

RESEARCH ARTICLE

Open Access



Intelligent diagnosis with Chinese electronic medical records based on convolutional neural networks

Xiaozheng Li¹, Huazhen Wang^{1*} , Huixin He¹, Jixiang Du¹, Jian Chen² and Jinzhun Wu³

Abstract

Background: Benefiting from big data, powerful computation and new algorithmic techniques, we have been witnessing the renaissance of deep learning, particularly the combination of natural language processing (NLP) and deep neural networks. The advent of electronic medical records (EMRs) has not only changed the format of medical records but also helped users to obtain information faster. However, there are many challenges regarding researching directly using Chinese EMRs, such as low quality, huge quantity, imbalance, semi-structure and non-structure, particularly the high density of the Chinese language compared with English. Therefore, effective word segmentation, word representation and model architecture are the core technologies in the literature on Chinese EMRs.

Results: In this paper, we propose a deep learning framework to study intelligent diagnosis using Chinese EMR data, which incorporates a convolutional neural network (CNN) into an EMR classification application. The novelty of this paper is reflected in the following: (1) We construct a pediatric medical dictionary based on Chinese EMRs. (2) Word2vec adopted in word embedding is used to achieve the semantic description of the content of Chinese EMRs. (3) A fine-tuning CNN model is constructed to feed the pediatric diagnosis with Chinese EMR data. Our results on real-world pediatric Chinese EMRs demonstrate that the average accuracy and F1-score of the CNN models are up to 81%, which indicates the effectiveness of the CNN model for the classification of EMRs. Particularly, a fine-tuning one-layer CNN performs best among all CNNs, recurrent neural network (RNN) (long short-term memory, gated recurrent unit) and CNN-RNN models, and the average accuracy and F1-score are both up to 83%.

Conclusion: The CNN framework that includes word segmentation, word embedding and model training can serve as an intelligent auxiliary diagnosis tool for pediatricians. Particularly, a fine-tuning one-layer CNN performs well, which indicates that word order does not appear to have a useful effect on our Chinese EMRs.

Keywords: Chinese electronic medical records, Convolutional neural networks, Natural language processing

Background

Challenges of diagnosing using EMR data

An integrated electronic medical record system is becoming an essential part of the fabric of modern healthcare, which can collect, store, display, transmit and reproduce patient information [1, 2]. The current studies show that medical information provided by Electronic Medical Records (EMRs) is more complete and faster to retrieve than traditional paper records [3]. Nowadays, EMRs are

becoming the main source of medical information about patients [4]. The degree of health information sharing has become one of the indicators of hospital information construction in various countries. Therefore, the research and application of EMRs have certain scales and experiences in the world. How to use the rapidly growing EMR data to support biomedical research and clinical research is an important research content [5].

Due to their semi-structured and unstructured form, the study of EMRs belongs to the specific domain of Natural Language Processing (NLP). Notably, recent years have witnessed a surge of interests in data analytics with

*Correspondence: wanghuazhen@hqu.edu.cn

¹College of Computer Science and Technology, Huaqiao University, 361021 Xiamen, China

Full list of author information is available at the end of the article



patient EMRs using NLP. Ananthkrishnan et al. [6] developed a robust electronic medical record–based model for classification of inflammatory bowel disease leveraging the combination of codified data and information from clinical text notes using natural language processing. Katherine et al. [7] assessed whether a classification algorithm incorporating narrative EMR data (typed physician notes) more accurately classifies subjects with rheumatoid arthritis (RA) compared with an algorithm using codified EMR data alone. The work by Ruben et al. [8] studied a real-time electronic predictive model that identifies hospitalized heart failure (HF) patients at high risk for readmission or death, which may be valuable to clinicians and hospitals who care for these patients. Although some effective NLP methods have been proposed for EMRs, lots of challenges still remain, to list a few among the most relevant ones:

(1) *Low-Quality*. Owing to the constraint of electronic medical record template, the EMRs data are similar in a large scale, especially the content of EMRs. What's more, the medical records writing is not standardized which sometimes shows inconsistency between records and doctor's diagnosis.

(2) *Huge-Quantity*. With the increasing popularity of medical information construction, EMRs data have been growing rapidly in scale and species. There is a great intensive knowledge to explore in the EMRs databases.

(3) *Imbalance*. Due to the wide variety of diseases (e.g., there are more than 14,000 different diagnosis codes in terms of International Classification of Diseases - 9th Version (ICD-9)) in EMRs data, the sample distribution is expected to remain rather imbalance.

(4) *Semi-structure and non-structure*. The EMRs data include front sheet, progress notes, test results, medical orders, surgical records, nursing records and so on. These documents include structured information, unstructured texts and graphic image information.

Despite the above challenges, one must address the additional challenges posed by the high density of the Chinese language compared with other languages [9]. Most of words in Chinese corpus cannot be expressed independently. Therefore, the word segmentation is a necessary preprocessing step, and its effect directly affects the following series NLP operations for EMRs [10].

Intelligent diagnosis using EMR data

In practice, a great deal of information is used to determine the disease, such as the patient's chief complaint, current history, past history, relevant examinations. However, the diagnostic accuracy not only depends on individual medical knowledge but also clinical experience. Different doctors may have different diagnoses on the same patient. In particular, doctors with poor skills or in remote areas have lower diagnostic accuracy. Therefore,

it's very important and realistic to establish a intelligent diagnosis model for EMRs.

Chen et al. [11] applied machine learning methods, including support vector machine (SVM), decision forest, and a novel summed similarity measure to automatically classify the breast cancer texts on their Semantic Space models. Ekong et al. [12] proposed the use of fuzzy clustering algorithm for a clinical study on liver dysfunction symptoms. Xu et al. [13] designed and implemented a medical information text classification system based on a KNN. Many researchers at home and abroad, who use EMRs for disease prediction, always focus on a particular department as well as a specific disease. At present, the algorithms used by researchers mostly focus on machine learning methods, such as KNN, SVM, DT. Due to the particularity of medical field and the key role of professional medical knowledge, common text classification methods often fail to achieve good classification performance and cannot meet the requirement of clinical practice [14].

Benefiting from big data, powerful computation and new algorithmic techniques, we have been witnessing the renaissance of deep learning, especially the combination of natural language processing and deep neural networks. Dong et al. [15] presented a CNN based multiclass classification method for mining named entities with EMRs. A transfer bi-directional Recurrent Neural Networks was proposed for named entity recognition (NER) in Chinese EMRs that aims to extract medical knowledge such as phrases recording diseases and treatments automatically [16]. SA [17] marked the prediction of heart disease as a multi-level problem of different features or signs and constructed an IHDPS (Intelligent Heart Disease Prediction System) based on neural networks.

However, to the best of our knowledge, few significant models based on deep learning have been employed for the intelligent diagnosis with Chinese EMRs. Rajkomar et al. [18] demonstrated that deep learning methods outperformed state-of-art traditional predictive models in all cases with electronic health record (EHR) data, which is probably the first research on using deep learning methods in EHR model analysis.

Deep learning for natural language processing

NLP is a theory-motivated range of computational techniques for the automatic analysis and representation of human language, which enables computers to perform a variety of natural language related tasks at all levels, ranging from parsing and part-of-speech (POS) tagging, to dialog systems and machine translation. In recent years, Deep learning algorithms and architectures have already won numerous contests in fields such as computer vision and pattern recognition. Following this trend, recent NLP research is now increasingly focusing on the use of deep learning methods [19].

In a deep learning with NLP model, word embedding is usually used as the first data preprocessing layer. It's because the learnt word vectors can capture general semantic and syntactical information, that word embedding produces state-of-art results on various NLP tasks [20–22]. Following the success of word embedding [23, 24], CNNs turned out to be the natural choice in view of their effectiveness in computer vision and pattern recognition tasks [25–27]. In 2014, Kim [28] explored using the CNNs for various sentence classification tasks, and CNNs was quickly adapted by some researchers due to its simple and effective network. Poria et al. [29] proposed a multi-level deep CNN to tag each word in a sentence, which coupled with a group of linguistic patterns and finally performed well in aspect detection.

Besides text classification, CNN models are also suitable for other NLP tasks. For example, Denil et al. [30] applied DCNN to map meanings of words that constitute a sentence to that of documents for summarization, which provided insights in automatic summarization of texts and the learning process. In the domain of Question and Answer (QA), the work by Yih et al. [31] presented a CNN architecture to measure the semantic similarity between a question and entries in a knowledge base (KB), which determined what supporting fact in the KB to look for when answering a question. In the domain of Information and Retrieval (IR), Chen et al. [32] proposed a dynamic multi-pooling CNN (DMCNN) strategy to overcome the loss of information for multiple-event modeling. In the speech recognition, Palaz et al. [33] performed extensive analysis based on a speech recognition systems with CNN framework and finally created a robust automatic speech recognition system. In general, CNNs are extremely effective in mining semantic clues in contextual windows.

It is well known that pediatric patients are generally depauperate, traversing from newborns to adolescents. Correspondingly, the treatment and dosage of medicine are different from those given to adult patients. Thus, it is a great challenge to build a prediction model for pediatric diagnosis that is trained to “learn” expert medical knowledge to simulate the doctor’s thinking and diagnostic reasoning.

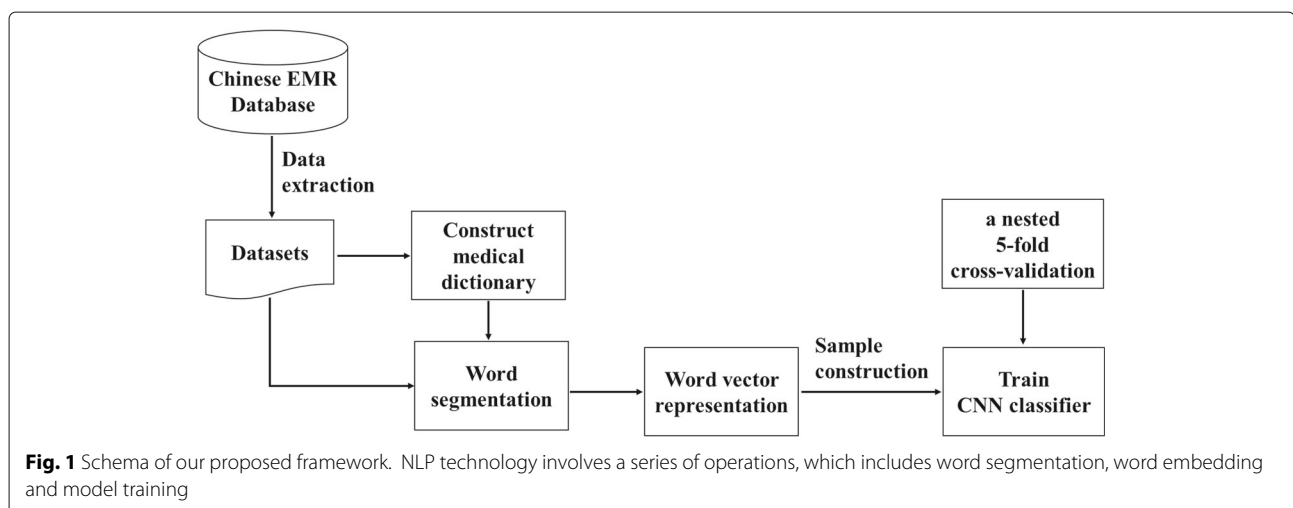
In this research, we propose a deep learning framework to study intelligent diagnosis using Chinese EMRs, which incorporates a convolutional neural network (CNN) into an EMR classification application. This framework involves a series of operations that includes word segmentation, word embedding and model training. In real pediatric Chinese EMR intelligent diagnosis applications, the proposed model has high accuracy and a high F1-score, and achieves good results. The novelty of this paper is reflected in the following:

- (1) We construct a pediatric medical dictionary based on Chinese EMRs.
- (2) Word2vec is used as a word embedding method to achieve the semantic description of the content of Chinese EMRs.
- (3) A fine-tuning CNN model is constructed to feed the pediatric diagnosis with Chinese EMR data.

Methods

Proposed framework

Our proposed framework is the incorporation of a CNN into the procedure of NLP with Chinese EMRs, and its schema is shown in Fig. 1, which includes word segmentation, word embedding and model training. First, the corpus is extracted from the Chinese EMR database. Then, a medical dictionary is constructed from the original corpus, which is used as external expert knowledge in word segmentation. Next, word embedding is executed. Finally,



word	The absence of our medical dictionary	The adoption of our medical dictionary
skin color (皮肤色泽)	skin\color (皮肤\色泽)	skin color (皮肤色泽)
three concave sign (三凹征)	three\concave sign (三\凹征)	three concave sign (三凹征)
bowel sound (肠鸣音)	bowel\sound (肠\鸣音)	bowel sound (肠鸣音)
knee reflex (膝反射)	knee\reflex (膝\反射)	knee reflex (膝反射)
no pericardial friction rub (无心包摩擦音)	no\pericardial friction\rub (无\心包\摩擦音)	no pericardial friction rub (无心包摩擦音)
muscle strength (肌张力)	muscle\strength (肌\张力)	muscle strength (肌张力)

Fig. 2 Semantic rationality of whether to use our medical dictionary

an optimal one-layer CNN. The relationship between accuracy and dimension is shown in Table 2.

Consider the seven-classification application as an example. Each word is embedded into 50-dimensional vector space. For instance, the word “cough” is expressed as [-3.982, -0.670, -1.754, , 3.048]₅₀ and the word “fever” is expressed as [-4.487, -5.976, -5.417, , 1.216]₅₀. Additionally, the word vector representation using word2vec can use the cosine distance to measure the degree of semantic similarity [10]. The cosine distance of words between “cough” are given in Table 3, which indicates that the smaller the cosine value, the more similar the semantics.

Convolutional neural networks

CNNs proposed by Lecun in 1989 [44] enable automatic feature representation learning. Different from the traditional feed-forward neural network, a CNN is a multi-layer neural network that includes four parts, embedding layer, convolution layer, pooling layer and fully connected layer, as illustrated in Fig. 3 [45].

The first layer is the input layer, which is an embedding matrix $I \in \mathbb{R}^{S \times N}$ that corresponds to the symptom text to be classified. Number of rows S is the number of words in the sentence and number of columns N is the dimension of the word vector. Consider the description of “cough for a week, a mild headache and runny nose” as an example. The sentence is divided into “cough + a + week + a mild + headache + runny nose” when the dictionary-based word segmentation method is used. Then each word is converted into a vector using word2vec, subsequently

Table 2 One-layer CNN accuracy for different dimensions with respect to four types of classification applications

Text classification	50 (%)	80 (%)	100 (%)
7 classes	83.72	83.65	83.63
8 classes	82.26	82.55	82.51
32 classes	73.13	73.44	73.54
63 classes	70.39	71.06	71.2

Boldface represents the best

forming embedding matrix I as the input layer of the CNN [45].

Then different filters are applied to different layers and the result is downsampled using the pooling layer. CNNs realize automatic feature representation learning through multiple layers of networks, the core of which lies in the convolutional layer and pooling layer. The convolution layer extracts local features, whereas the pooling layer reduces the dimension of the structured feature [46, 47].

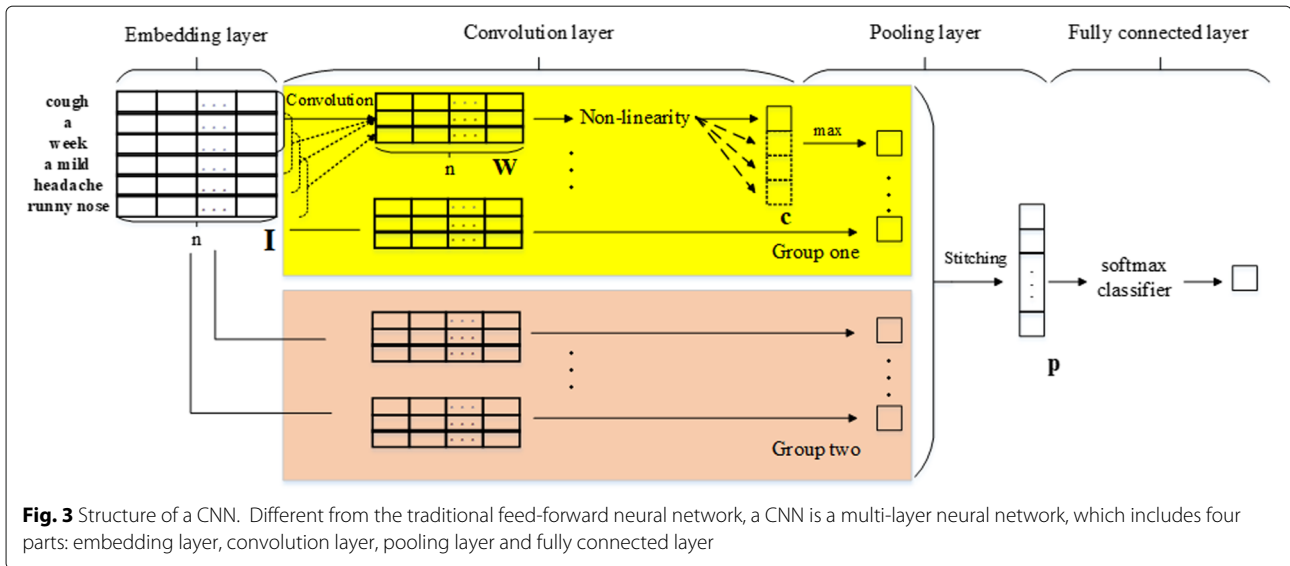
Additionally, the depth of neural networks plays a decisive role in the performance of a CNN model, and is regarded as one of the most investigated approaches used to increase its accuracy. For instance, Wang et al. [48] discussed the influence of the varied depth on the validation set of ILSVRC and proposed that “going deeper” is an effective and competitive approach to increase the accuracy of classification. The work by Hussam et al. [49] proposed a deep neural network comprised of 16 convolutional layers compressed with the Fire module adapted from the SqueezeNet model.

Hyperparameter setup

The architecture of CNN needs fine-tuning to obtain optimal performance on specific datasets. Generally, hyperparameter setup refers to the grid-search of several parameters, which include size of filter windows, number of feature

Table 3 Semantic similarity of word vectors

Word	Cosine distance
Recurrent cough	0.6350
Quiet cough	0.6196
Bad cough	0.5433
Little cough	0.5204
Dry cough	0.5208
Nasal obstruction	0.5914
Phlegm	0.5434
Vomiting	29.48
Afternoon	23.41
Muscular stiffness	22.83



maps, dropout rate, activation function, mini-batch size, and so on [28]. Practically, the hyperparameter setup of CNN refers the filter windows of 7, 6, 5, 4 and 3, the feature maps of 128, 100, 64, 50, 32 and 16, the mini-batch size of 100, 95, 64, 50 and 32. In our experiments, a nested 5-fold cross-validation approach was applied on the seven-classification dataset, where the inner cross-validation was used for the grid-search to tune the hyperparameters, and the outer cross-validation was adopted to evaluate the performance of different models mentioned in this paper. As a result, we found that the one-layer CNN outperformed on the EMR-based pediatric diagnosis, whose hyperparameters included the filter windows of 7, the feature maps of 100, the dropout rate of 0.5, activation of relu and mini-batch size of 64, and the update rule of AdaMax. All the experiments were conducted using Python 3.5 with Python packages.

Results

Evaluation

In this paper, we study the effectiveness of our proposed framework on real-world pediatric Chinese EMR data. For each dataset, three metrics were used to evaluate the effectiveness and performance of algorithms: accuracy, precision and F1-score. Precision and recall were often combined to obtain a better understanding of the performance of the classifier. Their formulas for calculation are as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

where

true positive (TP): scenario in text classification in which the classifier correctly classifies a positive test case into a positive class;

true negative (TN): scenario in text classification in which the classifier correctly classifies a negative test case into a negative class;

false positive (FP): scenario in text classification in which the classifier incorrectly classifies a negative test case into a positive class;

false negative (FN): scenario in text classification in which the classifier incorrectly classifies a positive test case into a negative class.

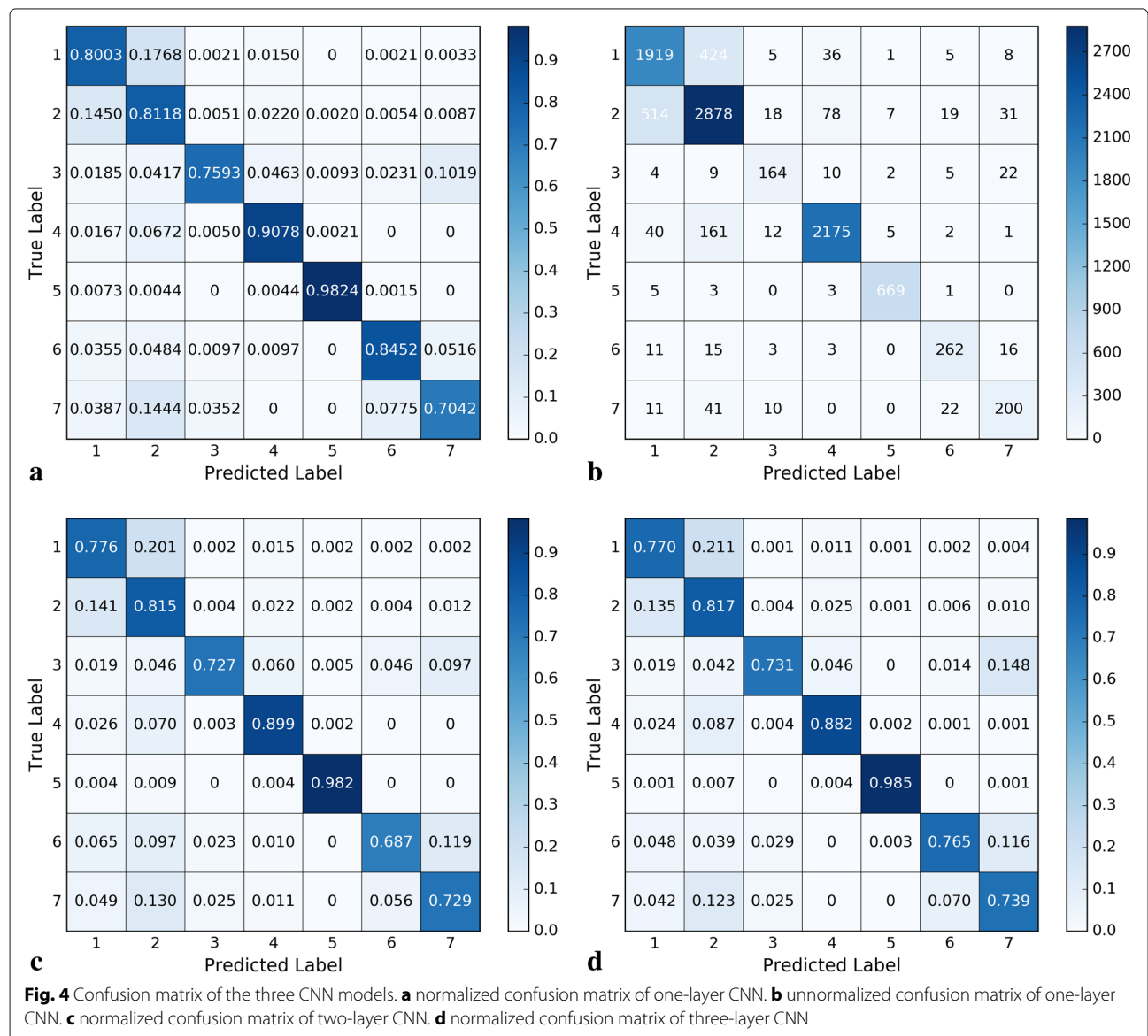
Performance of the CNN models

In the CNN experiments, we focused on the impact of depth on our application, that is, three different depths, depth 1, depth 2 and depth 3, were explored to obtain an optimal solution. Subsequently, the comparative results with respect to the seven-classification application are presented in Table 4, which contains the precision, accuracy and F1-score of each fold.

It can be seen from Table 4 that the accuracies of the three CNN models were all higher than 81%, and the same is true for other metrics. This result indicates the effectiveness of CNN for the classification of Chinese EMRs. Furthermore, one-layer CNN had the best performance among all the CNN models, which makes it the most

Table 4 Comparative results of the CNN model with the seven-classification application

Depth Fold\metrics	One-layer CNN(%)			Two-layer CNN(%)			Three-layer CNN(%)		
	Precision	Accuracy	F1-score	Precision	Accuracy	F1-score	Precision	Accuracy	F1-score
1	84.26	84.1	84.16	83.13	82.9	82.97	83.05	82.74	82.84
2	83.63	83.5	83.52	82.65	82.42	82.5	82.32	81.53	81.66
3	83.86	83.55	83.61	82.54	82.26	82.35	79.09	78.89	78.94
4	84.07	83.75	83.84	82.78	82.51	82.58	82.28	82.02	82.05
5	83.87	83.71	83.76	82.97	82.81	82.85	82.6	82.37	82.4
Average	83.94	83.72	83.78	82.81	82.58	82.65	81.87	81.51	81.58



practicable tool in pediatric diagnosis. Because the experimental datasets were more than two classes and imbalanced, the confusion matrix of the three CNN models are shown in Fig. 4, where Fig. 4a and b show the first-fold normalized confusion matrix and its non-normalized confusion matrix for the one-layer CNN model in the outer 5-fold cross-validation, respectively. The first-fold normalized confusion matrix of the two-layer CNN model and three-layer CNN model can be observed in Fig. 4c and d, respectively.

CNN vs. RNN models

The results of our CNN models against other methods are presented in Table 5. The model of long short-term memory (LSTM) did not perform well. The average accuracy and F1-score of the CNN models are up to 81%, which indicates the effectiveness of the CNN model for the classification of EMRs. Particularly, a fine-tuning one-layer CNN performs best among all CNN, recurrent neural network (RNN) (LSTM, gated recurrent unit (GRU)) and CNN-RNN models, and the average accuracy and F1-score are both up to 83%.

Based on the best CNN model architecture (one-layer CNN), the other classification applications, i.e., eight-classification application, 32-classification application, and 63-classification application, were evaluated by the 5-fold cross-validation. Table 6 shows the model accuracies of four types of pediatric diagnosis applications. It can be seen that (1) the highest accuracy was exhibited in the seven-classification application, which may have been caused by the small scale and somewhat balanced distribution of sample data; and (2) with the increase of disease types, the accuracy of the one-layer CNN model decreased. The main reason was that, because of the constraint of the EMR template, the content of the EMRs were similar on a large scale. Furthermore, there were not

Table 5 Results of our CNN models against other methods

Model	Precision(%)	Accuracy(%)	F1-score(%)
1-layer CNN	83.94	83.72	83.78
1-layer LSTM	43.97	46.33	38.18
1-layer GRU	82.95	82.2	82.37
2-layers CNN	82.81	82.58	82.65
2-layers LSTM	23.01	34.12	19.57
2-layers GRU	83.03	82.4	82.57
3-layers CNN	81.87	81.51	81.58
CNN-1LSTM	83.86	83.55	83.62
CNN-2LSTM	83.63	83.18	83.33
CNN-1GRU	83.42	83.02	83.13
CNN-2GRU	83.52	82.95	83.1

Boldface represents the best

Table 6 Accuracies of fine-tuning the one-layer CNN model with respect to four types of classification applications

The number of diseases	precision(%)	accuracy(%)	F1-score(%)
7 classes	83.94	83.72	83.78
8 classes	82.35	82.55	82.27
32 classes	73.09	73.54	72.5
63 classes	70.59	71.2	69.61

Boldface represents the best

sufficient samples to train for so many different types of diseases.

Discussion

Impact of the Chinese medical dictionary on word segmentation

With the dictionary-based word segmentation method incorporating our pediatric medical dictionary, the corpus can be separated by "\". Fig. 2 shows the semantic rationality of whether to use our medical dictionary. The second column shows the segmentation result with the absence of our medical dictionary and the third column shows the segmentation result with the adoption of our medical dictionary. This shows that adopting the medical dictionary as expert knowledge accurately divided the corpus into the smallest semantic independent medical expressions, which was very helpful for the subsequent model construction.

Impact of various example constructions

A typical medical record always contains a set of entries, such as *age*, *gender*, *current status*, *chief complaint*, *present history*, *previous history*, *family history*, *physical examination and diagnosis*. An example of a medical record from the pediatric Chinese EMRs is shown in Fig. 5.

Based on Fig. 5, the entry of *age*, *gender*, *current status*, *chief complaint*, *present history*, *previous history*, *family history* and *physical examination* are designated as the corpus, and *the initial diagnosis* is designated as the label.

When applying a CNN model, it is necessary to convert a medical record corpus into a fixed-size matrix. Considering the seven-classification application as an example, the corpus shown in Fig. 5 should be converted into a 120×50 matrix for training, and the number of words in each corpus is regularized to be 120 and the vector dimension of each word is 50. However, because the length of different medical records is different, that is, the number of words in the shortest corpus is 21 and the number of words in the longest corpus is 271, a corpus that contains records of various lengths should be truncated or filled to make the records even. If the shortest medical record is chosen as the regularized length, then important information in a longer corpus may be truncated. Conversely,

title	English content	Chinese content
Age	one year old	一岁
Gender	female	女
Current status	in general	一般
Chief complaint	runny nose with low fever for one day	流涕伴低烧一天
Present history	1 day ago there was a clear runny nose, sneezing with fever, the highest temperature 37.9 °C, no cough, no vomiting	1天前流涕,打喷嚏并发烧,体温最高37.9度,无咳嗽,无呕吐
Previous history	no special	无特殊
Family history	no clear history of allergies	无明确过敏史
Physical examination	sanity, respiration is smooth, reaction is normal, the whole body skin and mucous membrane are without yellow dye, no rash, no bleeding and ecchymosis. The whole body superficial lymph nodes did not touch swollen. Bilateral pupils are equal and round, which are sensitive to light reflection. No agitation of the nose, no cyanosis of lips, pharynx congestion. Three concave sign is negative. The two lungs have rough breathing sounds. Rhonchus and moist rales could not be heard. Heart sound is strong, regular rate, no murmurs. Stomach is soft and cannot touch block. Liver and spleen ribs don't touch swollen. Muscular strength and muscular ton are normal. Pathological sign is negative.	神清,气平,反应可,全身皮肤、粘膜无黄染,无皮疹,无出血点及瘀斑,全身浅表淋巴结未触及肿大。瞳孔等大等圆,对光反射存在。口唇无发绀,咽充血,三凹征阴性。双肺呼吸音粗,未闻及干湿性罗音,心音有力,心律齐,未闻及杂音。腹平软,无压痛,未触及包块,肝脾肋下未触及肿大,四肢肌张力正常,病理征阴性。
Initial diagnosis	acute upper respiratory tract infections	急性上呼吸道感染

Fig. 5 Description of a typical pediatric Chinese EMR datum

choosing the length of the longest medical record can add too many unwanted messages (fill 0) to a shorter corpus, and increase the complex of model training.

Therefore, we attempted to explore how three types of setup, that is, a regularized length of corpus, the truncation approach and the filling mode of the medical record, affect the performance of the CNN model. For the parameter of a regularized length, we attempted 90, 100, 110, 120, 130 and 140; for the parameter of the filling

mode, we considered two alternatives, that is, head-filling and tail-filling; and for the parameter of the truncation approach, we also considered two candidates, that is, head-truncation and tail-truncation. Thus, a grid-search method was adopted to determine an optimal parameter setup for the aforementioned best performing CNN model (one-layer CNN).

Because of the limited length of this paper, the performance of the seven-classification CNN model is

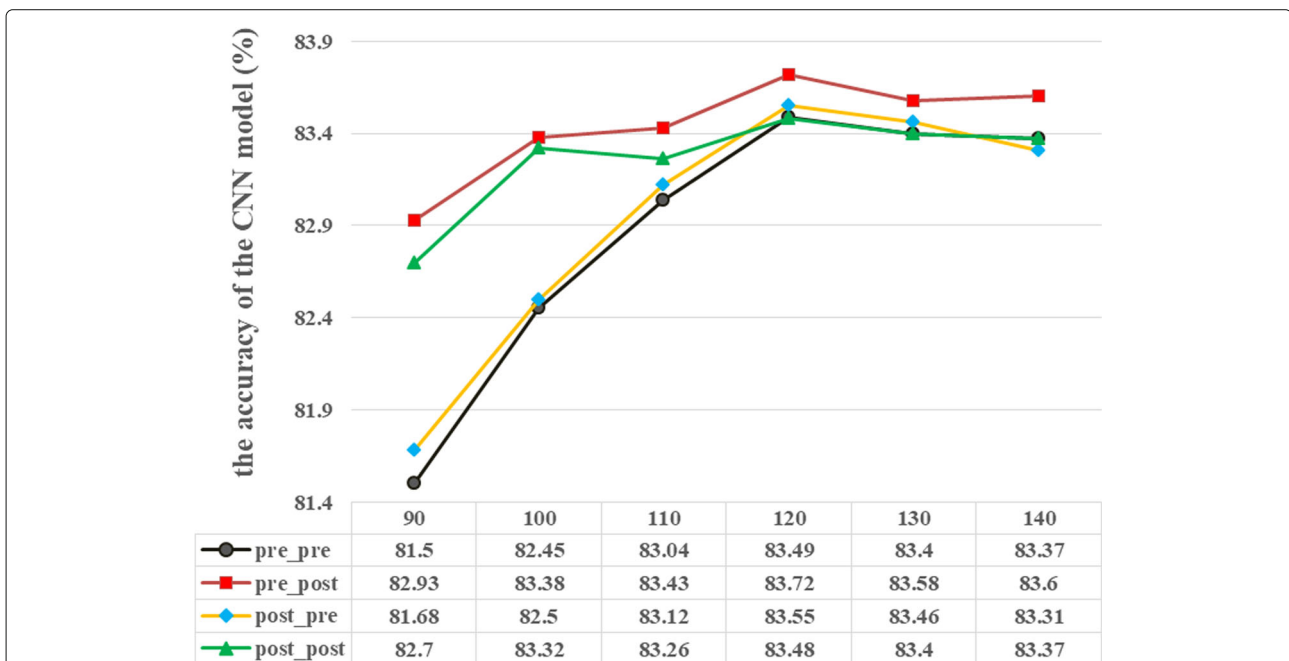


Fig. 6 Impact of three types of parameter on the accuracy of the CNN model. Note: “pre” refers to head-filling or head-truncation and “post” refers to tail-filling or tail-truncation. For example, “pre_post” means that short text is filled by head and long text is truncated by tail

Table 7 Comparative accuracies with respect to the seven-classification application and the eight-classification application of whether to use class weights

Class\metrics	Name of class	Sample size	Seven-classification		Eight-classification	
			Without class weight	With class weight	Without class weight	With class weight
Class1	Allergic rhinitis	1079	71.09	80.1	59.68	77.85
Class2	Respiratory disease	11980	90.37	87.92	85.28	86.3
Class3	Cough variant asthma	1418	70.31	80.74	67.12	81.45
Class4	Acute bronchitis	11990	77.5	80.00	65.56	81.78
Class5	Bronchialasthma, no critical	1550	79.23	83.56	78.82	80.77
Class6	Bronchitis	17726	82.79	73.42	66.94	51.42
Class7	Diarrhea	3405	97.91	98.7	94.9	97.06
Class8	Acute upper respiratory tract infection	43596	NA	NA	92.94	84.11

Boldface represents the best

illustrated in Fig. 6. The results of other classification applications were similar to those of Fig. 6. From Fig. 6, we can see that the model had very robust superiority for the configuration that had the corpus length of 120, in addition to using head-filling for shorter text and tail-truncation for the longer text, which indicates that head information for longer medical records is more important than tail information, and head-filling for shorter medical records is better than tail-filling. Therefore, for this optimal configuration, that is, where the regularized length of the corpus is 120, a head-filling mode and a tail-truncation approach for the medical record were adopted in our application.

Impact of the class weights in training

In order to improve the class accuracy of small-number class caused by the unbalance distribution, different class weights serves as error-recognition penalty were introduced.

$$class_weights = \frac{n_samples}{n_classes * n_class_samples} \quad (5)$$

where $n_samples$ is the number of samples, $n_classes$ is the class number of samples and $n_class_samples$ is the sample number of one class.

Table 8 Comparative results with respect to the seven-classification application and the eight-classification application of whether to use different class weights

Metrics	Seven-classification		Eight-classification	
	Without class weight	With class weight	Without class weight	With class weight
Precision (%)	83.94	82.27	82.35	80.97
Accuracy (%)	83.72	80.99	82.55	78.15
F1-score (%)	83.78	81.25	82.27	78.45

Boldface represents the best

Based on the best CNN model architecture (one-layer CNN), Table 7 shows the comparative accuracies of each class with respect to the seven-classification application and the eight-classification application, and Table 8 shows the three model evaluation indices. It can be seen that: (1) the class accuracy of small number of samples has promoted a lot when using class weights, at the same time, the class accuracy of large sample size has put down a lot; and (2) In a comprehensive view, it performs well in all three metrics than using the class weights. Therefore, we do not use class weights in our article.

Conclusions

Considering the advantage of CNNs in local feature extraction and modeling performance, we attempted to explore a framework based on a CNN model for intelligent diagnosis with pediatric Chinese EMRs. Our framework was composed of three parts: word segmentation, word embedding and model training. With an expert dictionary based on collected Chinese EMR data used in word segmentation, and the word vector representation of the medical records using word2vec, we validated the effectiveness of our proposed framework on real-world EMR data. A wide range of models, which included CNN models, RNN models (LSTM, GRU) and CNN-RNN hybrid architecture, were explored to determine an optimal model. The comparative experimental results indicate the effectiveness of the CNN model for the classification of Chinese EMR data, which indicates that word order does not appear to have a useful effect on our Chinese EMRs. Furthermore, one-layer CNN performed best among all the classification applications. To conclude, the one-layer CNN model might contribute to the diagnosis of pediatric Chinese EMRs.

In this study, we only used EMR data and did not integrate medical images into the model. Therefore, future research will focus on how to integrate multiple types of

medical information to improve the prediction effect for pediatric Chinese EMRs.

Additional file

Additional file 1: Distribution of datasets with respect to four types of classification applications for pediatric Chinese EMRs. (PDF 142 kb)

Abbreviations

CNN: Convolutional neural network; DMCNN: Multi-pooling CNN; EHR: Electronic health record; EMRs: Electronic medicine records; GRU: Gated recurrent unit; HAL: Hyperspace analog to language; HF: Heart failure; IHDPs: Intelligent heart disease prediction system; KB: Knowledge base; LSTM: Long short-term memory; MCCNN: Multi-column CNN; NER: Named entity recognition; NLP: Natural language processing; QA: Question and answer; RNN: Recurrent neural networks

Acknowledgements

We thank Professor Bicheng Li for his helpful guidance during the writing of the manuscript.

Funding

This work is partially supported by National Natural Science Foundation of China under Grant No. 61673186, the Natural Science Foundation of Fujian Province in China under Grant No. 2012J01274. The funders did not play any roles in the design of the study, in the collection, analysis, or interpretation of data, or in writing the manuscript.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available, because all EMR data are from the First Affiliated Hospital of Xiamen University and the constraints of the privacy policies, but they are available from the corresponding author on reasonable request.

Authors' contributions

HZW, HXH and JZW conceived the study. XZL completed experimental and wrote the initial draft of the manuscript. HZW, HXH, JXD, JZW and JC gave their helpful guidance during the analysis and writing of the manuscript. All authors contributed to analysing the data, writing and revising the manuscript. All authors read and approved the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹College of Computer Science and Technology, Huaqiao University, 361021 Xiamen, China. ²Research Department, Zhiye software, 361021 Xiamen, China. ³Pediatric Department, The First Affiliated Hospital of Xiamen University, 361003 Xiamen, China.

Received: 17 March 2018 Accepted: 10 January 2019

Published online: 01 February 2019

References

1. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. *BMC Health Serv Res.* 2010;10(1):231.

2. MacKinnon W, Wasserman M. Integrated electronic medical record systems: Critical success factors for implementation. 2009 42nd Hawaii Int Conf Syst Sci. 2009;1–10.
3. Tsai J, Bond GG. A comparison of electronic records to paper records in mental health centers. *Int J Qual Health Care J Int Soc Qual Health Care.* 2008;20(2):136–43.
4. Hu Y. Research on the information diagnostic technology based on medical information. University of Electronic Science and Technology of China. 2015.
5. Yang J, Guan Y, He B, Qu C, Yu Q, Liu Y, Zhao Y. Corpus construction for named entities and entity relations on chinese electronic medical records. *J Softw.* 2016;27(11):2725–46.
6. Ananthkrishnan AN, Cai T, Savova G, Cheng S-C, Chen P, Perez RG, Gainer VS, Murphy SN, Szolovits P, Xia Z, Shaw S, Churchill S, Karlson EW, Kohane I, Plenge RM, Liao KP. Improving case definition of crohn's disease and ulcerative colitis in electronic medical records using natural language processing a novel informatics approach. *Inflamm Bowel Dis.* 2013;19(7):1411–20.
7. Liao KP, Cai T, Gainer VS, Goryachev S, Zeng-Treitler Q, Raychaudhuri S, Szolovits P, Churchill SE, Murphy SP, Kohane IS, Karlson EW, Plenge RM. Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res.* 2010;62 8:1120–7.
8. Amarasingham R, Moore B, Tabak YP, Drazner MH, Clark CA, Zhang S, Reed W, Swanson TS, Ma Y, Halm EA. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. *Med Care.* 2010;48(11):981–8.
9. Hoosain R. Psycholinguistic implications for linguistic relativity: A case study of chinese. *J Neurolinguistics.* 1991;8(2):157–61.
10. Zhao M, Du H, Dong C, Chen C. Diet health text classification based on word2vec and lstm. *Trans Chin Soc Agric Mach.* 2017;48(10):202–8.
11. Chen G, Warren J, Riddle P. Semantic space models for classification of consumer webpages on metadata attributes. *J Biomed Inform.* 2010;43(5):725–35.
12. Ekong VE, Onibere EA, Imianvan AA. Fuzzy cluster means system for the diagnosis of liver diseases. *J Comput Sci Technol.* 2011;2(3):205–9.
13. Xu X, Zhang Q. Research of medical information text categorization based on knn algorithm. *Comput Technol Dev.* 2009;19(4):206–209.
14. Cao J. A text classifier about high blood pressure based on naive bayes. Taiyuan University of Technology; 2015.
15. Dong X, Qian L, Guan Y, Huang L, Yu Q, Yang J. A multiclass classification method based on deep learning for named entity recognition in electronic medical records. 2016 N Y Sci Data Summit (NYSDS). 2016;1–10.
16. Dong X, Chowdhury S, Qian L, Guan Y, Yang J, Yu Q. Transfer bi-directional lstm rnn for named entity recognition in chinese electronic medical records. 2017 IEEE 19th Int Conf e-Health Netw Appl Serv (Healthcom). 2017;1–4.
17. Sanap SA. Intelligent heart disease prediction system using data mining techniques. *International Journal of Healthcare & Biomedical Research.* 2013;94–101.
18. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Liu PJ, Liu X, Sun M, Sundberg P, Yee H, Zhang K, Duggan GE, Flores G,hardt M, Irvine J, Le QV, Litsch K, Marcus J, Mossin A, Tansuwan J, Wang D, Wexler J, Wilson J, Ludwig D, Volchenboum SL, Chou K, Pearson M, Madabushi S, Shah NH, Butte AJ, Howell M, Cui C, Corrado GS, Dean J. Scalable and accurate deep learning for electronic health records. *npj Digital Medicine.* 2018;1(1):18.
19. Younga T, Hazarikab D, Poriac S, Cambriad E. Recent trends in deep learning based natural language processing. *IEEE Comput Intell Mag.* 2018;13(3):55–75.
20. Weston J, Bengio S, Usunier N. Wsabie: Scaling up to large vocabulary image annotation. *IJCAI.* 2011:2764–2770.
21. Socher R, Lin CC-Y, Ng AY, Manning CD. Parsing natural scenes and natural language with recursive neural networks. In: *ICML.* 2011. p. 129–136.
22. Turney PD, Pantel P. From frequency to meaning: Vector space models of semantics. *J Artif Intell Res.* 2010;37:141–88.
23. Mikolov T, Karafiat M, Burget L, Cernocky J, Khudanpur S. Recurrent neural network based language model. In: *Eleventh Annual Conference of the International Speech Communication Association.* 2010.
24. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. *Adv Neural Inf Proces Syst.* 2013:3111–3119.

25. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems. 2012. p. 1097–105.
26. Razavian AS, Azizpour H, Sullivan J, Carlsson S. Cnn features off-the-shelf: An astounding baseline for recognition. 2014 IEEE Conf Comput Vis Pattern Recognit Workshops. 2014. p. 512–519.
27. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick RB, Guadarrama S, Darrell T. Caffe: Convolutional architecture for fast feature embedding. *ACM Multimedia*. 2014;675–678.
28. Kim Y. Convolutional neural networks for sentence classification. In: Eprint Arxiv. 2014.
29. Poria S, Cambria E, Gelbukh AF. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowl-Based Syst*. 2016;108:42–9.
30. Denil M, Demiraj A, Kalchbrenner N, Blunsom P, de Freitas N. Modelling, visualising and summarising documents with a single convolutional neural network. *Computer Science*. 2014. abs/1406.3830.
31. Yih W-t, He X, Meek C. Semantic parsing for single-relation question answering. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. 2014:643–648.
32. Chen Y, Xu L, Liu K, Zeng D, Zhao J. Event extraction via dynamic multi-pooling convolutional neural networks. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015. p. 167–176.
33. Palaz D, Magimai-Doss M, Collobert R. Analysis of cnn-based speech recognition system using raw speech as input. In: Sixteenth Annual Conference of the International Speech Communication Association. 2015.
34. Guo T. Research on automatic segmentation based on dictionary: Harbin University of Science and Technology; 2010.
35. Xiong H, Xia L. The review of chinese automatic word segmentation technology. *Libr Inf Serv*. 2008;52(4):81–4.
36. Xu D, Zhang M, Zhao T, Ge C, Gao W, Wei J, Zhu KQ. Data-driven information extraction from chinese electronic medical records. *Plos ONE*. 2015;10(8):e0136270.
37. HUang R, Zhang W. Study on sentiment analyzing of internet commodities review based on word2vec. *Comput Sci*. 2016;43(s1):387–9.
38. Xiong F, Deng Y, Tang X. The architecture of word2vec and its applications. *J Nanjing Normal Univ*. 2015;1:43–48.
39. Gutmann M, Hyvarinen A. Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *J Mach Learn Res*. 2012;13:307–61.
40. Mnih A, Teh YW. A fast and simple algorithm for training neural probabilistic language models. In: ICML. 2012.
41. Morin F, Bengio Y. Hierarchical probabilistic neural network language model. In: *AIStats*. 2005. p. 246–252.
42. Rumelhart DE, McClelland JL, Group TP. Parallel distributed processing: Explorations in the microstructures of cognition. *Language*. 1986;63(4).
43. Mikolov T, Kopecky J, Burget L, Glembek O, Cernocky J. Neural network based language models for highly inflective languages. 2009 IEEE Int Conf Acoust Speech Signal Process. 2009. p. 4725–8.
44. LeCun Y. Gradient-based learning applied to document recognition. In: Proceedings of the IEEE ,vol. 86. 1998. p. 4725–4728.
45. Liu Z, Wang H, Cao J, Qiu J. Power equipment defect text classification model based on convolutional neural network. *Power Syst Technol*. 2018;2:644–650.
46. Liu X, Zhang Y, Zheng Q. Sentiment classification of short texts on internet based on convolutional neural networks model. *Computer & Modernization*. 2017;4:73–77.
47. Yu B, Zhang L. Chinese short text classification based on cp-cnn. *Appl Res Lang Comput*. 2018;35(4):1001–1004.
48. Wang L, Lee C-Y, Tu Z, Lazebnik S. Training deeper convolutional networks with deep supervision. Cornell University. 2015. p. abs/1505.02496.
49. Qassim H, Feinzimer D, Verma A. Residual squeeze vgg16. Cornell University. 2017. p. abs/1705.03004.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

