TCM Function Multi-classification Approach Using Deep Learning Models

Quanying Ren[1], Keqian Li[2], Dongshen Yang[1], Yan Zhu[3],
Keyu Yao[3], and Xiangfu Meng[1(✉)]

[1] College of Electronic and Information, Engineering Liaoning Technical University,
Huludao, China
`marxi@126.com`
[2] School of Medical Information, Changchun University of Chinese Medicine,
Changchun, China
[3] Information Institute of Traditional Chinese Medicine,
Chinese Academy of Traditional Chinese Medicine, Beijing, China

Abstract. Traditional Chinese Medicine prescriptions are regarded as an important resource that brings together the treatment experience and wisdom of doctors of all ages. Effectively sorting out and excavating them, especially combining with their functional indications and medication rules, can improve clinical efficacy and new drug research and development. Because the classification system of prescriptions is not completely unified, the data of Chinese patent medicines, national standard formulae and the seventh editions of Chinese formula textbooks are manually integrated, and a quadrat data set containing 21 efficacy classifications is constructed. Since on the text description of prescription information (prescription name, composition, indications and efficacy) in the data set, a variety of deep learning text classification models are used to automatically judge prescription classification, so as to establish an efficient and accurate classification model, and finally construct a prescription efficacy classification data set. The experimental results show that the pre-trained Bert-CNN model has the best effect, with the accuracy rate of 77.87%, and the weighted accuracy rate, weighted recall rate and weighted F1 value of 79.46%, 77.87% and 77.44%, respectively. This study provides a useful reference for further realizing the automatic information processing of ancient formulas.

Keywords: Text classification · Deep learning · Prescription efficacy classification · Bert · CNN

1 Introduction

Through thousands of years of clinical practice of traditional Chinese medicine, a large number of formulae have been selected and accumulated. By the end

of the late Qing Dynasty, there were more than 100,000 ancient formulae. In addition, the new formulae developed and created, the self-made formulae of hospitals, and the Chinese patent medicine formulas on the market, they were even more "vast". These formulas embody the therapeutic experience and wisdom of ancient and modern physicians. If they can be effectively sorted out and excavated, especially their functions and indications, they will provide important support for improving clinical efficacy and developing new drugs.

In order to sort out and study the formulae of past dynasties, colleges and research institutes around the country have established various prescription databases and analysis systems [1, 2], which still face problems such as insufficient standardization of data and difficulty in efficient retrieval and analysis [3]. Among all kinds of information contained in formulae, in addition to the structure and standardization of drug composition information, more attention has been paid and better solutions have been achieved [4, 5]. However, there are few studies on information extraction and standardization in terms of the information of the opposite side (including indications and efficacy). The main reason is that its description covers a long history, and there are many phenomena such as polysemy, synonymy, ambiguity, cross meaning. In addition, the classification system of formula is not completely unified, and there is a lack of high-quality and large-scale training corpus. At present, there are few researches on information extraction and standardization based on natural language processing technology of deep learning model.

In recent years, with the extensive application of depth model in the medical field and the release and improvement of the national standards related to formulae, conditions have been provided for the automatic processing of prescription information. In this paper, the classification system and prescription data of the seventh edition of National Medical Insurance Catalogue [6], GBT 31773-2015 Coding Rules and Coding of Traditional Chinese Medicine Prescriptions [7] and Prescription Textbook [8] are manually integrated to form a prescription efficacy classification data set, and a variety of deep learning text classification models are adopted to realize text description based on prescription name, composition and main efficacy, and automatically judge the classification of the prescription. The classification efficiency of various models is compared, analyzed and discussed, which provides a useful reference for the further realization of the automatic information processing of ancient formulae.

2   Related Work

Text classification refers to the process of automatically determining text categories based on text content under a given classification system. The core issue is to extract the features of the classified data from the text, and then select the appropriate classification algorithms and models to model the features, thereby achieving classification. Figure 1 shows the processing procedure of text classification.
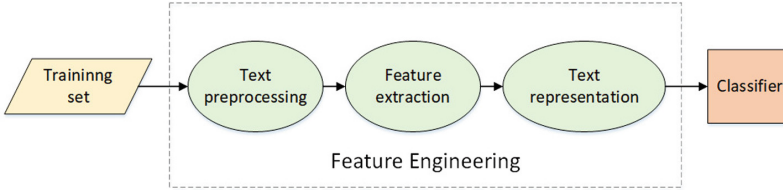
Fig. 1. Processing procedure of text classification.

Text classification algorithm model can be divided into traditional machine learning model and deep learning model. Among them, the commonly used traditional machine learning models include Naive Bayes [9], Support Vector Machine (SVM) [10], Decision Tree [11], etc. Deep learning models include TextCNN, TextRNN, TextRCNN, HAN [12], Bert [13], etc.

At present, text classification has been widely used in many fields, such as spam filtering [14], public opinion analysis [15] and news classification [16]. In the medical field, Yu [17] proposed a named entity recognition model to automatically identify the time and place information in the COVID-19 patient trajectory text. Li [18] proposed a three-stage hybrid method based on gated attention bidirectional long short-term memory (ABLSTM) and regular expression classifier for medical text classification tasks to improve the quality and transparency of medical text classification solutions. Prabhakar [19] proposed a medical text classification paradigm, using two novel deep learning architecture to alleviate human efforts. Cui [20] developed a new text classifier based on regular expressions. Machine-generated regular expressions can effectively combine machine learning technologies to perform medical text classification tasks, and have potential practical application value. Zheng [21] proposed a deep neural network model called ALBERT-TextCNN for multi-label medical text classification. The overall F1 value of the model classification reached 90.5 %, which can effectively improve the multi-label classification effect of medical texts. Li [22] proposed a two-level text classification model based on attention mechanism, which is used to classify biomedical texts effectively.

Generally, text classification has also been widely used in the medical field. In this paper, we use the deep learning model to classify the efficacy of prescription. The paper introduces the model used and analyzes the experimental results in detail.

3   Methodology

In this study, we aim to classify the efficacy of prescription into multiple categories using deep learning method which includes TextCNN, TextRCNN, RNN Attention, Bert and their combination models.

3.1   Convolutional Neural Network Model

CNN (Convolutional Neural Network) [23] is widely used in the field of image recognition. The structure of CNN model can be divided into three layers, i.e., convolutional layer, pooling layer and fully connected layer. The main function of convolution layer is to extract features. The pooling layer aims to down sampling but with no damage of the recognition results. The main role of the fully connected layer is classification.

TextCNN is a model proposed by Kim [24] in 2014, which pioneered the use of CNN to encode n-gram features for text classification. The convolutions in the image is two-dimensional, whereas TextCNN uses one-dimensional convolution (filter_size * embedding_dim), with one dimension equal to embedding. This can extract the information of filter_size grams. After inputting data, the Embedding layer converts words into word vectors and generates a two-dimensional matrix. Then, sentence features are extracted in one-dimensional convolution layer. After that, the sentences of different lengths are represented by fixed length in Max-Pooling layer. Finally, the probability distribution is obtained in Fully connected layer. The structure of TextCNN is shown in Fig. 2.
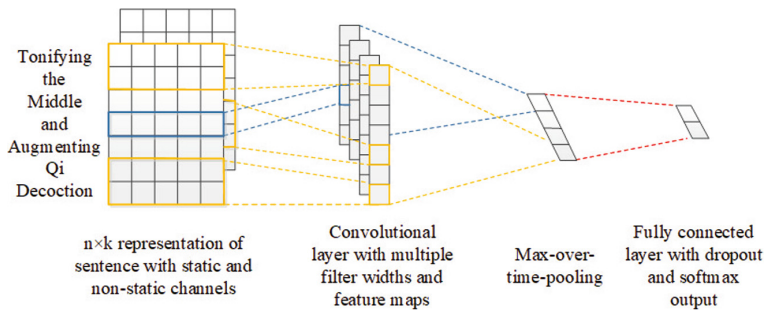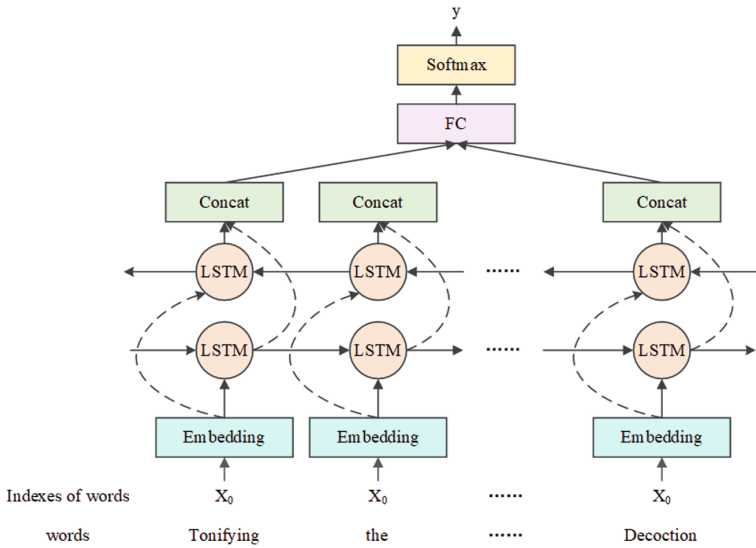

Fig. 2. The structure of TextCNN.

3.2   Recurrent Neural Network Model

RNN (Recurrent Neural Network) [25] is a kind of neural network with shortterm memory. In the RNN models, neurons cannot only accept the information of other neurons, but also accept their own information to form a network structure with loops. Compared with feedforward neural networks, RNNs are more in line with the structure of biological neural networks. RNN has been widely used in speech recognition, language models and natural language generation tasks.

TextRNN [26] takes advantages of RNN to solve text classification problems, trying to infer the label or label set of a given text (sentence, document, etc.). TextRNN has a variety of structures and its classical structure includes embedding layer, Bi-LSTM layer, concat output, FC layer and softmax. The structure of TextRNN is shown in Fig. 3.

Fig. 3. The structure of TextRNN.

TextRCNN [27] (TextRNN+CNN) first uses bi-directional RNN to obtain the forward and backward context representations of each word, so that the representation of the word becomes the form of splicing word vectors and forward and backward context vectors, and finally connects the same convolution layer as TextCNN and pooling layer.

RNN_Attention [28] introduces an attention mechanism based on RNN to reduce the loss of detailed information when processing long text information.

3.3 Bidirectional Encoder Representation from Transformers Model

BERT (Bidirectional Encoder Representation from Transformers) [13] is a pretrained language representation model that uses MLM (Masked Language Model) for pre-training and deep bidirectional Transformer component to construct the entire model to generate deep bidirectional language representations that fuse contextual information.

Each token of the input information (the yellow block in Fig. 4) has a corresponding representation, including three parts, i.e., Token Embedding, Segment Embedding, and Position Embedding. The vectors of the final input model are obtained by adding their corresponding positions, and then classified. The structure of BERT-classify model is shown in Fig. 4.

Bert-CNN, Bert-RNN, Bert-RCNN and Bert-DPCNN are combination models using Bert pre-trained models. First, the data is put into Bert model for pretraining, and then the output of BERT (that is, the output of the last layer of the transformer) is used as the input of the convolution (embedding_inputs) to access other models.
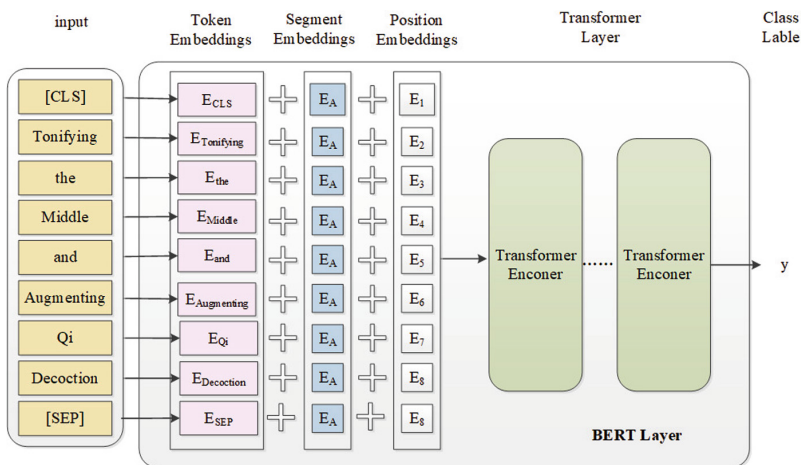
Fig. 4. The structure of BERT-classify model. (Color figure online)

4 Experiment

4.1 Data Sources

The experimental data were collected from Chinese patent medicines [6], national standard formulae [7] and the seventh editions of Chinese formula textbooks [8]. A total of 2,618 prescription data were manually integrated. Among them, there are 1,391 pieces of Chinese patent medicine data, 1,089 pieces of national standard formulae data and 138 pieces of Chinese formula textbook data. Each data contains the name, composition, efficacy and indications of the prescription. The experimental data format is as follow: {Prescription: Bufei Huoxue Capsule} {Composition: Membranous milkvetch root 720 g, radix paeonia rubra 720 g, malaytea scurfpea fruit 360 g} {Efficiency: benefiting qi for activating blood circulation, invigorating lung and nourishing kidney} {Indication: cor pulmonale (remission stage) diagnosed as qi deficiency and blood stasis syndrome. The symptoms include cough, shortness of breath, asthma, chest tightness, palpitation, cold and weak limbs, soreness and weakness of waist or knees, cyanosis of lips, pale tongue with white coating or dark purple tongue.} {Real label: 1} {Category: reinforcing agents}

4.2 Data Integration and Standardization

We exclude Tibetan medicine, Mongolian medicine, Uygur medicine and other ethnic medicines. We refer to the national classification standard and integrate the efficacy classification based on departmental division. For example, some traditional Chinese patent medicines are classified according to the names of modern diseases, such as drugs for nasal diseases, ear diseases, and anti-tumor drugs. Their efficacy indications are mainly described by modern diseases, which

deviates greatly from the terminology of efficacy indications of conventional formula. Referring to the primary and secondary classifications in the national standard prescription classification, the fine curative efficacy classification of Chinese patent medicines is integrated into the primary classification of the same efficacy.

A total of 2,618 prescription data were manually integrated. We retain the data based on efficacy classification, remove duplicate data and the data of modern disease naming efficacy, resulting in total of 2,368 prescription data. Then we sift out 12 items of incomplete efficacy and treatment and 2 items of emetic (too few data in this category), leaving a total of 2,354 prescription data. After data integration and screening, the final prescription efficacy can be divided into 21 categories. The distribution of sample categories is shown in Table 1.

Table 1. Distribution of sample categories.

Label	Category	Sample size
0	tranquilizing formulae	21
1	reinforcing formulae	411
2	astringent preparations	55
3	settlement formulae	42
4	diaphoretic formulae	150
5	resuscitating formulae	25
6	regulating qi formulae	228
7	blood regulating formulae	253
8	antipyretic agents	402
9	anthelmintic	8
10	dampness-dispelling formulae	227
11	summer-heat clearing formulae	20
12	interior-warming formulae	76
13	digestive formulae	37
14	purgative formulae	53
15	carbuncle therapeutic formulae	25
16	wind-calming medicine	113
17	moistening formulae	14
18	turbidity lipid-lowering formulae	16
19	detumescence formulae	28
20	phlegm, cough, asthma formulae	150

4.3   Experimental Parameters

The 2,354 prescription data were randomly disrupted and divided into training set, validation set and test set at a ratio of 8: 1: 1.

The model parameters used in this paper are shown in Table 2. To reduce the risk of over-fitting of the model, the training follows the principle of early stopping, and sets the detection parameter detect_imp = 1,000, dropout = 0.1.

Table 2. Model parameters.

Model	epoch	batch size	pad size	learning rate	num filters	hidden size
TextCNN	20	32	128	1e−3	256	128
TextRCNN	20	32	128	1e−3	256	256
RNN_Attention	20	32	128	1e−3	256	128
Bert, Bert-CNN and Bert-RNN	7	32	128	5e−3	256	768
Bert-RNN	7	32	128	5e−3	250	768

4.4 Statistics for Model Evaluation Measures

To evaluate the performance, our experiment used the classic evaluation indexes in text classification, namely Accuracy, Precision, Recall and F1-Measure. Due to the limited number of multi category data in the research, and the imbalanced samples in each category, weighted recall and weighted F1 are used as the overall evaluation indicators. The formulas are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{3}$$

$$F1 - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \tag{4}$$

TP, TN, FP and FN are positive samples with correct prediction, negative samples with correct prediction, positive samples with wrong prediction and negative samples with wrong prediction, respectively. In multi-classification task, the currently tested classification is treated as a positive sample, and other classifications are treated as negative samples. Firstly, the accuracy rate, recall and F1 value of each category are measured, and then the weighted value is obtained (different weights are given according to the proportion of each category).

4.5 Results

Included the 3 non-pretrained models, 8 models were used to classify the efficacy of formula. The experimental results are shown in Table 3.

In the non-pre-trained model, the TextRCNN model outperforms the other two models. The accuracy is 3 to 8% points higher, reaching 72.77%, with a loss value of 0.87, slightly worse than TextCNN's 0.82. Weighted-Precision, Weighted-Recall and Weighted-F1 are the best of the three models. RNN_Attention model performed worst, with an accuracy of only 64.68%.

Table 3. Experimental results.

Type	Model	Accuracy	Loss	Weighted Precision	Weighted Recall	Weighted F1
Non-pre-trained model	TextCNN [24]	70.64%	**0.82**	71.03%	70.64%	69.63%
	TextRCNN [27]	72.77%	0.87	73.83%	72.77%	71.64%
	RNN_Attention [28]	64.68%	1.2	64.79%	64.68%	63.35%
Bert and Bert combination model	Bert [13]	77.87%	0.83	77.06%	77.87%	76.75%
	Bert-CNN	**77.87%**	0.84	**79.46%**	**77.87%**	**77.44%**
	Bert-RNN	67.66%	1.2	63.84%	67.66%	63.19%
	Bert-RCNN	69.79%	0.9	69.00%	69.79%	68.42%
	Bert-DPCNN	73.19%	1.1	68.44%	73.19%	69.50%

The combination models using the Bert pre-trained language model outperform non-pre-trained models. Among them, the Bert-CNN model has the best effect in the efficacy classification of experimental prescriptions in this paper. Accuracy reaches 77.87% and Loss value is slightly inferior to the TextCNN model. Based on the above results, Bert-CNN has the best effect and is most suitable for this experiment.

The Precision of each efficacy category in the Bert-CNN model is as follows. Tranquilizing formulae: 66.67%; reinforcing formulae: 73.17%; astringent preparations: 100.00%; settlement formulae: 20.00%; diaphoretic formulae: 63.16%; resuscitating formulae: 50.00%; regulating qi formulae: 77.27%; blood regulating formulae: 88.00%; antipyretic agents: 76.60%; anthelmintic: 100.00%; dampnessdispelling formulae: 90.91%; summer-heat clearing formulae: 100.00%; interiorwarming formulae: 100.00%; digestive formulae: 50.00%; purgative formulae: 100.00%; carbuncle therapeutic formulae: 100.00%; wind-calming medicine: 85.71%; moistening formulae: 0.00%; turbidity lipid-lowering formulae: 100.00%; detumescence formulae: 75.00% and phlegm, cough, asthma formulae: 90.00%.

5   Discussion

5.1   Analysis of the Overall Results of Experimental Classification

Model Performance. The classification prediction accuracy of most models is above 70%. In non-pre-trained models, Text_RCNN performs best and RNN_Attention performs worst. The attention mechanism of RNN_Attention model is usually combined with RNN, which relies on the historical information of t-1 to calculate the information at time t, so it cannot be implemented in parallel, resulting in low prediction efficiency. In Text_RCNN model, the convolutional layer + pooling layer in CNN is replaced by two-way RNN + pooling layer, so the prediction result is better. In general, Bert pre-trained language model and its combination models are superior to the non-pre-trained models, and Bert-CNN has the best classification effect. The main model structure of Bert is Transformer encoder, which has stronger text editing ability than RNN and LSTM.

Classification Data. The accuracy of the larger sample size is about 80%, while the smaller sample size is unstable. In addition to the sample size, it is also related to the classifications' clear feature words, which are better for machine learning. 7 classifications out of the 21 categories had accuracy of 100%. The primary reason is that these classifications' defining phrases are comparatively plain, and machines are generally simple to learn. Moistening formulae has the lowest accuracy, followed by digestive formulae and resuscitating formulae (50%), which is easily confused with other classifications.

5.2   Summary and Cause Analysis of Typical Misclassified Data

By analyzing the cases of machine classification errors, we can draw the following conclusions, as is shown in Table 4. It shows that although the model features, recognition features, and learning features are different, the results of some models are very similar. But each model usually has different characteristics, and the characteristics of recognition and learning also have their own emphasis. Taking Dingkun Dan as an example, the results of the model recognition are mainly divided into "reinforcing formulae" (label is 1) and "blood regulating formulae" (label is 7), and the number of both is also relatively close.

Table 4. Typical error cases.

Model	Prescription(Real Label)		
	Bolus for Woman Disease (1)	Decoction of Ophiopogonis (17)	Xinsusan Decoction (17)
TextCNN	1	17	20
TextRCNN	7	17	20
RNN_Attention	7	14	4
Bert	7	8	4
Bert-CNN	1	8	4
Bert-RNN	1	8	20
Bert-RCNN	7	8	4
Bert-DPCNN	1	8	4

Text classification algorithm based on text feature is difficult to understand text connotation without considering TCM theory knowledge. For example, Maimendong Decoction, as a "moistening formulae" (label is 17), has fewer characteristic words related to "dryness" than "heat" and "fire", which causes most models to recognize it as a "antipyretic agents" (label is 8). Similarly, the model cannot understand that "Qingbanxia" can dry and damp phlegm, nor "ophiopogon" can nourish yin and moisten dryness.

The classification of existing standards is also controversial. For example, the standard classification of Xingsusan Powder is "moistening formulae" (label is

17), while most models classify it as "phlegm, cough, asthma formulae" (label is 20) and "diaphoretic formulae" (label is 4). From the perspective of traditional Chinese medicine theory, the classification results are also acceptable.

6   Conclusion

In this paper, a standard data set of prescription efficacy classification was constructed. By comparing and analyzing the efficiency of several deep learning text classification models, it is found that the Bert-CNN model is the best, which improves the accuracy of automatic classification of prescription function.

Future work aims to address the following issues: Collect more data. Using methods such as K-fold cross validation to improve data utilization; Adding medical pre-training model to improve the accuracy of the model to identify medical terms, and increase prior knowledge in the field of traditional Chinese medicine, ultimately improve the accuracy of model classification; More experts in relevant fields will be invited to participate in the discussion on the efficacy classification of prescription. On the one hand, the efficacy classification criteria could be optimized. On the other hand, the standard classification results could be optimized and adjusted.

References

1. Yan, Z., Bo, G., Meng, C.: Design and implementation of the analysis system of TCM prescription. China J. Tradit. Chin. Med. Pharm. **29**(5), 1543–1546 (2014)
2. Cui, Y., Gao, B., Liu, L., Liu, J., Zhu, Y.: AMFormulaS: an intelligent retrieval system for traditional Chinese medicine formulas. BMC Med. Inform. Decis. Mak. **21-S**(2), 56 (2021). https://doi.org/10.1186/s12911-021-01419-8
3. Yan, Z., Bo, G., Meng, C.: Statistical analysis on phenomenon of homonym and synonym of Chinese materia medica. China J. Tradit. Chin. Med. Pharm. **30**(12), 4422–4425 (2015)
4. Deshan, G., Wenyu, L., Bingzhu, Z., Xingguang, M.: Application of named entity recognition in the recognition of words for Chinese traditional medicines and Chinese medicine formulae. Chin. Pharmac. Affairs (2019)
5. Yan, Z., Ling, Z., Jun-Hui, W., Meng, C.: An efficient approach of acquiring knowledge from ancient prescriptions and medicines based on information extraction. China J. Tradit. Chin. Med. Pharm. **30**(5), 316–321 (2015)
6. Administration N H S, China M O H R a S S O T P S R O. Medicine catalogue of national basic medical insurance. Work-Related Injury Insurance and Maternity Insurance (2020). 2020-12-28 [2022-03-02]
7. Medicine N a O T C. Coding rules for Chinese medicinal formulae and their codes: GB/T 31773-2015. China Quality and Standards Publishing, Beijing (2015)
8. Zuo, Y., Li, Z.: Prescriptions of Chinese materia medica. China Press of Traditional Chinese Medicine (2021)

9. Xiao, L., Wang, G., Liu, Y.: Patent text classification based on naive Bayesian method. In: 11th International Symposium on Computational Intelligence and Design, ISCID 2018, Hangzhou, China, 8–9 December 2018, vol. 1, pp. 57–60. IEEE (2018). https://doi.org/10.1109/ISCID.2018.00020

10. Xia, H., Min, T., Yi, W.: Sentiment text classification of customers reviews on the Web based on SVM. In: International Conference on Natural Computation. IEEE (2010). https://doi.org/10.1109/ICNC.2010.5584077

11. Cañete-Sifuentes, L., Monroy, R., Medina-Pérez, M.A.: A review and experimental comparison of multivariate decision trees. IEEE Access **9**, 110451–110479 (2021). https://doi.org/10.1109/ACCESS.2021.3102239

12. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A.J., Hovy, E.H.: Hierarchical attention networks for document classification (2016). https://doi.org/10.18653/v1/n16-1174

13. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. CoRR abs/1810.04805 (2018). http://arxiv.org/abs/1810.04805

14. Nagwani, N.K., Sharaff, A.: SMS spam filtering and thread identification using bilevel text classification and clustering techniques. J. Inf. Sci. **43**(1), 75–87 (2017). https://doi.org/10.1177/0165551515616310

15. Hai-Bing, M.A., Jiu-Yang, B.I., Xin-Shun, G.: Applications of text classification in network public opinion system. Inf. Sci. **33**(05), 97–101 (2015)

16. Kapusta, J., Drlík, M., Munk, M.: Using of N-grams from morphological tags for fake news classification. PeerJ Comput. Sci. **7**, e624 (2021). https://doi.org/10.7717/peerj-cs.624

17. Yu, H., Pan, X., Zhao, D., Wen, Y., Yuan, X.: A hybrid model for spatio-temporal information recognition in COVID-19 trajectory text. In: Zhao, X., Yang, S., Wang, X., Li, J. (eds.) WISA 2022. LNCS, vol. 13579, pp. 267–279. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-20309-1_23

18. Li, X., Cui, M., Li, J., Bai, R., Lu, Z., Aickelin, U.: A hybrid medical text classification framework: Integrating attentive rule construction and neural network. Neurocomputing **443**, 345–355 (2021). https://doi.org/10.1016/j.neucom.2021.02.069

19. Prabhakar, S.K., Won, D.: Medical text classification using hybrid deep learning models with multihead attention. Comput. Intell. Neurosci. **2021**, 9425655:1–9425655:16 (2021). https://doi.org/10.1155/2021/9425655

20. Cui, M., Bai, R., Lu, Z., Li, X., Aickelin, U., Ge, P.: Regular expression based medical text classification using constructive heuristic approach. IEEE Access **7**, 147892–147904 (2019). https://doi.org/10.1109/ACCESS.2019.2946622

21. Zheng, C., Wang, X., Wang, T., et al.: Multi-label classification for medical text based on ALBERT-TextCNN model. J. Shandong Univ. (Sci. Edn.) **57**(04), 21–29 (2022)

22. Li, Q., Liao, W.: Biomedical text classification model based on attention mechanism. Chin. J. Med. Phys. **39**(04), 518–523 (2022)

23. Zhou, F., Jin, L., Dong, J.: Review of convolutional neural network. Chin. J. Comput. **40**(6), 23 (2017)

24. Kim, Y.: Convolutional neural networks for sentence classification. (2014). https://doi.org/10.3115/v1/d14-1181

25. Yang, L., Wu, Y., Wamg, J., Liu, Y.: Research on recurrent neural network. J. Comput. Appl. **38**(A02), 7 (2018)

26. Liu, P., Qiu, X., Huang, X.: Recurrent neural network for text classification with multi-task learning (2016)

27. Lai, S., Xu, L., Liu, K., Zhao, J.: Recurrent convolutional neural networks for text classification. In: National Conference on Artificial Intelligence (2015)
28. Zhou, P., Shi, W., Tian, J., Qi, Z., Xu, B.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (2016)