data Algorithms (EEBDA)*, Feb. 2023, pp. 229–232.
[23] Z. Wang, L. Wang, C. Huang, S. Sun, and X. Luo, "BERT-based Chinese text classification for emergency management with a novel loss function," *Int. J. Speech Technol.*, vol. 53, no. 9, pp. 10417–10428, May 2023.
[24] M. Imani, "Alzheimer's diseases diagnosis using fusion of high informative BiLSTM and CNN features of EEG signal," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105298.
[25] M. H. Park, D. Nan, Y. Kim, and J. H. Kim, "CBOE volatility index forecasting under COVID-19: An integrated BiLSTM-ARIMA-GARCH model," *Comput. Syst. Sci. Eng.*, vol. 47, no. 1, pp. 121–134, 2023.
[26] E. L. Abdellah, N. Benaya, T. Bakir, A. Mansouri, and A. E. L. Allat, "Prediction of U.S. 30-years-treasury-bonds mouvement and trading entry point using Robust 1DCNN-BiLSTM-XGBoost algorithm," Tech. Rep., 2023.
[27] J. Shin, J. Wang, S. Wang, and N. Nagappan, "Automatic static bug detection for machine learning libraries: Are we there yet?" 2023, *arXiv:2307.04080*.
[28] M. Rezaei, A. Mohammadifar, H. Gholami, M. Mina, M. J. P. M. Riksen, and C. Ritsema, "Mapping of the wind erodible fraction of soil by bidirectional gated recurrent unit (BiGRU) and bidirectional recurrent neural network (BiRNN) deep learning models," *Catena*, vol. 223, Mar. 2023, Art. no. 106953.
[29] M. Yu, D. Niu, T. Gao, K. Wang, L. Sun, M. Li, and X. Xu, "A novel framework for ultra-short-term interval wind power prediction based on RF-WOA-VMD and BiGRU optimized by the attention mechanism," *Energy*, vol. 269, Apr. 2023, Art. no. 126738.
[30] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2017, pp. 1597–1600.
[31] Q. Ye, R. Yang, C.-L. Cheng, L. Peng, and Y. Lan, "Combining the external medical knowledge graph embedding to improve the performance of syndrome differentiation model," *Evidence-Based Complementary Alternative Med.*, vol. 2023, pp. 1–11, Feb. 2023.
[32] Y. Li, Z. Li, K. Zhang, R. Dan, S. Jiang, and Y. Zhang, "ChatDoctor: A medical chat model fine-tuned on a large language model meta-AI (LLaMA) using medical domain knowledge," *Cureus*, vol. 15, no. 6, Jun. 2023.
[33] N. Zhang et al., "CBLUE: A Chinese biomedical language understanding evaluation benchmark," 2021, *arXiv:2106.08087*.
[34] X. Li, Y. Zhang, J. Jin, F. Sun, N. Li, and S. Liang, "A model of integrating convolution and BiGRU dual-channel mechanism for Chinese medical text classifications," *PLoS ONE*, vol. 18, no. 3, Mar. 2023, Art. no. e0282824.

GUAN BO was born in Qiqihar, Heilongjiang, China, in 1987. He received the degree in clinical medicine from the Chengde Medical College, and the master's degree from Tianjin University, China. He is currently pursuing the Ph.D. degree with Dong-A University, South Korea. He is also a Supervisor of the master's degree in hospital management with the Tianjin University of Traditional Chinese Medicine; the Co-Advisor of MPA with the School of Management, Tianjin University; and an Associate Professor and a Psychological Counselor of the Tianjin Hospital. His main research interests include social medicine, health care management, and hospital information management. He is the Deputy Director of the Tianjin Health Education Association Psychological Branch and a member of the Tianjin Association of Integrative Medicine Hospital Management Committee.

WANG SHANSHAN was born in Tangshan, Hebei, China, in 1979. She received the bachelor's degree in clinical medicine from Hebei Medical University and the master's degree in health care management from the North China Coal Medical College, China. She has engaged in medical quality management for 16 years. She is an Associate Researcher with the Tianjin Hospital. Her main research interests include medical quality management, medical safety, and hospital information management.

ZHANG QING was born in Tianjin, China, in 1982. She received the bachelor's degree in medicine from Tianjin Medical University, China. She has engaged in clinical medical work and hospital management for 16 years. She is currently an Assistant Researcher with the Tianjin Hospital. Her main research interest includes health care management.

PANG BO was born in Tianjin, China, in 1983. He received the master's degree in engineering from Tianjin University, China. He is currently the Deputy Director of the School of Medicine, Tianjin University. He is mainly engaged in medical professional development research and medical education research.

ZUO YAN was born in Tianjin, China, in 1982. She received the master's degree in education management from Tianjin Vocational and Technical Normal University, China. She is currently the Director of the Office of Legal Supervision and the Office of Hospital Quality Management and an Assistant Researcher with the Tianjin Nankai Hospital. She is also a member of the Tianjin Hospital Management Quality Control Center. She is mainly engaged in medical jurisprudence, legal practice supervision, hospital quality management, and management teaching and research. Her main research interests include medical jurisprudence and hospital management.