DL-PER: Deep Learning Model for Chinese Prehospital Emergency Record Classification

Xu Zhang¹, Haifang Zhang², Feng Tian³, Long Sheng¹,⁴*
¹School of Information and Electrical Engineering, Hebei University of Engineering, Handan Hebei 056038, China
²Handan Emergency Center, Handan Hebei 056002, China
³School of Medicine, Hebei University of Engineering, Handan Hebei 056038, China
⁴Hebei Key Laboratory of Security Protection Information Sensing and Processing, Hebei University of Engineering, Handan Hebei 056038, China

Corresponding author: Long Sheng (e-mail: shenglong@hebeu.edu.cn).

This work was supported in part by the Innovation Ability Promotion Project of Hebei Province under Grant 215576135D, and the Hebei Education Department under Grant SQ201023.

ABSTRACT

Prehospital emergency records contain much information about prehospital emergency patients. Extracting important patient information from many records has become the focus of all prehospital emergency personnel. The key to solving this problem is to achieve the automatic classification of prehospital emergency records. This study considers a deep learning-based prehospital emergency record classification model (DL-PER). The model uses a weighted text convolutional neural network to classify prehospital emergency records. First, we use prehospital emergency records to train a bi-directional encoder representation (BERT) model from the transformer. BERT obtains the word vectors. Then, we use a bi-directional long and short-term memory (BiLSTM) model to obtain text features from a global perspective. A weighted text convolutional neural network (WTextCNN) improves this model's local text feature extraction capability. We used activation functions instead of ReLu activation functions to improve the learning ability of the model. We conducted experiments using prehospital emergency records provided by the Handan Emergency Center. The results showed that the DL-PER model improved the F1 scores by up to 5.7%, 6.8%, 5.7%, and 4.9% on the four data sets, respectively, compared with the BiLSTM model.

INDEX TERMS
deep learning, text classification, prehospital emergency, weighted text convolutional neural network, bidirectional long short-term memory,

I. INTRODUCTION

The emergency medical service system is an essential part of the health service system in China, serving the task of rescuing patients with serious diseases [1]. Prehospital emergency plays a vital role in the emergency medical service system. Most patients treated in prehospital emergencies are critically ill with short survival times. For example, stroke [2] and cardiac arrest patients [3] have short resuscitation times. According to previous studies, only 21.5% of ischemic stroke patients in China reach the emergency room within three hours [4]. An emergency network in large and medium-sized cities has been established, whose professional level is higher and higher [5]. The prehospital emergency record contains all the information about the patient before arriving at the hospital. This information is important to improve the rate of emergency patient care. Finding useful information quickly and efficiently in the massive amount of data to improve the efficiency of emergency care is the challenge faced by prehospital emergency personnel. The key to solving this problem is implementing automatic classification of prehospital emergency records.

Text classification techniques can obtain the text features of each label from a large amount of text to classify the dataset. Researchers have used deep learning to achieve text classification in different medical fields, such as clinical records [6], electronic medical records [7], Chinese medical questions [8], and nursing texts [9], but research on prehospital emergency records is less common.

The prehospital emergency record is a piece of brief information about the patient's onset, the patient's identity information, past medical history, and other related contents recorded by the operator in a short period. The part of the prehospital emergency text describing the condition is similar to the general short text, but compared with the general Chinese short text, it has the following characteristics.

1) More medical terminology. The prehospital emergency text is recorded by professional medical personnel according to the information provided by patients about their illnesses. The text contains much medical
terminology, lacking the deformed words and irregular words found in texts of other fields.

(2) High text similarity. Texts are generally described around a theme. For example, short news texts record sports, finance, and other information, so there are significant differences in wording between various types of short texts, and the text-similarity is low. In contrast, prehospital emergency texts mainly describe patients’ illnesses, and the words used are repetitive, such as pain and discomfort. The presence of similar words makes the description of prehospital emergency texts highly similar.

(3) Fixed text format. The prehospital emergency text consists of several parts, such as gender, age, current medical history, and past medical history, each of which describes certain information. In contrast, texts in other fields are composed of various forms, and the information described in each part varies. Text specification is more applicable to text classification, but it is less studied.

Therefore, a DL-PRE text classification model based on the characteristics of prehospital emergency care was developed. The main contributions of this work are summarized as follows.

(1) To address the problem of many medical terms in prehospital emergency records, we use a pre-trained model, BERT, to obtain word vectors. The prehospital emergency texts train the BERT model to obtain the semantic information of medical terms. Moreover, according to the dataset, we adjust the pre-trained language model task to improve the training efficiency of the model.

(2) Based on the problem of high text similarity, we obtain text information from both global and local aspects by fusing BiLSTM and TextCNN models and using meta activate or not (Meta-ACON) instead of ReLu to further improve the feature extraction ability of the model.

(3) We improve the model’s local text feature extraction ability by weighted text convolutional neural network based on the fixed-text format phenomenon.

The rest of this paper is structured as follows. We review the development of deep learning-based text classification models in section 2. Then, we develop the DL-PER model in section 3 and conduct analytical experiments on text classification in section 4. Finally, concluding remarks are presented in section 5.

II. RELATED WORKS AND METHODS

A. RELATED WORKS

Research on medical text classification mainly focuses on extracting disease information content. Researchers have used fusion, modification, and other methods to improve the classification performance of models based on medical text features.

For example, Li et al. [10] established a three-stage hybrid method combining the gated attention-based bidirectional Long Short-Term Memory and the regular expression-based classifier proposed for medical text classification tasks. In order to improve the classification accuracy of electronic medical records, Mu et al. [11] proposed a neural network-based association classification algorithm. The classification performance of the text was improved by combining external medical knowledge sources [12]. To improve the classification efficiency of biomedical texts, Abdollahi et al. [13] used two new approaches to augment medical data to enrich the training data. Prabhakar et al. [14] used two novel deep learning architectures for medical text classification paradigms to alleviate human efforts. Moradi et al. [15] designed a new method called Biomedical Confidence Item Set Interpretation (BioCIE) to address the problem of black-box machine learning post hoc interpretation of biomedical text classification models. To address the ambiguous vocabulary problem in Chinese medical diagnosis, Liang et al. [16] improved the dual-channel mechanism as an essential enhancement of long short-term memory (LSTM). Ibrahim et al. [17] proposed a hybrid multi-label classification method based on generalized deep learning that can be used to classify different types of biomedical texts. Menon et al. [18] used an efficient classification algorithm for data reduction to attributes using PCA transformation instillation.

B. RELATED METHODS

Text classification is an essential component in the field of natural language processing. Research scholars have studied the word embedding model, feature extraction model, and activation function to improve the model’s classification accuracy.

In the field of word embedding models, word vectors generated by neural network language models can more accurately compute the similarity between words [19]. The improved word2vec word vector model obtains the relationship between words through bag-of-words [20, 21]. The word2vec-based unsupervised learning model doc2vec enhances the ability to learn the text of different lengths by adding paragraph vectors [22]. Bidirectional Encoder Representations from Transformers (BERT) [23] introduced Transformer [24]. The encoder part encodes the text, and through training and dynamic assignment to a large corpus, the feature representation capability of the word vector model is greatly improved. The problem of multiple meanings of a word is effectively solved.

In the field of feature extraction models, researchers have focused their research on recurrent neural networks (RNN), convolutional neural networks (CNN), and model fusion. RNNs are commonly used to solve text classification problems because they can obtain global structural information about the text. Long and short-term memory networks (LSTM) solve the problem of RNN gradient disappearance [25]. BiLSTM can obtain bidirectional semantic information and improve the feature extraction capability of the model [26].

RNN networks cannot extract local semantic information. CNN can ignore the global to extract locally important information. Kim [27] applied convolutional neural networks to text classification. The dynamic
convolutional neural network algorithm solves the problem of classifying text of different lengths [28]. Character-level CNN algorithms input characters rather than words into the algorithm to obtain more generality [29]. Conneau et al. [30] proposed a deep CNN algorithm to improve the model's classification. Although CNN is continuously improved, it still cannot escape the limitation of convolutional kernel size to obtain global text information.

To make up for the shortcomings of RNN and CNN, researchers have made significant achievements by fusing the two models. Jin et al. [31] fused multi-scale CNN and LSTM. The model improved the accuracy of multi-task multi-scale sentiment classification. A deep learning framework combining BiLSTM and CNN was used to identify sentiment labels in psychiatric social texts to improve the model's classification performance [32]. Jin et al. [33] used BiLSTM with an attention mechanism to obtain contextual semantic features of the text and CNN to obtain local semantic features.

The activation function affects the classification performance of the model. ReLu [34] function was proposed to solve the gradient disappearance problem of Sigmoid and Tanh. However, this activation function has the problem of "neuron death" to solve this problem, PReLU [35], Maxout [36], CReLU [37], Mish [38], and ELU [39] were proposed. However, these functions cannot control the activation state of each neuron adaptively according to the data characteristics. Meta-ACON, proposed by Ma et al. [40], can control the activation state of neurons flexibly by learning and highlighting the key information and has obtained higher classification results in image processing.

III. DL-PER Model

A. BERT WORD VECTOR

BERT can encode words flexibly according to the contextual semantics, making the word vector more consistent with the full-text information. Training the BERT model with domain-specific text can obtain specialized lexical information and word associations in the domain text [41, 42].

The BERT word vector composition is shown in Figure 1. The text is segmented using the particular classification marker ([CLS]) at the beginning of each sequence, using the particular marker ([SEP]). The [CLS] is a specific marker at the beginning of the sequence, and the [SEP] is placed later in the sequence to separate different sentences. The BERT word vector consists of a word vector, a sentence vector, and a position vector superimposed. The word vector converts words into a word vector of fixed dimension; the sentence vector distinguishes different sentences with zeros and ones; the position vector labels the sequential properties of the input text.

BERT consists of Masked Language Model (MLM) and Next Sentence Prediction (NSP), which automatically masks some words in the input utterance and predicts the words by the contextual background. The masking mechanism randomly selects a token position for prediction with a 15% probability in each training sequence. If the i-th token is selected, it will be replaced with one of three tokens.

1. 80% probability of replacing the word directly with [Mask]: "The patient fell after drinking more than 10 minutes ago." to "Fall injury. The patient was drunk more than 10 minutes ago and was inadvertently [Mask]."

2. 10% probability of replacing the word with another word: "The patient fell after drinking more than 10 minutes ago." to "The patient had an inadvertent headache after drinking more than 10 minutes ago."

3. 10% probability of not substituting: "The patient had a fall after drinking more than 10 minutes ago." to "The patient had an inadvertent fall after drinking more than 10 minutes earlier."

NSP determines whether sentence B is the following of sentence A. The input is a sentence pair of sentence A and sentence B. There is a 50% probability that sentence B is the following of sentence A. NSP enables BERT to obtain logical and causal relationships between contexts better.

The BERT model was fine-tuned to improve the ability of the model to extract features from prehospital emergency records.

1. Increase the percentage of NSP tasks. The content of prehospital emergency text consists of multiple parts, and the content connection between the parts is weak. Therefore, a weighted approach is used to increase the proportion of NSP tasks to make the pre-trained model more suitable for prehospital emergency text data.

2. Increase the Transformer hierarchy depth. The pre-trained language model using shallow feature representation is difficult to extract the rich content of the prehospital emergency text, which affects the final classification effect.

B. BiLSTM MODEL

BiLSTM can obtain the local semantic information of each part from the global perspective and then get more accurate global feature information of the prehospital emergency text. BiLSTM contains forward LSTM and backward LSTM and finally connects to the same layer of output, which solves the problem that LSTM cannot capture the long sequence of contextual semantic information. Therefore, BiLSTM can obtain prehospital emergency text feature information globally.
The process of encoding the sentences \([x_1, x_2, x_3]\) is shown in Figure 2. The forward LSTM inputs \(x_1, x_2, x_3\), in turn, obtain two hidden state vectors \(h_{1l}, h_{2l}, h_{3l}\), and the reverse LSTM inputs \(x_1, x_2, x_3\), in turn, obtain two hidden state vectors \(h_{1r}, h_{2r}, h_{3r}\) and finally the vectors in both directions are spliced to obtain \(h_1, h_2, h_3\).

BiLSTM contains both forward LSTM and backward LSTM. Three gates in BiLSTM influence other neuron states through the activation function. Assuming that the input sequence is \(X = (x_1, x_2, ..., x_t)\), and \(t\) denotes the \(t\)th word, the sequence is computed through the BiLSTM layers as (1-6).

\[
i_t = \sigma(W_i [h_{t-1}, x_t] + b_i) \\
f_t = \sigma(W_f [h_{t-1}, x_t] + b_f) \\
o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \\
c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \\
h_t = o_t * \tanh(c_t) \\
h_t = h_t + h_{tr}
\]

\(i_t\) is the input gate, \(f_t\) is the forget gate, \(o_t\) is the exit gate, \(C_t\) is the cell state, \(\tilde{C}_t\) is the temporary cell state, \(h_t\) is the hidden layer state, \(\sigma\) is the activation function sigmoid, \(W_i, W_f, W_o\) are the weights, and \(b_i, b_f, b_o\) are the bias term. \(h_t\) denotes the hidden layer state of the \(t\)th word, \(h_{tr}\) denotes the hidden layer state of the \(t\)th word obtained in the forward direction and \(h_t\) denotes the hidden layer state of the \(t\)th word obtained in the reverse direction.

C. WTextCNN Model

TextCNN can extract local text features but has certain limitations, so this paper proposes weighted TextCNN to improve the text feature extraction ability of the model.

The WTextCNN is shown in Figure 3. It is layer consists of two convolutional layers and a pooling layer. The output space of WTextCNN consists of n-gram convolutional filters that have global learning capability and are invariant to the input vector. The first convolutional layer extracts n-grams of features at local locations. Each convolutional kernel can learn the n-gram pattern of the text during training. Then, the convolution result of the first layer is used as the weight. The convolution kernel of the second convolution layer is weighted using the weight. The convolution kernels with important n-gram features will have larger values during the second convolution. In comparison, irrelevant n-grams will have smaller values, effectively capturing the important local features appearing in the text. Pooling can result in more accurate semantic features after combining with the global semantic information highlighted in the activation section.

![Figure 3. Structure of the WTextCNN model](image)

Let the input text be \(X = [x_1, x_2, ..., x_j]\), where \(j\) is the length of the input text, and let \(C = [c_1, c_2, ..., c_m]\) be \(m\) convolution kernels, and use convolution \(C\) to compute the \(X\) text with the formula (7).

\[
M = X \ast C
\]

Where \(\ast\) is the convolution operation. For the \(j\)th convolution kernel \(c_j\). The convolution at the \(j\)th position is computed as (8).

\[
m_{ij} = f(c_t^T \cdot \text{cat}(x_1, x_2, ..., x_{j+n-j}) + b)
\]

where \(f\) is the activation function, \(\text{cat}\) is the splicing operation, and \(n_t\) is the size of the convolution kernel. Each convolutional kernel \(c_j\) can learn some patterns of the n-gram, and the learned patterns are used to perform the weighted summation operation on the convolutional kernels, which is calculated as (9), (10).

\[
C^w = C \cdot M
\]

\[
O = \text{MaxPooling}(C^w)
\]
where $C^w$ is the weighted representation of the convolution kernel, which contains the n-gram local information of the text. Where $O = [o_1, o_2, ..., o_j]$ is the output vector after the maximization pooling operation.

**D. ACTIVATION FUNCTION**

Compared with the ReLu function, the upper and lower bounds of the Meta-ACON function can be obtained by $\beta$ learning, and the activation state of neurons can be controlled by $p_1, p_2$, which solves the problem of “neuron death” and can better ensure the flow of feature information. A comparison of the activation function functions is shown in Figure 4.

![Comparison of ReLu activation function and Meta-ACON activation function](image)

The yellow part is the activated neuron, the white part with $x$ is the inactivated neuron.

**FIGURE 4.** Comparison of ReLu activation function and Meta-ACON activation function

Meta-ACON function is an activation function that improves the ReLu function and can learn whether to activate neurons or not. Meta-ACON function can process the input vector to highlight key information. The formula is as (11).

$$f(x) = (p_1 - p_2) x \sigma \left[ \beta (p_1 - p_2) x \right] + p_2 x$$  \hspace{1cm} (11)

$p_1, p_2$ are parameters that can be learned and are used to adjust the upper and lower bounds of the function by self-adaptation. $\beta$ is the activation factor that controls whether the neuron is activated or not.

$p_1, p_2$ can control the upper and lower bounds of the function to get a more suitable smooth curvature by learning. The image of its activation function is shown in Figure 5.

![Activation function graph](image)

**FIGURE 5.** Activation function graph

$\beta$ is responsible for controlling whether to activate the neuron or not. The calculation formula is shown as (12).

$$\beta_c = \sigma W_1 W_2 \sum_{h=1}^{H} \sum_{w=1}^{W} x_{c,h,w}$$  \hspace{1cm} (12)

$\sigma$ is the sigmoid function. $W_1, W_2$ are the two convolution operations, and $c \cdot H$, and $W$ are the three channels of the image respectively.

In natural language processing, $c$ is the number of input texts, $H$ is the number of word vectors of the input texts, and $W$ is the dimensionality of the word vectors. Based on the characteristics of the text input, $\beta$ is modified to the formula (13)

$$\beta_c = \sigma W_1 x_{c,h}$$  \hspace{1cm} (13)

The convolution operation is first performed on each word vector, and then the result is obtained by the sigmoid function.

The Meta-ACON activation function replaces the ReLu function further to improve the feature extraction capability of the model.

**F. STRUCT OF DL-PER MODEL**

This section mainly introduces the framework structure of the DL-PER model, as shown in Figure 6. The DL-PER model can be described as follows:

Preprocess the original text to get the segmented text as the input layer of DL-PER models and extract text
features. The model uses BERT to obtain the word vector matrix. Then the global aspect semantic information is obtained using BiLSTM, and the local aspect semantic information is extracted using WTextCNN. Finally, the prediction results are output by the classifier. The detailed procedures of our proposed DL-PER model are given in visualized algorithm format, as shown in Table 1.

**TABLE 1. DESCRIPTION OF PREHOSPITAL EMERGENCY RECORD CLASSIFICATION ALGORITHM BASED ON DL-PER MODEL.**

| Algorithm: DL-PER model. The prehospital emergency record classification is based on feature fusion. |
|--------------------------------------------------|
| Input: \( \{ x_1 , x_2 , \ldots , x_l \} \), a set of \( l \) prehospital emergency records. |
| Output: \( \{ y_1 , y_2 , \ldots , y_n \} \), the category label of each record. |
| Methods: (The specific processing steps of the algorithm are introduced as follows) |
| Begin |
| 1. \( X^b_i = BERT(x_i) \); /* BERT outputs word vectors based on the text information, processes each text into \( n_i \times d \) dimension vector matrix, where \( n_i \) is the number of words and \( d \) is the dimension of BERT vector. */ |
| 2. \( X^c_i = BiLSTM(X^b_i) \); /*Deploy the two-layers BiLSTM and one-layer Dropout to output the text vector*/ |
| 3. \( X^e_i = WTextCNN(X^c_i) \); /* Deploy two layers of TextCNN and one layer of Pooling to output local feature vectors and use Meta-ACON to increase the ability of the model to acquire features. */ |
| 4. \( y_i = classifier(X^e_i) \); /*Classify the text and output the category label. */ |
| End |

**FIGURE 6.** The structure of the DL-PER model.

---

**A. DATA PROCESSING**

1) THE SOURCE AND CONTENT OF PREHOSPITAL EMERGENCY RECORDS

In this study, prehospital emergency records were retrieved from the Emergency Medical Rescue Command Centre of Handan, a municipal city in the middle plain area of China with over ten million. Prehospital emergency records include all prehospital emergency records from 2018 to 2019. A senior physician labels each prehospital diagnosis record according to the 10th International Classification of Diseases (ICD-10). This system classifies diseases according to rules based on specific characteristics of the disease and represents them using codes.

The prehospital emergency text consists of multi-part information. The layperson provides preliminary information about the patient, including age, gender, chief complaint, current medical history, and past medical history. The chief complaint and current medical history information are complex, containing the patient's current status and etiology. Some patients have no etiology in their chief complaint and current medical history. The professional makes a preliminary diagnosis of the patient based on the information provided. The prehospital emergency records are shown in Table 2.
TABLE 2. EXAMPLE OF PREHOSPITAL EMERGENCY RECORD

| age | gender | chief complaint | current medical history | past medical history | diagnosis |
|-----|--------|----------------|-------------------------|---------------------|-----------|
| 77  | man    | Coma, the patient appeared to be in respiratory and cardiac arrest more than 10 minutes ago. | The patient was seen lying flat on the bed with loss of consciousness, no response to call, loss of carotid artery pulsation, no respiration, and no heartbeat. She called 120 and was retrieved by 120 of our hospital. After retrieval, the arteriovenous ultrasound of the left lower extremity showed: venous thrombosis of the left lower extremity. | Previous history of cerebral infarction, hypertension, and coronary artery disease. Previously treated with surgery for lower extremity fracture 9 days ago | Respiratory and cardiac arrest |
| 65  | woman  | The fracture was 10 days old and swollen, due to the swelling and pain in the left lower limb found by the family half a day ago. | |

2) PREHOSPITAL EMERGENCY RECORDS PREPROCESSING

The prehospital emergency records need to be preprocessed before inputting into the model, as shown in Figure 7. First, medical experts re-label the prehospital emergency records to ensure the correct classification. Secondly, the information in each part of the prehospital emergency record is stitched together into a sentence. Then, unlike English, Chinese words are constructed differently, requiring a word splitting tool to segment the entire sentence. Finally, words that are not semantically relevant, such as "ah", were deleted.

3) GROUPING OF PREHOSPITAL EMERGENCY RECORDS

Prehospital emergency records are of various types and are unevenly distributed. They were divided into four groups according to the distribution of the number of types, as shown in Table 3.

The number of disease types in each group was not the same and contained different disease types, both trauma and cerebrovascular diseases. Group 1 has the most significant amount of data and contains more disease types, and group 4 has the least amount of data.

B. MODEL EVALUATION INDICATORS

In this paper, the F1 score was used as the evaluation index criterion for the classification effect of the model. The recall rate statistic predicts how many of the results with A are correctly classified; the recall rate statistic labels how many of the samples with A are correctly classified; recall and accuracy are considered from two different perspectives, but they are often contradictory.

The formula for calculating the precision can be expressed as (16):

\[ P = \frac{TP}{TP + FP} \]  

The formula for calculating the recall can be expressed as (17):

\[ R = \frac{TP}{TP + FN} \]  

The formula for the calculation of the F1 score can be expressed as (18):

\[ F1 = \frac{2 \times P \times R}{P + R} \]

In the confusion matrix, TP, FP, TN, and FN are the number of true cases, false-positive cases, true negative cases, and false-negative cases, respectively. The range of the F1 score is between 0 and 1, and the larger the F1 score, the higher the algorithm's performance.

TABLE 3. PREHOSPITAL EMERGENCY RECORD GROUPING INFORMATION

| Group | Category of disease                      | Number of each category | Number of categories |
|-------|----------------------------------------|-------------------------|---------------------|
| 1     | Respiratory arrest, headache, anterior epidermal scalp injury, injury to the head, lumbar injury, multiple injuries, upper extremity injury, lower extremity fracture, lower extremity injury, alcohol intoxication | 1700                    | 10                  |
| 2     | Stroke, cerebrovascular disease, palpitations, dyspnea, chest pain, abdominal pain, coma, fever, malaise and fatigue, syncope and collapse, convulsions, other and unspecified, fractures of the upper extremities, examination observation received after a traffic accident | 1000                    | 13                  |
| 3     | Heart disease, hypertensive cerebral hemorrhage, gastrointestinal hemorrhage, vomiting, eye injury, localized brain injury, chest injury, injury to a person in a vehicle accident, mainly other and unspecified harmful effects affecting the cardiovascular system | 500                     | 9                   |
| 4     | Hypoglycemia, epilepsy, acute myocardial infarction, acute heart failure, lumbar fractures, poisoning by drugs, pharmaceuticals biologics, carbon monoxide poisoning, slips, trips and falls on the same plane, harmful effects of other and unspecified drugs that primarily affect the autonomic nervous system | 300                     | 9                   |

C. THE SETTING OF MODEL PARAMETERS

This subsection discussed the parameter setting in 4 links (BERT vector, BiLSTM, WTextCNN, and optimizer) while establishing the DL-PER model. Taking Group 1 as an...
example, we explained the detailed parameter configuration in each part below.

BERT vector part: Considering that the median length of the prehospital emergency text is 50, the length of the text is unified to 50, and the dimension of each word vector is 768.

BiLSTM part: Peters et al. [43] proposed a two-layers BiLSTM model and achieved a good representation of text features. The number of layers of the BiLSTM can affect the training factor and accuracy of the model. The number of layers of BiLSTM is set to [1, 2, 3, 4], respectively, and group 1 data is used for testing. The final test results are shown in Figure 8. As the number of BiLSTM layers increases, the F1 score gradually becomes more extensive, while the time for each training also becomes longer. Then a two-layer BiLSTM with the highest F1 score and faster training time was selected.

![Figure 8. F1 score and training speed of the number of BiLSTM layers](image)

WTexCNN part: The convolutional kernel size of WTextCNN can affect the training and classification results of the model. The convolutional kernel sizes were set to [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], and the data from the first group were used for training and testing. The test results are shown in Fig. 9. Li et al. [44] proposed that using the convolutional kernel size with the highest accuracy and close to the classification effect can improve. So 3, 4, and 5 were chosen as the size of the convolution kernel.

![Figure 9. F1 score of the size of the convolution kernel](image)

Dropout can avoid overfitting and affects the performance of the model. Dropout layers are available in BiLSTM and WTextCNN. The dropout rates were set to [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9], and the data from the first group were trained and tested in the BiLSTM and WTextCNN models, respectively. The test results are shown in Figure 10. When the dropout rate is 0.5, the BiLSTM model achieves the highest value. The WTextCNN model achieved the highest value when the dropout rate was 0.2. Then, 0.5 was chosen as the dropout value for BiLSTM, and 0.2 was chosen as the dropout value for WTextCNN.

![Figure 10. F1 score of the dropout value](image)

Optimizer part: In this paper, the model chooses Nadam as the optimizer, which is Adam's optimization algorithm with high computational efficiency and fast convergence. As a super essential parameter in supervised learning and deep learning, the learning rate determines if and when the objective function converges to a local minimum. An appropriate learning rate allows the objective function to converge to a local minimum in a suitable time. The learning rate is set to [10⁻¹, 10⁻², 10⁻³, 10⁻⁴, 10⁻⁵], and the model is trained and tested using set 1 data, respectively. The test results are shown in Figure 11. Then 10⁻⁴ was chosen as the learning rate.

![Figure 11. F1 score of the learning rate](image)
According to the experimental test results, the hyperparameters of BERT, BiLSTM, WTextCNN, and optimizer are set as shown in Table 4.

| parameter           | value  |
|---------------------|--------|
| BERT                |        |
| number of transformer blocks | 12     |
| hidden layer unit   | 768    |
| number of attention head | 12     |
| word vector dimension | 768    |
| BiLSTM              |        |
| embedding size      | 400    |
| units               | 128    |
| dropout             | 0.5    |
| WTextCNN            |        |
| filter              | (3, 4, 5) |
| activation          | ReLU   |
| pool                | 1-max pooling |
| dropout             | 0.2    |
| optimizers          |        |
| optimizers          | Nadam  |
| learning rate       | 10^4   |
| beta1               | 0.9    |
| beta2               | 0.999  |
| epsilon             | 10^4   |
| schedule decay      | 0.004  |
| loss                | sparse categorical crossentropy |
| batch               | 512    |

### TABLE 4. HYPERPARAMETER SETTING.

D. RESULTS & MODEL COMPARISON

To validate the performance of the DL-PER model for prehospital emergency record classification, we built a model environment and learned the model under the training text. The classification effectiveness of the model based on the test text is then evaluated by the evaluation metrics mentioned above. The DL-PER model is compared with the baseline models of LSTM, BiLSTM, TextCNN, BiLSTM-Attention, TextCNN-Attention, and BiLSTM-TextCNN, respectively. The word2vec word vector model is used to compare with BERT. Comparison with MetaCON using ReLU, PReLU functions.

In this paper, the DL-PER model is compared with models such as BiLSTM, TextCNN, and BiLSTM-Attention. The test results are shown in Table 5. DL-PER achieves better classification results on all four datasets. Compared with the BiLSTM model with the lowest F1 score, the DL-PER model improved the F1 scores by 5.7%, 6.8%, 5.7%, and 4.9% in the four groups. It proves that the DL-PER model is more accurate in acquiring text features and improving the text classification performance of the model.

The experimental results show that the same model has different classification effects on four data groups. Group 4 had the highest F1 score, and group 2 had the lowest F1 score. The data of group 2 contained many identical words, such as "dizziness" and "pain" appearing in each label of group 2. The occurrence of identical words increases classification difficulty and demands the model to obtain deeper text features.

The DL-PER model achieves higher F1 scores than other comparison classification models in the four sets of test results. The model can effectively extract local semantic information from prehospital emergency text. The difference between the F1 scores of BiLSTM and TextCNN is minor. It may be because BiLSTM has difficulty extracting local text features from deep levels. In contrast, TextCNN has limited feature extraction ability due to the limitation of convolutional kernel size, so these two models have difficulty achieving high F1 scores. The F1 scores of BiLSTM-Attention, TextCNN-Attention, and BiLSTM-TextCNN are higher than BiLSTM and TextCNN models. It indicates that the fusion model can accurately acquire text features and improve the model's classification performance. The DL-PER model further enhances the ability to acquire local information in the text using WTextCNN after acquiring distant text information. Compared with the TextCNN layer, the WTextCNN layer is more capable of acquiring local text hidden information, thus improving the text classification performance of the model.

### TABLE 5. COMPARISON OF EXPERIMENTAL RESULTS

| model                   | Group 1 | Group 2 | Group 3 | Group 4 |
|-------------------------|---------|---------|---------|---------|
| BiLSTM                  | 67.44   | 61.69   | 65.11   | 78.78   |
| TextCNN                 | 67.01   | 61.48   | 65.49   | 78.84   |
| BiLSTM-Attention        | 69.23   | 62.34   | 66.24   | 79.53   |
| TextCNN-Attention       | 69.34   | 63.78   | 66.11   | 79.21   |
| BiLSTM-TextCNN          | 69.77   | 63.48   | 67.63   | 80.22   |
| DL-PER                  | 71.53   | 66.21   | 69.03   | 82.89   |

This paper uses word2vec (WV) and BERT models to obtain text word vectors. The word vectors of word2vec and BERT are input to the BiLSTM, TextCNN, and BiLSTM-Attention models, respectively, for training. The final test results are shown in Figure 12. The experimental results show that the F1 score of the model using BERT word vectors is significantly higher than that of word2vec. It proves that BERT word vectors can fully exploit the semantic information between texts and improve the model's classification performance than word2vec word vectors.

The activation functions of WTextCNN in the DL-PER model were experimented with using ReLU, PReLU, and MetaACON excitation, respectively. The experimental results are shown in Table 6. The experimental results show that the F1 score of the MetaACON model is higher than that of ReLU and PReLU. The MetaACON function can determine the activation state of each neuron by learning, which can highlight the key semantic information more than ReLU and PReLU. The accuracy variation of the model training process is shown in Figure 13. In Figure 13 (a), (b), and (c), the images using the MetaACON activation function model converge faster and with higher accuracy.
than ReLu and PReLu. It indicates that Meta-ACON can speed up the training and improve the model’s classification performance. However, Figure 13(d) shows that the model converges at the same rate for each activation function when training for group 4. It may be because the amount of data in group 4 is too small to highlight the advantage of Meta-CON.

**TABLE 6. COMPARISON EXPERIMENTAL RESULTS OF DIFFERENT ACTIVATION FUNCTIONS**

| model     | Group 1 | Group 2 | Group 3 | Group 4 |
|-----------|---------|---------|---------|---------|
| ReLu      | 69.78   | 64.83   | 67.82   | 81.76   |
| PReLu     | 70.98   | 65.13   | 68.93   | 82.06   |
| Meta-ACON | 71.83   | 66.28   | 69.03   | 82.89   |

**FIGURE 12. Effect of word vector model on F1 score**

**FIGURE 13. Effect of different activation functions on model training**

V. CONCLUSIONS

Prehospital emergency texts have problems such as many medical terms, the short length, and standardized format. In this paper, we propose a DL-PER text classification model to solve the problem of prehospital emergency text classification. Through the experiment, we have made the following findings.

1. The BERT model is better than the word2vec model for obtaining semantic information in specialized domains. Improving the feature extraction ability of the model solves the problem of more terms in the professional domain.

2. The weighted convolution operation effectively improves the local feature extraction ability of the model. After combining with the global text features extracted by BiLSTM, the feature extraction ability of the model is significantly improved.

3. The activation function Meta-ACON can improve the training efficiency of the model and improve the feature extraction ability of the model. However, its advantage is less evident in the case of a small amount of data.

The DL-PER model obtains the contextual semantic information of long-range text through BiLSTM. Then the local information extraction capability is improved by WTextCNN. In this way, the model can obtain more accurate semantic information about the critical semantics of the text. We use the Meta-ACON activation function instead of the ReLu function to improve the model's performance. The effectiveness of DL-PER is proved...
through experiments, and the model is helpful for reference in professional domain text classification.

REFERENCES

[1] Q. Zhou, W. Y. Tian, R. Y. Wu, C. Z. Qin, H. J. Zhang, H. Y. Zhang, S. D. Zhou, S. W. Li, Y. Z. Jin and Z. J. Zheng, “Quantity and Quality of Healthcare Professionals, Transfer Delay and In-hospital Mortality Among ST-Segment Elevation Myocardial Infarction: A Mixed-Method Cross-Sectional Study of 89 Emergency Medical Stations in China.” Frontiers in public health, vol. 9, Jan 2022, doi:10.3389/fpubh.2021.812355.

[2] J. Emberson, K. R. Lee, P. Lyden, L. Blackwell, G. Albers, E. Bluhmki, T. Brott, G. Cohen, S. Davis, G. Donnan, J. Grotta, G. Howard, M. Kaste, M. Koga, R. v. Kummer, M. Lansberg, R. I Lindley, G. Murray, J. M. Olivot, M. Parsons, B. Tilley, D. Toni, K. Toyoda, N. Wahlgren, J. Wardiaw, W. Whiteley, G. J d. Zoppo, C. Bajgant, P. Sandercoc, and W. Hacke. Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials. , vol. 384, no. 9958, pp. 1929-1935, 2014.

[3] H. A. Park, K. O. Ahn, E. J. Lee and J. O. Park. Association between survival and time of on-scene resuscitation in refractory out-of-hospital cardiac arrest: A cross-sectional retrospective study. International journal of environmental research and public health, vol. 18, no. 2, pp. 496, 2021.

[4] W. Wang, M. L. Zhu, Y. J. Wang, Z. S. Wu, R. L. Gao, L. Z. Kong and S. S. Hu. Summary of the China Cardiovascular Disease Report 2012. Chinese circulatio journal, vol. 28, no. 6, pp. 408-412, 2013.

[5] X. F. Shi, J. T. Bao, H. L. Zhang, H. Wang, Y. Wang, L. Li, and P. Hou. Emergency medicine in China: A review of the history of progress and current and future challenges after 40 years of reform. Am J Emerg Med, vol. 38, no. 3, pp. 662-669. Mar 2020, doi: 10.1016/j.ajem.2019.11.008.

[6] A. Baghne, A. Samman, V. D. H. PG, F. W. Asselbergs, and D. L. and Oberski. ETM: Enrichment by topic modeling for automated clinical sentence classification to detect patients' disease history. Journal of Intelligent Information Systems, vol. 2020, no. 1, pp. 1-21, 2020.

[7] M. Ormerod, J. Martinez-Del-Rincon, N. Robertson, J. M. D. Rincon and B. Devereux. Analysing Representations of Memory Impairment in a Clinical Notes Classification Model, 18th BioNLP Workshop and Shared Task (BioNLP 2019), 2019.

[8] S. Mao, L. Zhanglu, G. Guanzhen. An LSTM&Topic-CNN Model for Classification of Online Chinese Medical Questions. IEEE Access, vol. 9, pp. 52580-52589, April 2021. doi:10.1109/access.2021.3070375.

[9] H Tohira, J Finn, S Ball, D. Brink and P. Buzzacott. Machine learning and natural language processing to identify falls in electronic patient care records from ambulance attendances. Inform Health Soc Care., vol. 55, no. 2, pp. 329-349, 2021.

[10] X. Li, M. L. Cui, J. P. Li, and R. Bai. A hybrid medical text classification framework: Integrating attentive rule construction and neural network. Neurocomputing, vol. 443, pp. 345-355, July, 2021.

[11] X. L. Mu, H. Y. Zhang. Embedded electronic medical record text data mining using neural network association classification algorithm. Expert systems, Dec., 2021, doi: 10.1111/esxy.12874.

[12] Q. Zhou, W. Peng, D. Tang. Automatic recommendation of medical departments to outpatients based on text analyzes and medical knowledge graph. Journal of intelligent & fuzzy systems; vol. 41, no. 2, pp. 3289-3299, Oct. 2021.

[13] M. Abdollahi, X. Gao, Y. Mei, S. Ghosh, J. Y. Li and M. Narag. Substituting clinical features using synthetic medical phrases: Medical text data augmentation techniques. Artificial Intelligence in Medicine, vol. 120, pp. 102167. Oct. 2021, doi: 10.1016/j.artmed.2021.102167.

[14] S. K. Prabhakar, D. Won. Medical text classification using hybrid deep learning models with multilhead attention. Computational intelligence and neuroscience, vol. 2021, pp. 9425655, Sep. 2002, doi: 10.1155/2021/9425655.

[15] M. Moradi, M. Samwaid. Explaining black-box models for biomedical text classification. IEEE journal of biomedical and health informatics, vol. 25, no. 8, pp. 31120-3120, Aug. 2021. doi: 10.1109/JBHI.2021.3056748.

[16] S. B. Liang, X. N. Chen, J. X. Ma, W. C. Du and H. W. Ma. An improved double channel long short-term memory model for medical text classification. Journal of healthcare engineering, vol. 2021, pp. 6664893, Feb. 2021, doi:10.1155/2021/6664893.

[17] A. M. Ibrahim, M. U. G. Khan, F. Mehmoond, M. N. Aisim and W. Mahmood. A gene-based hybridized shallow neural network for multi-label biomedical text classification. Journal of Biomedical informatics, vol. 116, pp. 103699, Mar. 2021, doi: 10.1016/j.jbi.2021.103699.

[18] S. M. Menon, P. Rajarajawari. A novel approach for multi variant classification of medical data in short text. Journal of scientific & industrial research, vol. 80, no. 5, pp. 457-462, May, 2021.

[19] Y. Benjio, R. Ducharme, P. Vincent and C. Jauvin. A neural probabilistic language model. Journal of Machine Learning Research, vol. 3, pp. 1137-1155, 2003.

[20] T. Mikolov, K. Chen, G. Corrado and J. Dean. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.

[21] T. Mikolov, I. Sutskever, K. Chen, C. Greg and D. Jeffrey. Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, vol. 26, pp. 3111-3119, 2013.

[22] Q. LE, T. MIKOLOV. Distributed representations of sentences and documents. Proceedings of the 31st International Conference on Machine Learning, pp. 1188-1196, 2014.

[23] J. Devlin, M. W. Chang, K. Lee and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805, 2018.

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin. Attention Is All You Need. arXiv:1706.03762, 2017.

[25] Hochreiter, S. Schmidhuber J. Long Short-Term Memory. Neural Computation, vol. 8, no. 9, pp. 1735-1780, 1997.

[26] M Schuster, K. K. Paliwal. Bidirectional recurrent neural network. IEEE transactions on Signal Processing, vol. 45, no. 11, pp. 2673-2681, 1997.

[27] Y. KIM. Convolutional neural networks for sentence classification. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Pao Alto, USA:AAAI Press, pp. 1746-1751, 2014.

[28] N. Kalchbrenner, E. Grefenstette and P. Blunsom. A convolutional neural network for modeling sentences. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistic. Stroudsburg, USA:ACL, pp. 655-665, 2014.

[29] X. Zhang, J. Zhao, Y. Lecun. Character-level Convolutional Networks for Text Classification. Neural Information Processing Systems. MIT Press, 2015.

[30] A. Conneau, H. Schwenk, L. Barrault and Y. Lecun. Very Deep Convolutional Networks for Text Classification. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics; vol. 1, 2017.

[31] N. Jin, X. Wujia, X. Ma, K. Yan and Y. C. Mo. Multi-task learning model based on multi-scale CNN and LSTM for sentiment classification. IEEE Access, vol. 8, pp. 77060-77072, 2020. doi: 10.1109/ACCESS.2020.2989428.

[32] H. L. Wu, Y. Y. He, L. C. Yu and K. R. Lai. Identifying emotion labels from psychiatric social texts using a bidirectional LSTM-CNN model. IEEE Access, vol. 8, pp. 66638-66646, 2020. doi: 10.1109/ACCESS.2020.2985228.

[33] B. B. Jin, X. R. Xue, W. J. Peng, W. Cai, Y. M. Zhang and L. Zhang. TBL-c:Attention: A deep neural network model for recognizing the emotional tendency of Chinese medical
comment. *IEEE Access*, vol. 8, pp. 96811-96828, 2020. doi: 10.1109/ACCESS.2020.2994252.

[34] X. Glorot, A. Bordes, Y. Bengio. Deep Sparse Rectifier Neural Networks. *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 15, pp. 315-323, 2011.

[35] B. Xu, N. Wang, T. Chen, et al. Empirical evaluation of rectified activations in convolutional network. Available: arXiv preprint arXiv: 1505.00853, 2015.

[36] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville and Y. Bengio. Maxout networks. *International Conference on Machine Learning*, pp. 1319-1327, 2013.

[37] W. Shang, K. Sohn, D. Almeida and H. Lee. Understanding and improving convolutional neural networks via concatenated rectified linear units. *International Conference on Machine Learning*, pp. 2217-2225, 2016.

[38] M. Diganta. Mish: A Self Regularized Non-Mo-notonic Neural Activation Function[Online]. 2019. Available:https://arXiv.org/pdf/1908.08681.pdf.

[39] D. Clevert, T. Unterthiner, S. Hochreiter. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *International Conference on Learning Representations*, pp. 1-14, 2016.

XU ZHANG. He was born in Hebei, China, in 1996. He received his bachelor’s degree from Hebei University of Engineering in 2019. He is currently pursuing his master’s degree in the School of Information and Electrical Engineering at the Hebei University of Engineering. His research interests include natural language processing and deep learning.

HAIFANG ZHANG was born in Hebei, China, in 1978. She graduated from Tongji University School of Medicine in Shanghai in 2003, received her bachelor’s degree from Hebei Engineering University in 2009, and her master’s degree from Xinxiang Medical College in 2013. Since 2007, she has been working at the Emergency Rescue Command Center in Handan, Hebei Province. Her research interests include emergency first aid and emergency rescue.

FENG TIAN. He was born in Tianjin, China, in 1981. He received his B.S. and Ph.D. from Tsinghua University and was a visiting scholar at the National Center for Data Mining from 2007 to 2008. Since 2019, he has been the Dean of the School of Medicine at Hebei University of Engineering. His research interests include bioinformatics and health big data analytics.

LONG SHENG. He was born in Hebei, China, in 1982. He received his bachelor’s degree from Central South University in 2004 and his Ph.D. from the University of Electronic Science and Technology in 2012. Then he worked at the University of Information and Electrical Engineering, Hebei University of Engineering. He has published 9 papers. His research interests are natural language processing and machine learning.