applied sciences

MDPI

Article

Prediction of Venous Thrombosis Chinese Electronic Medical Records Based on Deep Learning and Rule Reasoning

Jiawei Chen, Jianhua Yang and Jianfeng He *

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650032, China
* Correspondence: jfenghe@kmust.edu.cn

Abstract: Aiming at the problems of heavy workload of medical staff in the process of venous thrombosis prevention and treatment, error evaluation, missed evaluation, and inconsistent evaluation, we propose a joint extraction model of Chinese electronic medical records based on deep learning. The approach was to first construct the handshake annotation, then use bidirectional encoder representations from transformers (BERT) as the word vector embedding, then use the bidirectional long short-term memory network (BiLSTM) to extract the contextual features, and then integrate the contextual information into the process of normalizing the word vector. Experiments show that our proposed method achieves 93.3% and 94.3% of entity and relation F1 on the constructed electronic medical record dataset, which effectively improves the effect of medical information extraction. At the same time, the venous thromboembolism (VTE) risk factors extracted from the electronic medical record were used to judge the risk factors of venous thrombosis by means of rule reasoning. Compared with the assessment of clinicians on the Wells and Geneva scales, the accuracy rates of 84.7% and 86.1% were obtained.

Keywords: venous thromboembolism (VTE); deep learning; information extraction; electronic medical record (EMR); joint extraction

check for updates

Citation: Chen, J.; Yang, J.; He, J. Prediction of Venous Thrombosis Chinese Electronic Medical Records Based on Deep Learning and Rule Reasoning. *Appl. Sci.* **2022**, *12*, 10824. https://doi.org/10.3390/app122110824

Academic Editor: Jan Egger

Received: 10 September 2022
Accepted: 15 October 2022
Published: 25 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Venous thromboembolism (VTE) is a group of diseases including deep vein thrombosis and pulmonary embolism. VTE is one of the three most common vascular diseases, and its mortality is second only to tumors and myocardial infarction [1]. Surveys in 32 countries show that more than half of hospitalized patients are at risk of VTE, and only half of high-risk patients have access to effective preventive measures [2]. The occurrence of VTE in hospitalized patients is often related to a series of factors such as the patient's advanced age, genetic predisposition, and surgical treatment [3], VTE is also an important cause of unexpected deaths in hospitals, and has become a serious problem faced by hospital administrators and clinical medical staff. Early identification and diagnosis can effectively reduce the risk of venous thromboembolism [4]. At present, early screening is mainly performed by using the VTE risk assessment scale. The more commonly used assessment scales are Padua, Caprini, Wells, etc. The above risk assessment scales all use the linear weighting method, assigning different scores to different risk factors in the assessment scale, finally adding the scores to obtain the risk assessment score, and then refer to VTE guidelines to determine a patient's current VTE risk level and what precautions should be taken. However, there are too many items in the clinically used VTE assessment scale. Taking the Caprini scale as an example, there are as many as 37 items. The way of filling in the form manually not only increases the workload of medical staff, but also uses the evaluation method of a table item, which will lead to different results due to the different experience of medical staff, making them therefore prone to inconsistency in evaluation, resulting in patients being unable to obtain correct preventive measures.

With the development of artificial intelligence technology and EHR storage technology [5], some scholars use artificial intelligence technology to intelligently assess the risk of VET. For example, Qatawneh et al. [6] divided the risk level of VTE into 5 categories, and extracted 35 risk factors from the Caprini assessment scale; 35 risk factors were input into a multilayer perceptron feedforward neural network to predict the VET risk level of hospitalized patients. Inconsistencies in assessments can be addressed, but risk factors entered into the model need to be captured by the physician from the electronic medical record. Vilhena et al. [7] developed a VTE diagnosis decision support system based on knowledge representation and logical reasoning, supplemented by case-based reasoning. Case-based reasoning method requires time-consuming and laborious construction of case base. Yang et al. [8] proposed an ontology-based assessment model for VTE, constructing an ontology including comprehensive information such as symptoms, diagnoses, drugs, and surgeries, and using rule-based and greedy matching methods to extract data from electronic medical records risk factors, and finally using machine learning to judge a risk level. This method can identify the information related to VTE from the electronic medical record, but it takes time and effort for domain experts to build the ontology set. Elias et al. [9] developed an algorithm to automatically calculate the VTE risk score for the Padua assessment scale and calculate the Padua assessment scale score by obtaining the risk factors related to the Padua assessment scale from the admission record. This method requires building an Enterprise Data Warehouse (EDW) to identify risk factors. In Table 1, we summarize the recent work.

Table 1. The recent work.

Paper	Method	Evaluation Scale	Result
Qatawneh et al. [6]	Multilayer perceptron feedforward neural network is used to predict the VET risk level of hospitalized patients.	Caprini	Accuracy: 81%
Vilhena et al. [7]	On the basis of knowledge representation and logical reasoning, case-based reasoning is used to predict the risk level of VTE.	Thrombophilia predisposition risk	Accuracy: 90%
Yang et al. [8]	An ontology-based assessment model for venous thromboembolism.	Padua	AUC: 97.3%
Elias et al. [9]	An algorithm for automatically calculating VTE risk score for Padua assessment scale.	Padua	AUC: 81%

We can find that although the current VTE intelligent assessment method uses artificial intelligence technology, the acquisition of risk factors still relies on the extraction of medical staff, and the workload of medical staff has not been greatly liberated. As a precious medical resource, electronic medical records record various information of the patient's medical treatment process. We can use the method of information extraction to extract the risk factors related to VTE for intelligent evaluation of VTE. For example, Zhou et al. [10] developed an AICDSS system that was connected to the hospital's electronic medical record system to obtain information from the electronic medical record system to perform VTE assessment on patients.

At present, information extraction can be divided into two types: pipeline extraction and joint extraction. The pipeline-based entity relation extraction divides entity recognition and relation extraction into two independent sub-tasks. First, the identification of named entities is performed, and then the relation extraction is performed on the basis of entities. Although the structure of the pipeline method is simple, the error of entity recognition will be transmitted to the relation extraction model, resulting in poor final results. In order to overcome this problem, some scholars have proposed a model of entity-relation joint extraction. The joint extraction approach refers to using a model to identify entities and relations from text. The joint extraction model is divided into two methods based on parameter sharing and labeling strategy. The method based on parameter sharing [11–13]

is to pass two tasks through the same encoding layer and let both tasks find the best parameters during training. Based on the method of parameter sharing, entity recognition and relation extraction use independent decoding methods, resulting in weak interaction between the two models. For this reason, some researchers proposed a new labeling strategy [14], and proposed using a unified way to encode and decode entities and relations to solve the problem of weak interaction, thereby improving the effect of joint extraction. The method of labeling strategy is often problematic. There is a problem in that the model parameters are too large, and the inference speed is slow.

Based on the current problems in VTE intelligent assessment and the problem of joint model error transmission in information extraction, we propose a joint extraction model to extract VTE risk factors from electronic medical records, and then formulate rules in the Wells and Geneva assessment scales of pulmonary embolism, and then use rule reasoning to judge the risk level of pulmonary embolism.

2. Method

This paper proposes a joint extraction model, which can be divided into a bidirectional encoder representations from transformers (BERT) word embedding layer, Bidirectional Long Short-Term Memory (BiLSTM) context feature extraction layer, Conditional Layer Normalization (CLN) [15], handshake marked, and Linear layer. As shown in Figure 1, for the input sentence, the word vector is first obtained through the BERT word embedding layer, and then the obtained word vector is obtained through BiLSTM to obtain the contextual semantic information. Then, the above and below information are used as conditions to perform conditional normalization on the word vector and integrate the contextual information into the normalized feature space. After the fused features are obtained, each word in a sentence is concatenated with the remaining words to form handshake marked [16], finally, a linear layer is used for feature classification. Each part of the model is described in detail below.

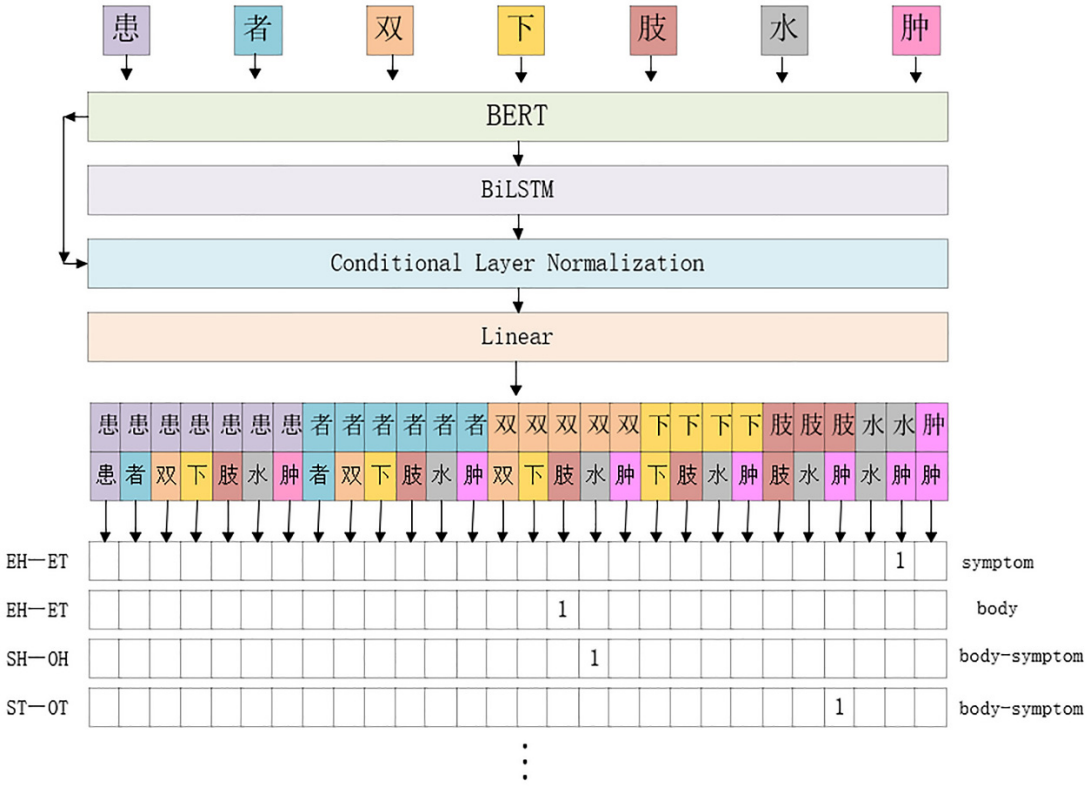


Figure 1. The model structure. In this example, the input sentence is ‘ 患者双下肢水肿’ (The patient had edema of both lower limbs), and we use a different color for each word.

2.1. BERT Word Embedding Layer

Text data cannot be directly input into the neural network for training, and it needs to be vectorized for training. There are currently many vectorization methods, such as Word2vec [17], ELMo [18], and BERT [19], The vectorization method of Word2vec obtains static word vectors through a large amount of text training, which cannot solve the problem of ambiguity of the same word in different contexts. The ELMo model solves the problem of polysemy by dynamically adjusting the context of the word vector, but the LSTM used by the ELMo model has insufficient ability to extract features. BERT is currently the most commonly used pre-training model. BERT uses the Transformer's structure, represented by bidirectional encoding, which has powerful feature extraction capabilities. At the same time, it uses the "masked language model" to randomly mask 15% of the words during training. This enables BERT to overcome the problems of Word2vec and ELMo. Figure 2 is the structure diagram of BERT. The input of BERT is composed of the addition of three embedding features: word embedding, position embedding, and segmentation embedding. Trm represents Transformers, E represents BERT's input, and T represents the word vector output after bidirectional coding of Transformers.

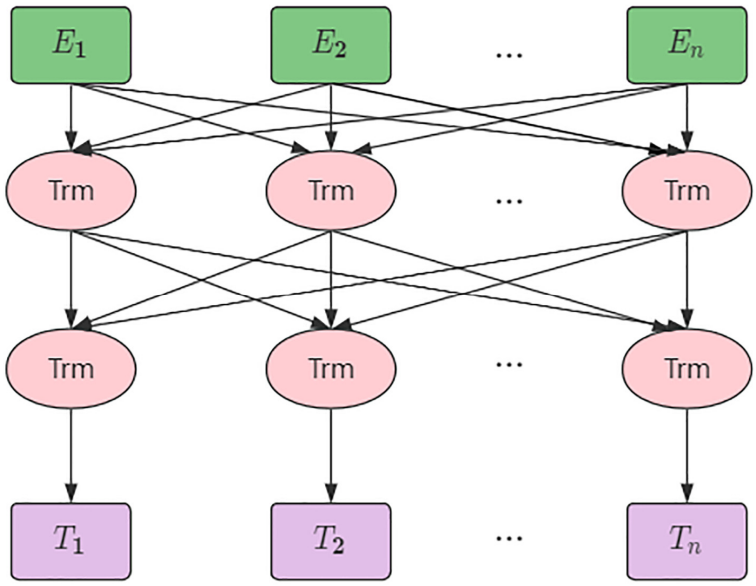


Figure 2. BERT structure.

2.2. BiLSTM

LSTM controls the sequence information through the input gate, forget gate, and output gate, which can well solve the gradient disappearance or gradient explosion problem in RNN. Figure 3 is the structure of BiLSTM. BiLSTM is composed of two LSTM units, one is the forward LSTM and the other is the backward LSTM. Two hidden layer vectors with different direction information can be obtained through the two LSTMs, and then the two vectors are spliced together to obtain the final output. BiLSTM can extract forward and backward information of words and words in a sentence, and better obtain semantic features and long-distance information.

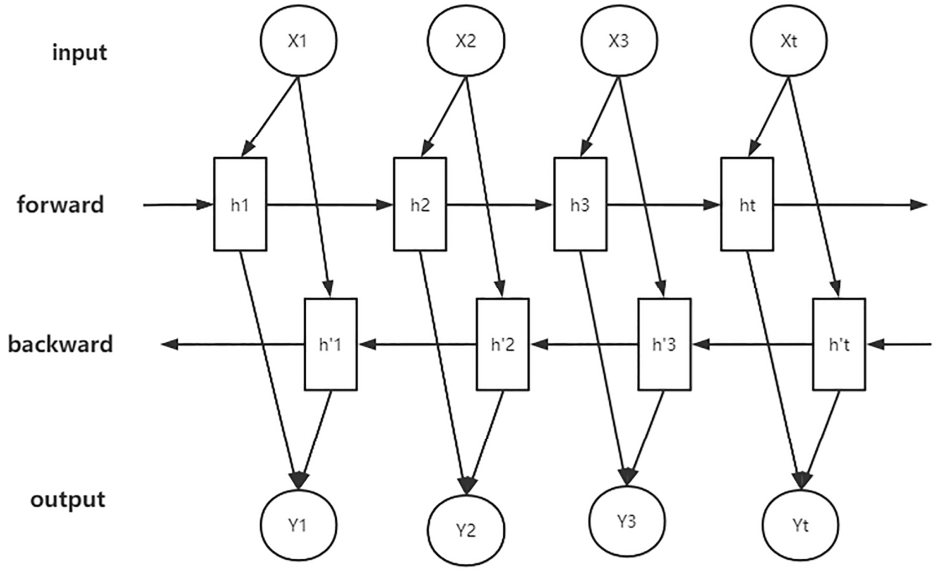


Figure 3. BiLSTM structure.

2.3. CLN

Normalization normalizes the word vector encoded by BERT by calculating the mean and variance of the data, which solves the problem of gradient explosion and gradient disappearance to a certain extent, and the conditional normalization here refers to normalization in the process of integrating other features into the normalized results. Figure 4 is a structural diagram of conditional layer normalization.

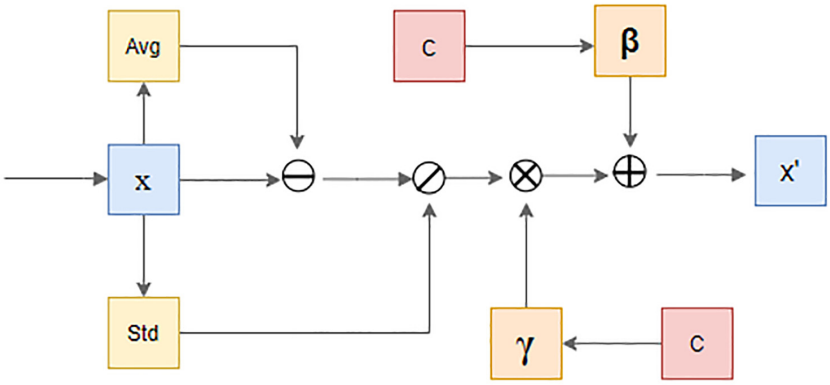


Figure 4. Conditional layer normalization structure.

In this paper, the context features extracted by BiLSTM are incorporated into variables β and γ as conditions, so as to integrate the context information into the normalized word vector space. The formula for CLN is as follows.

$$\mathrm{CLN}(x) = \gamma_c \left(\frac{x - E(x)}{\sqrt[2]{Var(x) + \varepsilon}} \right) + \beta_c \tag{1}$$

$$\gamma_c = W_\gamma + b_\gamma \tag{2}$$

$$\beta_c = W_\beta + b_\beta \tag{3}$$

2.4. Handshake Marked

Inspired by Wang et al. [15], we construct a marker matrix for the input sentences. For example, given the sentence ‘ 患者双下肢水肿’ (The patient had edema of both lower limbs.), we constructed the labeling matrix as shown in Figure 5 EH-ET, which was used to

represent the link between the head and tail of an entity, and SH-OH was used to represent the link between the head of the subject entity and the head of the object entity in the relation. Use ST-OT to represent the link between the tail of the Subject entity and the tail of the Object entity in the relation. In Figure 5, two entities, ‘双下肢’ (double lower limbs) and ‘水肿’ (edema), can be seen, as well as the relation link between these two entities. In the actual calculation process, due to the symmetry of the matrix along the diagonal, we convert the shaded part of the marking matrix into the marking sequence shown in Figure 6, which is called the handshake marked. The example sentence: The patient had edema of both lower limbs.

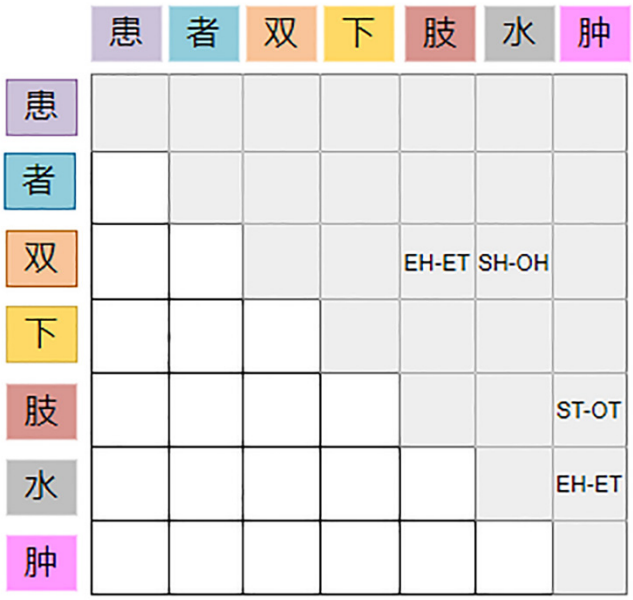


Figure 5. The marker matrix.

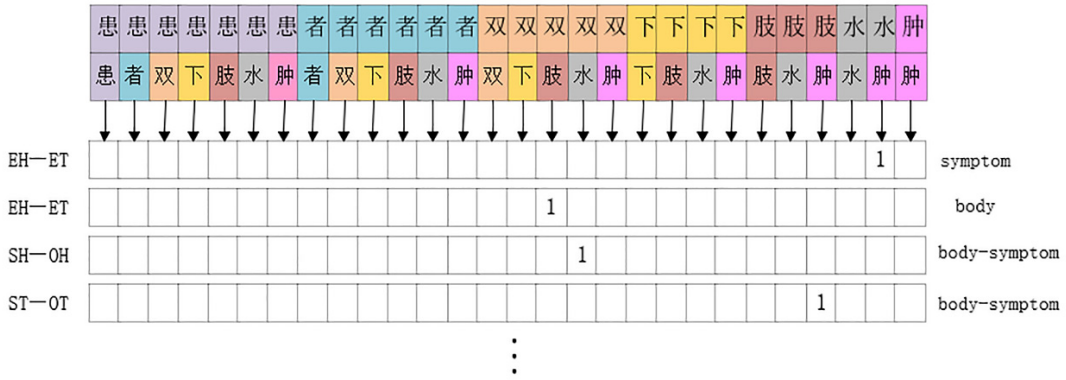


Figure 6. The handshake marked. For illustrative of word combinations, we use a different color for each word in the input sentence.

Figure 6 is the handshake annotation, each type of entity has an EH–ET tag, and each type of relation has an SH–OH and ST–OT tag. Our model is classified by extracting the features between words in the handshake marked which, as shown in Figure 6 (双, 肢), are assigned to the EH–ET of the entity ‘body’, indicating that ‘双下肢’ (double lower limbs) is an entity of ‘body’.

3. Results

3.1. Dataset and Experiment

The data used in this paper are from the clinical electronic medical records of a hospital in Yunnan Province, including chief complaint, past history, examination results, doctor's diagnosis, diagnosis and treatment plan, preoperative diagnosis, and postoperative diagnosis. We collected 3000 electronic medical records and selected 1800 that met the criteria based on the presence or absence of risk factors associated with VTE. First of all, it is necessary to clean the data, including deleting duplicate data, and dealing with missing values and outliers. According to the entries in the risk assessment scale related to VTE, we summarize 12 categories of entities and 6 categories of relationship categories, and then manually annotated the preprocessed files.

After many iterations and revisions, a total of 1800 electronic medical records were marked, with a total of 18,076 sentences, 17,170 entities, and 5326 relations.

The experiment uses 1800 marked electronic medical records, which are divided into training set, validation set, and test set according to the ratio of 8:1:1. The distribution of entity types and relation types of the three datasets is shown in Tables 2 and 3:

Table 2. Statistics on the number of dataset entities.

Entity Type	Train	Validation	Test
Body	1128	152	158
Disease	2784	338	407
Drug	781	97	126
Method	288	33	40
Operation	323	54	58
pastHistory	950	132	105
Result	82	4	8
Symptom	2343	307	300
Test	2359	327	295
Time	137	28	23
Treatment	89	21	8
Value	2275	323	287

Table 3. Statistics on the number of relations in the dataset.

Relation Type	Train	Validation	Test
body–symptom	1207	162	168
method–drug	423	62	76
test–result	84	4	8
test–value	2275	323	287
time–disease	15	2	1
time–operation	171	32	26

Experimental environment: The experiment was run on a computer with Windows 10 operating system, and the CPU was Intel(R) Core (TM) i7-10700F CPU @ 2.90 GHz. The GPU was an NVIDIA GTX3060 and the space size was 12 GB. The computer ran with a memory size of 64 GB. The experiments were run on the environment of Python 3.7 and PyTorch 1.7.0.

3.2. Model Performance Comparison Analysis

The evaluation of the model uses commonly used evaluation metrics using precision (Precision, P), recall (Recall, R), and F1 value. The expressions for precision (P), recall (R), and F1 values are as follows:

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = \frac{2PR}{P + R} \tag{6}$$

Among them: (True Positive) means to predict the positive class as a positive class; FP (False Positive) means to predict the negative class as a positive class; FN (False Negative) means to predict a negative class as a negative class; (TP + FP) means to predict, in the results of the positive class, the proportion of the correct number; (TP + FN) indicates the proportion of the samples that are actually positive classes, correctly judged as positive classes.

In order to avoid the influence of dividing the dataset, we divided the data into ten equal fractions, and selected one of them as the test set and the rest as the training set, one at a time. The ten-cross validation results are shown in Table 4:

Table 4. Ten-cross validation.

Run Set	Entity			Relation		
	P	**R**	**F**	**P**	**R**	**F1**
1	0.905	0.943	0.924	0.934	0.949	0.941
2	0.894	0.956	0.924	0.905	0.937	0.921
3	0.884	0.956	0.919	0.907	0.955	0.930
4	0.880	0.938	0.908	0.905	0.912	0.909
5	0.873	0.948	0.909	0.923	0.945	0.934
6	0.880	0.928	0.903	0.891	0.928	0.909
7	0.905	0.934	0.919	0.924	0.917	0.921
8	0.890	0.953	0.920	0.917	0.919	0.918
9	0.888	0.955	0.920	0.936	0.942	0.939
10	0.915	0.953	0.933	0.929	0.957	0.943
Average	0.891	0.946	0.918	0.917	0.936	0.9265

As can be seen in Table 4, the average F1 of ten-cross validation entity recognition is 0.9179, and that of relation, recognition is 0.9265. The overall results were relatively stable.

In order to further verify the effect of the model proposed in this paper, the following baseline models are selected for comparison:

(1) Multi-Head [11]: Use BiLSTM-CRF for named entity recognition and incorporate the recognized results into relation extraction as soft labels.
(2) SpERT [12]: The core is the pre-train transformer network BERT. Entity recognition and relation classification for all spans in the input sentence via strong negative sampling, span filtering, and local context representation.
(3) Table-Sequence Encoders [13]: It is proposed to use table encoding and sequence encoding to extract features related to named entity and relation classification, respectively, and propose a mechanism to make the two encodings interact.
(4) UniRE [14]: Use a unified label space to represent entities and relations, classify them with a classifier, and eliminate the gap between the two subtasks.

From the data in Table 5, it can be seen that the improved model in this paper has achieved the best results in F1 score compared with other models. Compared with the Multi-Head model, the F1 score of entity recognition and relation extraction has been greatly improved, respectively, by 2.4% and 4.9%. The reason is that the relation extraction model of the Multi-Head model uses real labels in the training phase, but uses the results of entity recognition in the previous phase in the inference phase, resulting in inconsistency between training and inference. In the Multi-Head model, entity recognition and relation extraction are decoded separately, and the interaction information between the two is not enough. Compared with the SpERT model, the F1 score is increased by 1.3% and 3.0%, respectively. Through the pre-training model and the powerful negative sampling ability of SpERT, the SpERT model is better than the Multi-head model, but it is still a two-stage

model, and the decoding method that exists has the problem of error propagation. The same problem exists in the Table-Sequence Encoders model. Compared with the UniRE model, the F1 score entity recognition and relation extraction are improved by 0.9% and 2.3%.

Table 5. Comparative experimental results.

	Entity			Relation		
	P	**R**	**F**	**P**	**R**	**F1**
Multi-Head	0.918	0.899	0.909	0.888	0.900	0.894
SpERT	0.924	0.916	0.920	0.890	0.937	0.913
Table-Sequence Encoders	0.896	0.899	0.898	0.945	0.903	0.924
UniRE	0.921	0.926	0.924	0.936	0.904	0.920
Ours	0.915	0.953	**0.933**	0.929	0.957	**0.943**

Use the tenth partition of the dataset.

Tables 5 and 6 show that our proposed model is the best in terms of the effect of information extraction, the number of model parameters, and the reasoning speed, compared with other related works in recent years. Our proposed model is a one-stage model, which identifies entities and relations in the same label space and solves the problem of error transmission in the two-stage model. In terms of decoding, the two-stage model needs to decode entities and relations, respectively, resulting in too long an inference time, while the one-stage model only needs one decoding method, so the inference speed is improved. Our improved model has the least number of parameters while being the fastest in terms of inference speed, which can process 120 sentences per second.

Table 6. Model parameters and inference speed comparison.

	Model Parameters (M)	**Inference Speed (sent/s)**
Multi-Head	108.55	10
SpERT	108.34	29
Table-Sequence Encoders	111.35	12
UniRE	108.86	95
Ours	**108.20**	**120**

3.3. Overall Effect Analysis

In order to verify the accuracy of the model in the actual use process, the model is integrated in the developed VTE software system and put into clinical use. The scoring interface in the actual application process is shown in Figure 7. Our system is for Chinese electronic medical records, and the corresponding Chinese translations are in brackets.

We selected 137 medical records that were evaluated by clinicians using the Wells scale and the Geneva scale for risk level assessment of pulmonary embolism, input these 137 medical records into the model for information extraction, and then judged by rule inference to obtain the patient's risk assessment result. The risk level of embolism, the final statistical doctor's evaluation results, are compared with the evaluation results of our method, as shown in Table 7.

As can be seen from Table 7, when the clinician used the Wells scale for assessment, they got 73 low risks and 64 high risks, and when using the Geneva scale, they got 94 low risks and 43 high risks. The method proposed in this paper obtained 76 low risks and 61 high risks on the Wells scale, and 94 low risks and 43 high risks on the Geneva scale. Finally, the number of correct predictions in the Wells scale was 116, with an accuracy rate of 84.7%, and the number of correct predictions in the Geneva scale was 118, with an accuracy rate of 86.1%. It can be seen from the experimental results that we used the joint extraction model to extract the information related to VTE risk factors from the medical records, and then used the rule inference method to judge the risk level, which also has a good effect in clinical application.



Figure 7. The scoring interface in the actual application process.

Table 7. Comparison of evaluation results.

	Number of Patients	Low Risk	High Risk	Predict Low Risk	Predict High Risk	Correct	Accuracy
Wells	137	73	64	76	61	116	**84.7%**
Geneva	137	94	43	79	58	118	**86.1%**

4. Conclusions

In this paper, a joint extraction model for Chinese electronic medical records is proposed, which can extract the risk factors of venous thrombosis from electronic medical records. Our proposed model achieves the best F1 score on the constructed data sets, and also achieves the fastest inference speed. Finally, the extracted information was used to judge the VET risk level by rule reasoning. In the VTE software system developed and put into clinical use, the prediction results of the model were compared with the evaluation results of clinicians, and the accuracy of the Wells scale and the Geneva scale was 84.7% and 86.1%.

This study had several limitations. First, this is a single-center study. Since a single center is not representative of the Chinese patient population, a multicenter study is needed to verify whether our approach is applicable to a broader Chinese patient population. Second, VTE risk factors are only obtained from text, and part of the information may be in the patient's medical image. It does not do a good job of diagnosing correctly.

Finally, in a future study, we will consider the combination of text and image to diagnose patients from multiple aspects. We will collect and label more data to train the model, and at the same time, we will optimize the model structure in one step to make the model achieve better results.

Author Contributions: Conceptualization, J.H.; methodology, J.C. software, J.Y.; validation, J.C., J.Y. and J.H.; formal analysis, J.H.; investigation, J.C.; resources, J.H.; data curation, J.Y.; writing—original draft preparation, J.C.; writing—review and editing, J.H.; visualization, J.Y.; supervision, J.H.; project administration, J.H.; funding acquisition, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study has received funding from the National Natural Science Foundation of China, No. 82160347; Yunnan Key Laboratory of Smart City in Cyberspace Security, No. 202102AE090031.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of The First People Hospital of Anning City (protocol code No. 2017YYLH035 and date of approval).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: We sincerely appreciate the anonymous reviewers for their precious comments and valuable suggestions. Moreover, we genuinely thank my teacher for his constructive guidance. We sincerely thank Hongjiang Zhang who provided a lot of medical knowledge in the writing process of the paper and gave guidance in the data annotation process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Silverstein, M.D.; Heit, J.A.; Mohr, D.N.; Petterson, T.M.; O'Fallon, W.M.; Melton, L.J. Trends in the Incidence of Deep Vein Thrombosis and Pulmonary Embolism: A 25-Year Population-Based Study. *Arch Intern Med.* **1998**, *158*, 585–593. [CrossRef] [PubMed]
2. Liu, X.; Qiao, A.; Lu, G. Study on comprehensive prevention and management of venous thromboembolism. *Chin. Nurs. Res.* **2015**, *9*, 3124–3127.
3. Zeng, Y.; Si, H.; Wu, Y.; Yang, J.; Zhou, Z.; Kang, P.; Pei, F.; Shen, B. The incidence of symptomatic in-hospital VTEs in Asian patients undergoing joint arthroplasty was low: A prospective, multicenter, 17,660-patient-enrolled cohort study. *Knee Surg. Sport. Traumatol. Arthrosc.* **2019**, *27*, 1075–1082. [CrossRef] [PubMed]
4. Zhang, J.; Liu, Z. Management and intervention of venous thromboembolism in hospital. *Chin. J. Hosp. Ad-Minist.* **2017**, *33*, 51–53.
5. Mubashar, A.; Asghar, K.; Javed, A.R.; Rizwan, M.; Srivastava, G.; Gadekallu, T.R.; Wang, D.; Shabbir, M. Storage and Proximity Management for Centralized Personal Health Records Using an IPFS-Based Optimization Algorithm. *J. Circuits Syst. Comput.* **2021**, *31*, 2250010. [CrossRef]
6. Qatawneh, Z.; Alshraideh, M.; Almasri, N.; Tahat, L.; Awidi, A. Clinical Decision Support System for Venous Thromboembolism Risk Classi-fication. *Appl. Comput. Inform.* **2017**, *15*, 12–18. [CrossRef]
7. Vilhena, J.; Vicente, H.; Martins, M.R.; Grañeda, J.M.; Caldeira, F.; Gusmão, R.; Neves, J.; Neves, J. A case-based reasoning view of thrombophilia risk. *J. Biomed. Inform.* **2016**, *62*, 265–275. [CrossRef] [PubMed]
8. Yang, Y.; Wang, X.; Huang, Y.; Chen, N.; Shi, J.; Chen, T. Ontology-based venous thromboembolism risk assessment model developing from medical records. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 1–13. [CrossRef] [PubMed]
9. Elias, P.; Khanna, R.; Dudley, A.; Davies, J.; Jacolbia, R.; McArthur, K.; Auerbach, A.D. Automating Venous Thromboembolism Risk Calculation Using Electronic Health Record Data Upon Hospital Admission: The Automated Padua Prediction Score. *J. Hosp. Med.* **2013**, *12*, 231–237. [CrossRef] [PubMed]
10. Zhou, S.; Ma, X.; Jiang, S.; Huang, X.; You, Y.; Shang, H.; Lu, Y. A retrospective study on the effectiveness of Artificial Intelligence-based Clinical Decision Support System (AI-CDSS) to improve the incidence of hospital-related venous thromboembolism (VTE). *Ann. Transl. Med.* **2021**, *9*, 491. [CrossRef] [PubMed]
11. Giannis, B.; Johannes, D.; Thomas, D.; Develder, C. Joint entity recognition and relation extraction as a multi-head selection problem. *Expert. Syst. Appl.* **2018**, *114*, 34–45.
12. Eberts, M.; Ulges, A. Span-Based Joint Entity and Relation Extraction with Transformer Pre-Training. In Proceedings of the European Conference on Artificial Intelligence, Santiago de Compostela, Spain, 29 August–8 September 2020.
13. Wang, J.; Lu, W. Two Are Better than One: Joint Entity and Relation Extraction with Table-Sequence Encoders. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Online, 16–20 November 2020; pp. 1706–1721.
14. Wang, Y.; Sun, C.; Wu, Y.; Zhou, H.; Li, L.; Yan, J. UniRE: A Unified Label Space for Entity Relation Extraction. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Online, 1–6 August 2021; pp. 220–231.
15. Zhang, L.; Yin, L.; Ren, L.; Su, J.; Ming, R.; Bai, Y. BSLRel: A Binary Sequence Labeling Based Cascading Relation Triple Extraction. *J. Chin. Inf. Process.* **2021**, *35*, 74–84.
16. Wang, Y.; Yu, B.; Zhang, Y.; Liu, T.; Zhu, H.; Sun, L. TPLinker: Single-stage Joint Extraction of Entities and Relations through Token Pair Linking. In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020.
17. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–8 December 2013; Volume 26.
18. Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep contextualized word representations. *arXiv* **2018**, arXiv:1802.05365.
19. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.