Deep Cardiovascular Disease Prediction with Risk Factors Powered Bi-Attention

Yanlong Qiu (✉ lankyqiu@163.com)

Northwest Normal University <https://orcid.org/0000-0003-4940-5546>

Zhichang Zhang

Northwest Normal University

Xiaohui Qin

Northwest Normal University

Shengxin Tao

Northwest Normal University

Research article

Keywords: Chinese electronic medical record, CVD risk factors extraction, CVD prediction, Deep learning, Bi-directional attention, Information fusion

Posted Date: May 29th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-31061/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Deep Cardiovascular Disease Prediction with Risk Factors Powered Bi-Attention

Yanlong Qiu^{1§}, Zhichang Zhang^{1,2}, Xiaohui Qin³, Shengxin Tao⁴

¹ College of Computer Science and Engineering, Northwest Normal University, Lanzhou, China

² College of Computer Science and Engineering, Northwest Normal University, Lanzhou, China

³ College of Computer Science and Engineering, Northwest Normal University, Lanzhou, China

⁴ College of Computer Science and Engineering, Northwest Normal University, Lanzhou, China

[§]Corresponding author

Email addresses:

Yanlong Qiu: lankyqiu@163.com

Zhichang Zhang: zzc@nwnu.edu.cn

Xiaohui Qin: qinxh_qj@163.com

Shengxin Tao: Taoshengxin11@163.com

Deep Cardiovascular Disease Prediction with Risk Factors Powered Bi-Attention

Abstract

Background

Cardiovascular disease (CVD), as a chronic disease, has been perplexing human beings and is one of the serious diseases endangering life and health. Therefore, using the electronic medical record information of patients to automatically predict CVD has important application value in intelligent auxiliary diagnosis and treatment, and is a hot issue in intelligent medical research. In recent years, attention mechanism has been successfully extended to various tasks of natural language processing. Typically, these methods use attention to focus on a small part of the context and summarize it with a fixed-size vector, coupling attention in time, and/or often forming a uni-directional attention.

Methods

In this paper, we propose a CVD risk factors powered bi-directional attention (RFPBiA) network, which is a multi-stage hierarchical process, representing information fusion at different granularity levels, and uses the bi-directional attention to obtain the text representation of risk factors without early aggregation.

Results

The experimental results show that the proposed method can obviously improve the performance of CVD prediction, and the F-score reaches 0.9424, which is better than the existing related methods.

Conclusions

We propose to extract the risk factors leading to CVD by using the existing mature entity recognition technology, which provides a new idea for disease prediction tasks. Moreover, the memory-*less* attention mechanism in both directions in our proposed prediction model of RFPBiA can fuse the character sequence and the risk factors contained in the electronic medical record text to predict CVD.

Keywords: Chinese electronic medical record, CVD risk factors extraction, CVD prediction, Deep learning, Bi-directional attention, Information fusion

Background

Cardiovascular disease (CVD) continues to be a leading cause of morbidity and mortality in the world [1-3]. According to data released by the World Health Organization, CVD is the number one cause of death worldwide, with more deaths from CVD each year than any other cause of death. In 2016, an estimated 17.9 million people died of CVD, accounting for 31% of all deaths worldwide. In its 2018 report, China's National Center for CVD noted that CVD mortality remained at the top of 2016, higher than cancer and other diseases, and the number of patients was as high as 290 million. As a chronic disease, CVD will not obviously show the corresponding characteristics in daily life in the hidden stage. However, once the symptoms are manifested, it may affect the life safety of the patient. Therefore, we want to help doctors make rapid diagnosis in time by analyzing the Electronic Medical Record (EMR) of patients during their daily physical examination.

As CVD risk increases in China, interest in strategies to mitigate it is growing. However, information on the prevalence and treatment of CVD in daily life is limited. But in the medical field, many hospitals have been able to systematically accumulate medical records for a large number of patients by introducing an EMR system. Deep learning has been successfully applied to medical field based on accumulated EMR data [4, 5]. In particular, many studies have been conducted to predict the risk of cardiovascular disease in order to prevent cardiovascular diseases with a high mortality rate globally [6]. Because EMR data is based on patient records, it contains information on the pathogenesis of CVD. However, we found that there is a large amount of irrelevant information in most EMRs. For example, a complete medical record contains only 10 valid information records leading to diseases. The excessively irrelevant information not only reduces model's emphasis on effective disease information. In the field of natural language processing, such problems also exist in text classification tasks. In the article [7], the model proposed by Huang et al. avoids redundant information in the text by skipping the content in the text. Therefore, we intend to use the now steadily developing entity identification model architecture to extract key information. **Table 1** is the key information we considered, including 12 risk factors. In response, we propose to extract the risk factors of pathogenesis from EMRs and take the time attributes of these risk factors. In fact, although the training time of the model is reduced based on the extracted CVD risk factors, the experimental results are not optimistic. After experimental analysis, we think that the main reason is the lack of certain contextual information.

In view of this situation, we introduce the Risk Factors Powered Bi-Attention (RFPBiA) network, a hierarchical multi-stage architecture for modeling the representations of the EMR context paragraph at different levels of granularity (**Figure 3**). RFPBiA includes character-level and contextual embeddings, and uses bi-directional attention flow to obtain a risk factors representation. In this regard, we will predict that the network can take into account both the risk factor information and the context information in EMR. This is very critical for our prediction task. For example, the increased correlation between hypertension risk and controlling blood pressure risk can better predict whether patients suffer from CVD. The experimental results show that the F-score reaches 0.9424, which fully demonstrates the effectiveness of our proposed method and network architecture. To sum up, our contribution is four-fold, with the following conclusions:

- Some of the key information in patients' EMRs is to determine whether they have CVD. As we gave the example above, the number of these key factors appearing in the whole EMRs is limited. Just as our experimental data only take into account the 12 factors in **Table 1**, these factors have been determined

Table 1 Attributes of CVD

No.	Attributes Name	Description
1.	Overweight/Obesity (O2)	A diagnosis of patient overweight or obesity
2.	Hypertension	A diagnosis or history of hypertension
3.	Diabetes	A diagnosis or a history of diabetes
4.	Dyslipidemia	A diagnosis of dyslipidemia, hyperlipidemia or a history of hyperlipidemia
5.	Chronic Kidney Disease (CKD)	A diagnosis of CKD
6.	Atherosclerosis	A diagnosis of atherosclerosis or atherosclerotic plaque
7.	Obstructive Sleep Apnea Syndrome (OSAS)	A diagnosis of OSAS
8.	Smoking	Smoking or a patient history of smoking
9.	Alcohol Abuse (A2)	Alcohol abuse
10.	Family History of CVD (FHCVD)	Patient has a family history of CVD or has a first-degree relative (parents, siblings, or children) who has a history of CVD
11.	Age	The age of the patient
12.	Gender	The gender of patient

through long-term research by doctors. We define this key information as risk factors leading to CVD.

- Our attention layer is not used to summarize the context paragraph into a fixed-size vector. Instead, the attention is computed for every time step, and the attended vector at each time step, along with the representations from previous layers, is allowed to flow through to the subsequent modeling layer. This reduces the information loss caused by early summarization.
- We use a memory-*less* attention mechanism. That is, the attention at each time step is a function of only the risk factor and the EMR context paragraph at the current time step and does not directly depend on the attention at the previous time step. We hypothesize that this simplification leads to the division of labor between the attention layer and the modeling layer. It forces the attention layer to focus on learning the attention between the risk factor and the context, and enables the modeling layer to focus on learning the interaction within the factor-aware context representation (the output of the attention layer). It also allows the attention at each time step to be unaffected from incorrect attendances at previous time steps.
- We use attention mechanisms in both directions, factor-to-context and context-to-factor, which provide complimentary information to each other.

Methods

The main idea of this paper is to predict whether a patient has CVD by focusing on the risk factors in EMRs. First of all, we need to prepare the data we need. The user enters the appropriate values from his/her EMR report. After this, the historical dataset is uploaded. The fact that most medical dataset may contain missing values makes this accurate prediction difficult. So, for this missing data, we have to transform the missing data into structured data with the help of a data cleaning and data imputation process. After preparing the data, we mainly perform the following two steps. Firstly, the risk factors in the EMRs and their corresponding labels are extracted using the relatively mature entity recognition technology that has been developed. In addition to Age and

Figure 1 The main process of CVD prediction.

Gender, the labels for other risk factors include the type of the risk factor and its temporal attributes. We only use the CRF layer to identify the F-score of the extraction result to reach 0.8994. When we use bidirectional LSTM with a CRF layer (BiLSTM-CRF), the F-score identifying the extraction results reached 0.9073. In contrast, the BiLSTM-CRF model has better recognition performance, so we consider using it to extract the risk factors and corresponding labels in EMRs. These extracted risk factors that carry the corresponding labels serve as the basis for input and predict CVD. In the end, by using the RFPBiA model, we can predict whether a patient has CVD. **Figure 1** shows above-mentioned two main processes for predicting CVD.

Technical Details of BiLSTM-CRF Model

The architectures of BiLSTM-CRF model illustrated in **Figure 2**. In the model, the BIO (Begin, Inside, Outside) tagging scheme is used. The $Q = (q_1, \dots, q_{k-3}, \dots, q_k)$ represents the context information carried by the character embedding trained by Skip-gram. The Skip-gram model predicts surrounding words given the current word [8]. The HyC is represented as Hypertension and its temporal attribute is Continue. The HyD is represented as Hypertension and its temporal attribute is During. It is similar to the ones presented by Huang et al. [9], Lample et al. [10] and Ma et al. [11].

Given a sentence, the model predicts a label corresponding to each of the input tokens in the sentence. Firstly, through the embedding layer, the sentence is represented as a sequence of vectors $X = (x_1, \dots, x_t, \dots, x_n)$ where n is the length of the sentence. Next, the embeddings are given as input to a BiLSTM layer. In the BiLSTM layer, a forward LSTM computes a representation $\vec{h}_t$ of the sequence from left to right at every character t , and another backward LSTM computes a representation $\overleftarrow{h}_t$ of the same sequence in reverse. These two distinct networks use different parameters, and then the representation of a character $h_t = [\vec{h}_t; \overleftarrow{h}_t]$ is obtained by concatenating its left and right context representations. LSTM memory cell is implemented as Lample et al. [10] did.

Figure 2 The architecture of BiLSTM-CRF model.

Then a $\tanh$ layer on top of the BiLSTM is used to predict confidence scores for the character having each of the possible labels as the output scores of the network.

$$e_t = \tanh(W_e h_t), \quad (1)$$

where the weight matrix W_e is the parameter of the model to be learned in training.

Finally, instead of modeling tagging decisions independently, the CRF layer is added to decode the best tag path in all possible tag paths. We consider P to be the matrix of scores output by the network. The t^{th} column is the vector e_t obtained by the Equation (1). The element $P_{i,j}$ of the matrix is the score of the j^{th} tag of the i^{th} character in the sentence. We introduce a tagging transition matrix T , where $T_{i,j}$ represents the score of transition from tag i to tag j in successive characters and $T_{0,j}$ as the initial score for starting from tag j . This transition matrix will be trained as the parameter of model. The score of the sentence X along with a sequence of predictions $y = (y_1, \dots, y_t, \dots, y_n)$ is then given by the sum of transition scores and network scores:

$$s(X, y) = \sum_{i=1}^n (T_{y_{i-1}, y_i} + P_{i, y_i}), \quad (2)$$

Then a softmax function is used to yield the conditional probability of the path y by normalizing the above score over all possible tag paths $\tilde{y}$:

$$p(y|X) = \frac{e^{s(X, y)}}{\sum_{\tilde{y}} e^{s(X, \tilde{y})}}, \quad (3)$$

During the training phase, the objective of the model is to maximize the log-probability of the correct tag sequence. At inference time, we predict the best tag path that obtains the maximum score given by:

$$\operatorname{argmax}_{\tilde{y}} s(X, \tilde{y}). \quad (4)$$

This can be computed using dynamic programming, and the Viterbi algorithm [12] is chosen for this inference.

Figure 3 The architecture of RFPBiA model. (best viewed in color)

Technical Details of RFPBiA Model

As shown in **Figure 3**, the purpose of our work is to comprehensively model EMRs text by using the characteristics of text content and risk factors in EMRs text, thus further realizing CVD prediction task. Generally speaking, RFPBiA consists of five parts: character embedding layer, contextual embedding layer, bi-attention layer, modeling layer and predicting layer. The details are as follows.

Character Embedding Layer maps each character to a vector space using the pre-trained character vectors. Let $X = \{x_1, \dots, x_T\}$ and $W = \{word_1, \dots, word_J\}$ represent the input EMR context and risk factors in the EMR context, respectively. For risk factors, we add each character-level embedding vector that matches the word, and then average to obtain the embedding vector corresponding to risk factors. In this way, we get a new representation of risk factors $W' = \{w_1, \dots, w_J\}$.

Contextual Embedding Layer utilizes contextual cues from surrounding characters to refine the embedding of the characters. We use a Long Short-Term Memory Network (LSTM) [13] on top of the embeddings provided by the previous layers to model the temporal interactions between words. We place a LSTM in both directions, and concatenate the outputs of the two LSTMs. Hence we obtain $H \in \mathbb{R}^{2d \times T}$ from the EMR context character vectors X , and $U \in \mathbb{R}^{2d \times J}$ from risk factor vectors W' . Note that each column vector of H and U is $2d$ -dimensional because of the concatenation of the outputs of the forward and backward LSTMs, each with d -dimensional output. The first two layers apply to risk factors and the EMR context.

Bi-Attention Layer is responsible for linking and fusing information from the EMR context and the risk factors. Unlike previously popular attention mechanisms [14-16], the attention layer is not used to summarize the two kinds of information into single feature vectors. Instead, the attention vector at each time step, along with the embeddings from previous layers, are allowed to flow through to the subsequent modeling layer. This reduces the information loss

caused by early summarization.

The inputs to the layer are contextual vector representations of the EMR context H and the risk factors U . The outputs of the layer are the factor-aware vector representations of the EMR context characters, G , along with the contextual embeddings from the previous layer.

In this layer, we compute attentions in two directions: from EMR context to risk factors as well as from risk factors to EMR context. Both of these attentions, which will be discussed below, are derived from a shared similarity matrix, $S \in \mathbb{R}^{T \times J}$, between the contextual embeddings of the EMR context (H) and the risk factors (U), where S_{tj} indicates the similarity between t -th context character and j -th risk factor. The similarity matrix is computed by

$$S_{tj} = \alpha(H_{:t}, U_{:j}) \in \mathbb{R}, \quad (5)$$

where α is a trainable scalar function that encodes the similarity between its two input vectors, $H_{:t}$ is t -th column vector of H , and $U_{:j}$ is j -th column vector of U . We choose $\alpha(h, u) = w_{(s)}^\top [h; u; h \circ u]$, where $w_{(s)}^\top \in \mathbb{R}^{6d}$ is a trainable weight vector, $\circ$ is elementwise multiplication, $[;]$ is vector concatenation across row, and implicit multiplication is matrix multiplication. Now we use S to obtain the attentions and the attended vectors in both directions.

Context-to-factor Attention. Context-to-factor (C2F) attention signifies which risk factors are most relevant to each context character. Let $a_t \in \mathbb{R}^J$ represent the attention weights on the risk factors by t -th context character, $\sum_j a_{tj} = 1$ for all t . The attention weight is computed by $a_t = \text{softmax}(S_{t,:}) \in \mathbb{R}^J$, and subsequently each attended risk factor vector is $\tilde{U} = \sum_j a_{tj} U_{:j}$. Hence $\tilde{U}$ is a $2d$ -by- T matrix containing the attended risk factor vectors for the entire context.

Factor-to-context Attention. factor-to-context (F2C) attention signifies which context characters have the closest similarity to one of the risk factors and are hence critical for the predicting task. We obtain the attention weights on the context characters by $b = \text{softmax}(\max_{col}(S)) \in \mathbb{R}^T$, where the maximum function ($\max_{col}$) is performed across the column. Then the attended context vector is $\tilde{h} = \sum_t b_t H_{:t} \in \mathbb{R}^{2d}$. This vector indicates the weighted sum of the most important characters in the context with respect to the risk factors. $\tilde{h}$ is tiled T times across the column, thus giving $\tilde{H} \in \mathbb{R}^{2d \times T}$.

Finally, the contextual embeddings and the attention vectors are combined together to yield G , where each column vector can be considered as the factor-aware representation of each context character. We define G by

$$G_{:t} = \beta(H_{:t}, \tilde{U}_{:t}, \tilde{H}_{:t}) \in \mathbb{R}^{d_G}, \quad (6)$$

where $G_{:t}$ is the t -th column vector (corresponding to t -th context character), β is a trainable vector function that fuses its (three) input vectors, and d_G is the output dimension of the β function. While the β function can be an arbitrary trainable neural network, such as multi-layer perceptron, a simple concatenation as following still shows good performance in our experiments: $\beta(h, \tilde{u}, \tilde{h}) = [h; \tilde{u}; h \circ \tilde{u}; h \circ \tilde{h}] \in \mathbb{R}^{8d \times T}$ (i.e., $d_G = 8d$).

Modeling Layer employs a Recurrent Neural Network (RNN) to scan the context. The input to the modeling layer is G , which encodes the factor-aware representations of

context characters. The output of the modeling layer captures the interaction among the context characters conditioned on the risk factors. This is different from the contextual embedding layer, which captures the interaction among context characters independent of the risk factor. We use two layers of BiLSTM, with the output size of d for each direction. As a result, we take the final hidden layer states of BiLSTM (i.e., y_o) as the final output, then we redefine it as $M \in \mathbb{R}^d$, which is passed onto the predicting layer to predict the result. Here, M is exactly the ultimate representation of input EMR context and risk factors.

Predicting Layer provides a prediction to the CVD. As a result, we take the output of BiLSTM (i.e., M) as the final output. After that, we feed M into a fully-connected neural network to get an output vector $O \in \mathbb{R}^k$ (K is the number of classes):

$$O = \text{sigmoid}(M \times N), \quad (7)$$

where $N \in \mathbb{R}^{d \times k}$ is the weight matrix for dimension transformation, and $\text{sigmoid}(\cdot)$ is a non-linear activation function. Finally, we apply a softmax layer to map each value in O to conditional probability and realize the prediction as follows:

$$P = \text{argmax}(\text{softmax}(O)). \quad (8)$$

Model Training. Since what we are trying to solve is a prediction task, we follow the work in [17] to apply the cross-entropy loss function to train our model, and the goal is to minimize the following *Loss*:

$$\text{Loss} = - \sum_{T \in \text{Corpus}} \sum_{i=1}^K p_i(T) \log p_i(T), \quad (9)$$

where T is the input EMR text, *Corpus* denotes the training corpus and K is the number of classes. In the training process, we apply *Adagrad* as optimizer to update the parameters of RFPBiA, including α , N and all parameters (weights and biases) in each RNN. To avoid the overfitting problem, we apply the dropout mechanism at the end of the embedding layer.

Results

Dataset and evaluation metrics

Our dataset contains two corpora. The first corpus came from a hospital in Gansu Province with 800,000 unlabeled EMRs of internal medicine. The dataset was mainly used to train and generate our character embedding. In **Figure 4**, we also added a dictionary of risk factors during the training. In this way, the Skip-gram model in word2vec we use can better make each character in the risk factor more compact. The other one is from the Network Intelligence Research Laboratory of the Language Technology Research Center of the School of Computer Science, Harbin Institute of Technology, which contains 1186 EMRs. In the risk factor identification phase, BiLSTM-CRF model used all EMRs in the experiment, of which 830 were the training dataset, 237 were the test dataset and 119 were the development dataset. Then we will use the EMRs that need to be used for prediction to identify the risk factors through the model. This corpus intends to be used to develop a risk factor information extraction system that, in turn, can be applied as a foundation for the further study of the progress of risk factors and CVD [18]. For the corpus, we divided it into CVD and no CVD according to the clinically diagnosed disease

Figure 4 Generate the character embedding for experiments.

in the electronic medical record. In the corpus we used, there were 527 EMRs and 132 EMRs. The basis comes from the following two parts: On the one hand, according to the definition of CVD by the World Health Organization [19]. On the other hand, the first (*Symptoms*) and the third (*Diseases*) in the book of *Clinical Practical Cardiology* [20].

Our experiments involve 12 risk factors which are *Overweight/Obesity* (O2), *Hypertension*, *Diabetes*, *Dyslipidemia*, *Chronic Kidney Disease* (CKD), *Atherosclerosis*, *Obstructive Sleep Apnea Syndrome* (OSAS), *Smoking*, *Age*, *Gender*, *Alcohol Abuse* (A2), *Family History of CVD* (FHCVD), as shown in **Table 1**. In addition, the dataset also includes 4 temporal attributes which are *Continue*, *During*, *After*, *Before*. Since risk factors of Age and Gender have not consider the temporal attributes, we have added a temporal attribute: *None*. The relevant details of all EMRs are shown in **Table 2**.

In dataset, it consists of unstructured data, meaning data which is not in well-formed data. Mostly medical data is not in proper format. For the missing data, imputation and data cleaning are necessary. The unwanted data and noisy data must be removed from dataset, so that we get structured data.

In the experiment, the training dataset contains 461 EMRs, the test dataset contains 132 EMRs, and the development dataset contains 66 EMRs. Finally, we use *Accuracy* (*A*), *Precision* (*P*), *Recall* (*R*) and *F-score* (*F*) to evaluate the performance [21, 22]:

$$F = \frac{2PR}{P+R}. \quad (10)$$

Table 2 Distribution of CVD risk factors and their occurrence times

Risk factors	Before DHS	During DHS	After DHS	Continuing DHS	Total
O2	0	0	0	18	18
Hypertension	405	1909	10	1405	3729
Diabetes	60	57	13	877	1007
Dyslipidemia	4	287	6	75	372
CKD	0	0	0	26	26
Atherosclerosis	3	4	0	137	144
OSAS	0	0	0	1	1
Smoking	8	0	0	500	508
A2	9	0	0	86	95
FHCVD	0	0	0	10	10
Age	-	-	-	-	1859
Gender	-	-	-	-	1909

DHS duration of hospital stay, “-” denotes not considered.

Table 3 Hyper parameters of RFPBiA

Parameter	Description	Value
d_w	Dimension of character embedding	100
lr	Learning rate	1e-3
B	Batch size	10
d_p	Each neuron’s deactivation rate	0.5
dr	Decay rate for lr	0.99
ds	Number of decay steps	500
h	Each RNN’s hidden unit quantity	128
n	Number of epochs	28

Models and parameters

We carry out the experiments to compare the performance of our model with others described in the following.

CRF: This model was used by Mao et al. [23] to recognize the named entity in the electronic medical records based on Conditional Random Field.

BiLSTM-CRF: This model was used by Li et al. In order to realize automatic recognition and extraction of entities in unstructured medical texts, a model combining language model conditional random field algorithm (CRF) and Bi-directional Long Short-term Memory networks (BiLSTM) is proposed [24].

SVM: This model was used by S. Menaria et al. As a traditional machine learning method, the support vector machine algorithm performs well in [25].

ConvNets: This model was used by Xiang et al. [26] offering an empirical exploration on the use of character-level convolutional networks (ConvNets) for text classification.

LSTM/RFPBiA (no att): This model was used by Xin et al. [27], which proposed a LSTM network with fully connected layer and activation layers. We have further built a BiLSTM model for CVD prediction. The model also has a fully connected layer and an activation layer. In fact, this is the RFPBiA model without using the attention mechanism.

RFPBiA: This is the model proposed in this paper. **Table 3** gives the chosen hyper-parameters for all experiments. We tune the hyper-parameters on the development set by random search. We try to share as many hyper-parameters as possible in experiments.

Experimental results

We did a rich comparative experiment on our own model and other models:

In **Figure 5**, we performed a comparison of the CRF and BiLSTM-CRF models for

Figure 5 Comparison of CRF and BiLSTM-CRF models.

the identification of risk factors in EMRs. We did a lot of experiments and summarized why there was such a high F-score based on experimental and EMRs data analysis. Because there are 12 risk factors in the entire data, these risk factors are largely repeated in EMRs. In addition, BiLSTM can take contextual information into account well. This is greater for the system we have proposed.

In **Table 4**, we show the comparison between the previous model and our proposed RFPBiA model for *accuracy*, *precision*, *recall* and *F-score*. And the performance of each model when the dataset is the original EMRs, the risk factor with the label, or the risk factor without the label.

In **Table 5**, we compared four cases: (1) The performance of the ConvNets model in random embedding; (2) The performance of the LSTM model in random embedding; (3) The performance of our model without attention mechanism, that is, the performance of the BiLSTM model; (4) When our model is in random embedding.

In **Figure 6**, we have a visual example, which consists of the following three parts: (a) A sample of the case characteristics in EMR is that the risk factors have been marked according to the labeling rules [18]. (b) We translated the case characteristics into English. (c) We have made a visualization of attention matrix. The first point from the case characteristics module in EMR that we considered. It contains 8 risk factors and 61 characters.

Discussion

Table 4 shows the predictive performance of different models on our evaluation corpus. It can be seen that our RFPBiA model is superior to other methods in various evaluation indexes when pre-trained embedding and labels are used. In addition, the performance of the LSTM and BiLSTM models on the original EMRs text and the risk factor dataset,

Figure 6 a A sample of the case characteristics in EMR with marked CVD risk factors. b The sample of the case characteristics (English version). c Visualization of attention matrix for the factor-context tuple. Here we only consider the first point of case characteristics. The label on the x-axis in the figure is the characters. The label on the y-axis in the figure is the risk factor.

Table 4 The comparison of each model for CVD prediction results

Model	Accuracy%	Precision%	Recall%	F-score%
SVM (raw)	90.91	90.91	90.91	90.91
SVM (no labels)	89.39	89.03	89.39	89.21
ConvNets (raw)	92.83	92.64	92.83	92.73
ConvNets (risks with labels)	93.94	89.43	93.21	91.28
LSTM (raw)	92.24	93.46	92.73	93.09
LSTM (risks with labels)	82.58	81.35	83.01	82.17
RFPBiA (raw, no att)	93.91	93.83	93.91	93.86
RFPBiA (risks with labels, no att)	89.23	88.96	89.23	89.07
RFPBiA (no labels)	93.41	92.73	93.12	92.92
RFPBiA	94.87	93.97	94.52	94.24

“att” denotes the bi-directional attention, “no labels” denotes the risk factors without labels.

Table 5 The performance of each model at random embedding

Model	Accuracy%	Precision%	Recall%	F-score%
ConvNets (risks with labels)	91.67	90.87	91.24	91.05
LSTM (risks with labels)	81.82	79.45	82.36	80.88
RFPBiA (no att)	92.31	92.18	92.31	91.91
RFPBiA	94.22	93.57	94.61	94.09

respectively. We can find that the prediction effect is not optimistic without considering EMR text information. Compared with the ConvNets model proposed by Xiang et al., we find that the sequential model extracts EMR text information better than the model based on convolutional neural networks. Judging from the performance of our model in **Table 4** and **Table 5**, we can use the internal medicine EMRs pre-trained character embedding to help improve the performance of the model. Not only that, judging from the performance of our model in **Table 4** without the bi-directional attention mechanism, the importance of the attention mechanism to the performance of our model is shown clearly. This also shows that CVD prediction focusing on risk factors is more important. Moreover, through the presence or absence of corresponding labels of risk factors on our dataset, we can verify that the label information is beneficial to our prediction task.

For **Figure 6 (a)**, we can intuitively find that the number of risk factors present in EMR is relatively small. In other words, the key information in an EMR that can determine whether a patient has CVD is limited. For **Figure 6 (c)**, we emphasize that the learned attentions can be very useful to reduce a doctor's reading burden. It also emphasizes the correlation between the characters in EMR and the risk factors therein. For example, in this figure, the risk factor of "高血压病史 (*have a history of hypertension*)" and the risk factor of "高脂血症病史 (*have a history of hyperlipidemia*)" are clearly related to each other. This is exactly what we need, the risk factors are no longer independent of each other. We believe this can be attributed to our bi-directional attention and pre-trained character embedding. The performance comparison between the RFPBiA model and other models in **Table 4** and **Table 5** can also prove our analysis.

Conclusions

In this paper, the disease prediction experiment was carried out on the RFPBiA model using structured data. After a long-term study of EMRs, we focused on the key information leading to CVD, and we defined this key information as risk factors. So, we used BiLSTM-CRF model to identify the risk of CVD and its corresponding risk factors. what's more, we have done the next three aspects in the prediction task. On the one hand, we can use BiLSTM to obtain the context information of EMRs well. On the other hand, with the help of the bi-directional attention mechanism, we can integrate the risk factors

that leading to CVD and the original EMR information. In addition, we also pre-trained the character embedding vector with a large number of the internal medicine EMRs. It is worth mentioning that we carry temporal attribute information through labels, which enables our prediction model to consider the temporal attribute of risk factors to a certain extent. Finally, this gives us the ideal prediction effect. In the future, we will strengthen research on the pathogenic factors of CVD and improve the accuracy of CVD prediction as much as possible.

Abbreviations

CVD: Cardiovascular Disease; **RFPBiA**: the Risk Factors Powered Bi-Attention model; **LSTM**: Long Short-Term Memory; **CRF**: Conditional Random Field; **BiLSTM**: Bi-directional Long Short-term Memory networks; **BiLSTM-CRF**: bidirectional LSTM with a CRF layer; **SVM**: Support Vector Machine; **ConvNets**: Character-level Convolutional Networks; **RNN**: a Recurrent Neural Network

Declarations

Acknowledgements

We thank the anonymous reviewers for their insightful comments.

Funding

This study was supported by the National Natural Science Foundation of China (No. 61762081, No.61662067, No.61662068) and the Key Research and Development Project of Gansu Province (No.17YF1GA016).

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable requests.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Author's contributions

QY and ZZ had a part in conceiving the study, and have substantially contributed in writing and revising the manuscript. QX and TS analyzed the risk factors for disease in Chinese electronic medical records and were the main contributors to the processing of experimental data. All authors read and reviewed the final manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. Li, Y., Huo, C., Pan, T., Li, L., Jin, X., Lin, X., Chen, J., Zhang, J., Guo, Z., Xu, J., Li, X.: Systematic review regulatory principles of non-coding RNAs in cardiovascular diseases. *Briefings in Bioinformatics* **20**, 66–76 (2019)

2. Townsend, N., Nichols, M., Scarborough, P., Rayner, M.: Cardiovascular disease in Europe – epidemiological update 2015. *European Heart Journal* **36**, 2696–2705 (2015)
3. Townsend, N., Wilson, L., Bhatnagar, P., Wickramasinghe, K., Rayner, M., Nichols, M.: Corrigendum to: Cardiovascular disease in Europe: epidemiological update 2016. *European Heart Journal* **37**, 3232–3245 (2016)
4. Ng, J.Y., Hausknecht, M.J., Vijayanarasimhan, S., Vinyals, O., Monga, R., Toderici, G.: Beyond short snippets: Deep networks for video classification. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pp. 4694–4702 (2015). IEEE Computer Society
5. Liang, Z., Zhang, G., Huang, J.X., Hu, Q.V.: Deep learning for healthcare decision making with emrs. In: Zheng, H.J., Dubitzky, W., Hu, X., Hao, J., Berrar, D.P., Cho, K., Wang, Y., Gilbert, D.R. (eds.) *2014 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2014, Belfast, United Kingdom, November 2-5, 2014*, pp. 556–559 (2014). IEEE Computer Society
6. Wang, J., Ding, H., Bidgoli, F.A., Zhou, B., Iribarren, C., Molloy, S., Baldi, P.: Detecting cardiovascular disease from mammograms with deep learning. *IEEE Transactions on Medical Imaging* **36**, 1172–1181 (2017)
7. Huang, T., Shen, G., Deng, Z.: Leap-lstm: Enhancing long short-term memory for text categorization. In: Kraus, S. (ed.) *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pp. 5017–5023 (2019). ijcai.org
8. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. In: Bengio, Y., LeCun, Y. (eds.) *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings* (2013)
9. Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF models for sequence tagging. *CoRR* **abs/1508.01991** (2015).
10. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. In: Knight, K., Nenkova, A., Rambow, O. (eds.) *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pp. 260–270 (2016). The Association for Computational Linguistics
11. Ma, X., Hovy, E.H.: End-to-end sequence labeling via bi-directional lstm-cnns-crf. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers* (2016). The Association for Computer Linguistics
12. Viterbi, A.: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory* **13**, 260–269 (1967)
13. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Computation* **9**, 1735–1780 (1997)
14. Tao, H., Tong, S., Zhao, H., Xu, T., Jin, B., Liu, Q.: A radical-aware attention-based model for chinese text classification. In: *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pp. 5125–5132 (2019). AAAI Press
15. Sun, S., Sun, Q., Zhou, K., Lv, T.: Hierarchical attention prototypical networks for few-shot text classification. In: Inui, K., Jiang, J., Ng, V., Wan, X. (eds.) *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and*

- the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019, pp. 476–485 (2019). Association for Computational Linguistics
16. Wang, S., Huang, M., Deng, Z.: Densely connected CNN with multi-scale feature attention for text classification. In: Lang, J. (ed.) Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden, pp. 4468–4474 (2018). ijcai.org
 17. Zhou, Y., Xu, B., Xu, J., Yang, L., Li, C., Xu, B.: Compositional recurrent neural networks for chinese short text classification. In: 2016 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2016, Omaha, NE, USA, October 13-16, 2016, pp. 137–144 (2016). IEEE Computer Society
 18. Su, J., He, B., Guan, Y., Jiang, J., Yang, J.: Developing a cardiovascular disease risk factor annotated corpus of chinese electronic medical records. *BMC Med. Inf. & Decision Making* **17**, 117–111711 (2017)
 19. The Details of Cardiovascular Diseases (CVDs) Come from World Health Organization (WHO). 2017. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). Accessed 21 May 2020.
 20. Guo, J.: Clinical Practical Cardiology. Peking University Medical Press, New Haven (2015)
 21. Hotho, A., Nurnberger, A., Paass, G.: A brief survey of text mining. *LDV Forum* **20**, 19–62 (2005)
 22. Qiao, L., Zhao, H., Huang, X., Li, K., Chen, E.: A structure-enriched neural network for network embedding. *Expert Syst. Appl.* **117**, 300–311 (2019)
 23. Mao, X., Li, F., Duan, Y., Wang, H.: Named entity recognition of electronic medical record in ophthalmology based on crf model. In: 2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC), pp. 785–788 (2017). IEEE
 24. Li, W., Song, W., Jia, X., Yang, J., Wang, Q., Lei, Y., Huang, K., Li, J., Yang, T.: Drug specification named entity recognition base on bilstm-crf model. In: 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), vol. 2, pp. 429–433 (2019). IEEE
 25. Woldemichael, F.G., Menaria, S.: Prediction of diabetes using data mining techniques. In: 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 414–418 (2018). IEEE
 26. Zhang, X., Zhao, J.J., LeCun, Y.: Character-level convolutional networks for text classification. In: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (eds.) Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada, pp. 649–657 (2015)
 27. Hong, X., Lin, R., Yang, C., Zeng, N., Cai, C., Gou, J., Yang, J.: Predicting alzheimer’ disease using lstm. *IEEE Access* **7**, 80893–80901 (2019)

Figures

Figure 1

The main process of CVD prediction.

Figure 2

The architecture of BiLSTM-CRF model.

Figure 3

The architecture of RFPBiA model. (best viewed in color)

Figure 4

Generate the character embedding for experiments.

Figure 5

Comparison of CRF and BiLSTM-CRF models.

Figure 6

a A sample of the case characteristics in EMR with marked CVD risk factors. b The sample of the case characteristics (English version). c Visualization of attention matrix for the factor-context tuple. Here we only consider the first point of case characteristics. The label on the x-axis in the figure is the characters. The label on the y-axis in the figure is the risk factor