A Comparative Study of Natural Language Processing Algorithms Based on Cities Changing Diabetes' Vulnerability Data

Siting Wang

Tianjin Medical University

Fuman Song

Tianjin Medical University

Qinqun Qiao

Tianjin Medical University

Yuanyuan Liu

Tianjin Medical University

Jiageng Chen (✉ chenjiageng@tmu.edu.cn)

Tianjin Medical University

Jun Ma

Tianjin Medical University

Research Article

Keywords: Diabetes, NLP, ERNIE, BERT

Posted Date: October 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-959766/v1>

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

[Read Full License](#)

Abstract

Background: Diabetes has become a global public health priority resulting in significant workforce losses and health care expenditures. Therefore, research on diabetes vulnerability has become imperative. Current methods for studying disease vulnerability mainly use qualitative research methods represented by Thematic Analysis (TCA), which has the disadvantage of being staff-intensive for long periods of time. Natural Language Processing (NLP) could achieve efficient results in information mining tasks, but we didn't find many studies talking about NLP in non-infectious chronic diseases.

Methods: In this study, hyperparameters were adjusted to obtain more cost-effective model applicable to The Cities Changing Diabetes' vulnerability data by comparing Bidirectional Encoder Representation from Transformers (BERT) and Enhanced Language Representation with Informative Entities (ERNIE) in terms of test accuracy, completion time and evaluation metrics on classification.

Results: The results showed that BERT took less time for the same hyperparameter cases, and the test accuracy of ERNIE was slightly better than that of BERT. We further adjusted the Batch size of ERNIE as we found that ERNIE with the splitting ratio of 8:1:1 and Batch size of 64 had the better efficiency with the test accuracy was 97.67%, the completion time was 12min36s and Macro- F_1 score was 0.9734.

Conclusions: In this study, BERT overwhelmed ERNIE in terms of completion speed with the same hyperparameters. ERNIE showed higher accuracy, especially the ideal performance at the split ratio of 8:1:1 after enhancing the Batch size. From the point of view, we pursue a model with high accuracy and fast processing speed, which means that we can obtain the highest accuracy in the shortest time. It could be selected according to the actual situation in the application process.

Instruction

Diabetes has become the most influenced non-infectious chronic disease (NCD) which is the focal problem of global public health, has developed into "health killer" for mankind. In China, diabetes patients have tripled compared with the past ten years. The direct cost of diabetes in China has climbed over time with the annual cost associated with diabetes estimated at \$51 billion [1, 2]. Since 2013, China has made much effort on medical reform [3], so that diabetes patients can easier obtain medical support. China has gradually shifted the focus of early prevention and treatment of diabetes from large general hospitals to community health service centers, from simple clinical diagnosis and treatment to tertiary prevention of diabetes. The current focus of research in China is on the prevention and treatment effects of diabetes [4–5] and there is still a significant vulnerability in diabetes patients.

The self-management status of diabetes patients could obviously affect the strategy of Chinese public health. It is very important to locate the vulnerability factors of diabetes patients. It took too much time to handle a conversation between the NCD patient and the interviewer. This kind of texts contains a lot of information such as medical history and treatment methods, which is suitable for new-type data mining through Natural Language Processing [6]. The use of specialized language and labels in NLP-based

models can improve classification efficiency [7]. Pre-trained language models have represented the most advanced NLP, including BERT [8], ERNIE [9], XLNet [10], etc.

BERT leads a new wave of NLP and even Deep Learning. It could fully describe the syntactic semantic and other information of a text by mining multi-granularity feature relations [11]. ERNIE was a newer model based on BERT, which both Tsinghua research group and Baidu company chose ERNIE as the pre-trained NLP model name. These ERNIEs had little differences, but they both had same obvious affected in Chinese Language condition. Although ERNIE is a pre-trained model with Chinese corpus, it also outperformed BERT and XLNet in benchmarking the English corpus [12].

TCA was the traditional qualitative research method to locate disease vulnerability. In this study we used the BERT model and ERNIE model to classify The Cities Changing Diabetes' data. We explored the effectiveness of Deep Learning techniques in vulnerability studies and evaluated the BERT and ERNIE approaches. The Chinese government is still helping to find a balance between the interests of health insurance, hospitals, patients and other parties, so our relatively well-screened algorithms will be heavily trained and put to use to make recommendations for implementing standardized diabetes management.

Methods

All the methods we used were carried out in accordance with the relevant guidelines and regulations. Our experimental protocols were approved by the Ethics Committees of Tianjin Medical University. The pre-work divided into 3 parts, including planning, data collection and analysis. During the planning phase, an interview outline was developed. Interviewers were diabetes clinicians and they divided into twenty-six groups by interviewers' hospital. Interviewers received professional training during this phase. During the data collection phase, interviewers took at least 2 hours interview with participants. All participants signed informed consent form before the interview which means that we conducted the interview with the consent of all subjects. During the analysis phase, TCA, which is a systematic method of simplifying data to extract the most salient information from many qualitative interviews, was used to analyze the qualitative data.

Sample Population

To make sure we could gain vulnerable patients as much as possible, an inclusion criterion was developed by expert group. Case filter and its definition were shown in Table 1. All participants were ≥ 18 years old. In the end, a total of 259 documents were received, 229 of which were able to meet the requirements for the next step of data analysis. The final collection of 229 interviews was able to meet the requirements for the next step of data analysis. From the 229 interviews, field staff digitally recorded the interviews and collected demographic and clinical information about the participants.

Table 1
Case filter and its definition

Case filter	Definition
High BMI	BMI>28kg/m ²
High blood glucose levels	FBG>6.1 mmol/L, 2h PBG>7.8 mmol/L
Duration if diabetes/comorbidities	Have co-morbidities
Health insurance	No worker basic insurance, urban basic insurance or commercial health insurance
Employment status	Unemployment
Below poverty level	For urban residence, the per capita income is less than 705 Yuan, for rural residents, the per capita income is less than 540 Yuan
Body size and physical characteristics	Waistline: Male≥90cm, Female≥85cm
Distance between home and work	≥16km
Education background	Primary and illiteracy
Physical activity level	Low (Civil service and no exercise etc.)

Thematic Analysis

After discussed with the Cities Changing Diabetes Tianjin team and experts from the City University of London (UCL), we developed an initial code manual. The demographic and clinical overview of participants were developed alongside the code manual to generate “vulnerability matrix”. The coder coded two or three interviews according to the coding manual, and then the coder coded a transcript from another member of the team to ensure that the manual coding is valid. During coding, we opened several discussions to perfect the code manual. Finally, we identified 12 themes and 25 factors associated with patients’ vulnerability. All themes and factors were shown in Table 2.

Table 2
Themes and factors of vulnerability of diabetic patients in Tianjin

Themes	Factors
Financial constraints	Low income
	Unemployment
	No medical insurance/Low reimbursement ratio
	Significant family expenditure
Severity of disease	Appear symptoms, complications, comorbidities
	Poor disease control
Health literacy	Low literacy
Health beliefs	Perceived diabetes indifferently
	Acquire health knowledge passively
	Distrust of primary health services
Medical environment	Needs not met by medical services
Life restriction	Limited daily life behaviors
	Occupational restriction
Lifestyle change	Adherence to the traditional or unhealthy diet
	Lack of exercise
	Low-quality sleep
Time poverty	Healthcare seeking behaviors were limited by work/taking care of family issues
Mental condition	Appearance of negative emotions towards diabetes treatment or life
Levels of support	Lack of community support
	Lack of support from friends and family
	Lack of social support
Social integration	Low degree of social integration
	Faith in suffering alone
Experience of transitions	Diet transformation
	Dwelling Environment/Place of residence transformation

Sampling

Our study sampled from Cities Changing Diabetes data. The data set was related to “Levels of support” and “Health beliefs”. We compiled 239 sentences and 104 sentences respectively for a total of 343 sentences. The data resource we sampled were too small to affect the pre-training effect of the model. Referring to many researchers who faced similar problems [13], we expanded the number of texts through operations such as synonymous substitution, changing sentence structure, and sentence transcription. At last, we got 899 statements about “Levels of support” and 400 statements about “Health beliefs”.

BERT

The BERT model has “Bidirectional Transformer Mechanism”. This mechanism considers the semantic information implied in the context and can adequately extract features from long and complicated sentences [8]. It uses two unsupervised approaches including Masked Language Model (MLM) and Next Sentence Prediction (NSP) to jointly pre-train. The former type of MLM can perform random masking on 15% of the words in a sentence, and then use the context to predict the content of the masking. NSP is to determine the contextual relationship by predicting the coherence of the contextual sentence. BERT is an advanced pre-trained word embedding model based on transformer coding architecture [14] and its result output is a vector or multiple vectors.

The hidden layer, Transformer layer and algorithm input layer in the BERT model rely on the activation function to connect the three parts and the most common activation function is Sigmoid and ReLU (Rectified Linear Unit). In order to improve the convergence speed, alleviate the gradient disappearance problem caused by Sigmoid, and improve the computational efficiency, we choose the ReLU function in this study. ReLU is a segmented linear function where the input has the same values as the output if the input is positive, otherwise the output is zero. The sparse model implemented by ReLU can be more effective in mining target-related features and fitting the training data. The vectorization of the text is implemented in the input layer of the algorithm, which consists of three components: word vector, text vector and position vector. The core of Transformer is the Attention Mechanism, which is to make the vector corresponding to each word in each sentence incorporate the information of all words in that sentence [8]. Just take out a vector corresponding to the token, and it contains the information of the whole sentence. SoftMax classification in the Transformer layer is a multi-classification algorithm that maps inputs to probabilities that are favorable for classification.

ERNIE

Inspired by the BERT masking strategy, ERNIE was designed to enhance learning language representations through Knowledge Masking Strategies, including base level Masking, entity Masking, and phrase Masking [9]. The model consists of two main layers. The first layer is the lower text encoder which is responsible for capturing the basic vocabulary and information from the input tokens. Another layer was the upper knowledge encoder which is responsible for integrating the knowledge information into the text information in order to represent the heterogeneous information of tokens and entities into a unified feature space. ERNIE treats a phrase or an entity as a unit which usually consists of several words. During word representation training, all words in the same units were Masked instead of just one

word or character. ERNIE doesn't add knowledge embedding directly, instead, it implicitly learns knowledge and longer with semantic dependency information, which is used to guide word embedding learning. The first learning stage is basic level Masking. As with BERT, ERNIE randomly masks 15% of the basic language units and trains a Transformer to predict the mask units using the other basic units in the sentence as input. Basic word representations are available at this stage. The second learning stage is phrase Masking stage, which is unique to ERNIE unlike BERT. ERNIE uses basic linguistic units as training input, and it masks and predict all basic units in the same phrase for a random selection of several phrases in the sentence. The third learning phase is the entity Masking phase which is unavailable in BERT. Named entities can be both abstract and actual. After the three stages of Basic Level Masking, Entity Masking and Phrase Masking, we can obtain a more semantically informative representation of words.

Pre-training Processing

Before the training started, we deleted meaningless words in sentences, such as unnecessary tone words and responses (e.g., um, ah, uh, huh, etc.). We did this in order to improve the learning efficiency of the model. The final 1299 statements were compiled and stored in a Linux, UTF-8 BOM Txt file. In the second step, the data sets were divided into training, validation and testing dataset in the ratio of 8:1:1, 7:2:1 and 6:3:1. In the third step, the relevant hyperparameters were set. Batch size is the number of samples trained in each Batch. Epoch refers to the process of propagating a complete data set once in the forward and once in the reverse direction through a neural network. In this study, the hyperparameters of both models were set to the same values. The fine-tuned hyperparameters of the BERT and ERNIE models were shown in Table 3.

Table 3
Fine-tuned hyperparameters of BERT and ERNIE models

Name of the hyperparameters	Hyperparameters Value
Hidden size	768
Learning rate	5e ⁻⁵
Pad size	16
Require Improvement	1000
Epoch	100
Batch size	32

Performance metrics

In this study, we used the confusion matrix (Table 4) and its derived evaluation metrics including precision, recall, F_1 score, and test accuracy for comprehensive comparisons of classifier performance. One side of the confusion matrix represented the actual class of the dataset, and the other side represented the classification of the dataset by the classifier.

Table 4
Confusion matrix of testing dataset

Predicted class			
		Health beliefs	Level of support
Actual class	Health beliefs	True positive (TP)	False negative (FN)
	Level of support	False positive (FP)	True negative (TN)

- Using this mixed matrix as an example, identifying “health beliefs” as positive result and “Levels of support” as negative result. To avoid duplication of presentation, we only show the matrix with health beliefs as positive outcomes.
- In the process of calculating the relevant evaluation index of a factor, this factor is used as the positive result.
- Cells in bold font indicated the correct classification of the model for the dataset.

The test accuracy represented the proportion of the total number of cases correctly predicted by the pre-trained model.

$$TestAccuracy = \frac{TP + TN}{P + N}$$

1

Precision indicated the number of correctly classified cases as a percentage of the total number of cases classified into this class.

$$Precision = \frac{TP}{TP + FP}$$

2

Recall measured the classifier completeness which represented the percentage of those who correctly predicted positive result as a percentage of all those who were positive result support.

$$Recall = \frac{TP}{TP + FN}$$

3

$$F_{\beta} = \frac{\beta^2 + 1 \times Precision \times Recall}{Precision + Recall}$$

4

β as a parameter to adjust the weight between Precision and Recall. Generally, in the actual test we default Precision and Recall equally important and set it to 1.

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

5

To prevent the differences between the F_1 scores of various classes being difficult to distinguish, we calculated macro- F_1 scores for comparison.

$$Macro - F_1 = \frac{F(\text{Health beliefs}) + F(\text{Level of support})}{2}$$

6

Results

Table 5
Comparison of models' confusion matrix

Name	Batch size	Splitting ratio	Confusion matrix	
BERT	32	8:1:1	Actual class	Predicted class
			Health beliefs	Level of support
			86	2
			Level of support	3
				38
		7:2:1	Actual class	Predicted class
			Health beliefs	Level of support
			90	1
			Level of support	3
				37
		6:3:1	Actual class	Predicted class
			Health beliefs	Level of support
			83	0
			Level of support	5
				43
ERNIE	32	8:1:1	Actual class	Predicted class
			Health beliefs	Level of support
			86	2
			Level of support	1
				40
		7:2:1	Actual class	Predicted class
			Health beliefs	Level of support
			90	1
			Level of support	2
				38
		6:3:1	Actual class	Predicted class
			Health beliefs	Level of support
			81	2
			Level of support	2
				46
	64	8:1:1	Actual class	Predicted class
			Health beliefs	Level of support

Name	Batch size	Splitting ratio	Confusion matrix	
	7:2:1	Health beliefs	86	2
		Level of support	1	40
		Actual class	Predicted class	
			Health beliefs	Level of support
		Health beliefs	90	1
		Level of support	4	36
	6:3:1	Actual class	Predicted class	
			Health beliefs	Level of support
		Health beliefs	81	2
		Level of support	4	44

Table 6
Text Acc and completion time of models

Name	Batch size	Splitting ratio	Test Acc(%)	Completion time
BERT	32	8:1:1	96.12	0:07:38
		7:2:1	96.95	0:07:46
		6:3:1	96.18	0:09:27
ERNIE	32	8:1:1	97.67	0:35:36
		7:2:1	97.71	0:48:31
		6:3:1	96.95	1:29:05
	64	8:1:1	97.67	0:12:36
		7:2:1	96.18	0:10:16
		6:3:1	95.42	0:08:45

Table 5 and Table 6 showed the confusion matrixes, test accuracies and completion time of the two models with different splitting ratios. Text classification was fundamentally a mapping process, and the ideal state was to point all correct results to the correct set. The confusion matrix presented the classification results directly and explicitly, and the extended performance metrics could help us to select a better model. With the same Batch sizes, ERNIE showed better test accuracy in all three splitting ratios, especially in the splitting ratio of 7:2:1 the test accuracy was 97.71%. BERT cost much shorter time in same Batch size, and the test accuracy reaching over 96%. The splitting ratio of 7:2:1 was more suitable for both models, and the models showed relatively better performance both in test accuracy and in completion time with Batch size of 32.

Table 7
Comparison of performance metrics by class

Name	Batch size	Splitting ratio		Precision	Recall	F_1	Macro- F_1
BERT	32	8:1:1	Health beliefs	0.9663	0.9773	0.9718	0.9551
			Level of support	0.9500	0.9268	0.9383	
		7:2:1	Health beliefs	0.9677	0.9890	0.9783	0.9635
			Level of support	0.9737	0.9250	0.9487	
		6:3:1	Health beliefs	0.9432	1.0000	0.9708	0.9580
			Level of support	1.0000	0.8958	0.9451	
ERNIE	32	8:1:1	Health beliefs	0.9885	0.9773	0.9829	0.9734
			Level of support	0.9524	0.9756	0.9639	
		7:2:1	Health beliefs	0.9783	0.9890	0.9836	0.9728
			Level of support	0.9744	0.9500	0.9620	
		6:3:1	Health beliefs	0.9759	0.9759	0.9759	0.9671
			Level of support	0.9583	0.9583	0.9583	
ERNIE	64	8:1:1	Health beliefs	0.9885	0.9773	0.9829	0.9734
			Level of support	0.9524	0.9756	0.9639	
		7:2:1	Health beliefs	0.9574	0.9890	0.9730	0.9541
			Level of support	0.9730	0.9000	0.9351	
		6:3:1	Health beliefs	0.9529	0.9759	0.9643	0.9503
			Level of support	0.9565	0.9167	0.9362	

Table 7 showed the performance metrics of the two models for each class under different conditions. Under the same Batch size, ERNIE showed higher F_1 scores in both classes compared with BERT in the same splitting ratio, especially the F_1 score of “Health beliefs” was 0.9836 when the splitting ratio was 7:2:1. After increasing the Batch size, we found that ERNIE's Precision, Recall and F_1 score decreased in

the splitting ratio of 7:2:1 and 6:3:1. Since ERNIEs with different Batch sizes exhibit the same confusion matrix with the splitting ratio of 8:1:1, they had the same values of Precision, recall and F_1 score.

Combining the evaluation metrics of Table 6 and Table 7, the optimal model could be selected by comprehensive consideration. From the application point of view, what we pursue was to obtain the highest performance metrics' score in the shortest time. In this study, after choosing, we considered that ERNIE with the splitting ratio of 8:1:1 and Batch size of 64 had the better efficiency with the test accuracy was 97.67%, the completion time was 12min36s and Macro- F_1 score was 0.9734.

Discussion

In this study, Hyperparameters were adjusted to obtain an efficient model applicable to The Cities Changing Diabetes' vulnerability data by comparing two pre-trained models in terms of test accuracy, completion time and evaluation metrics on classification. Both models achieved excellent performance in performing the classification task of The Cities Changing Diabetes' vulnerability data.

According to our result, the ERNIE model with the splitting ratio of 7:2:1 had the best test accuracy, but it took longer to pre-train compared to the BERT model. Checking the literature [15], it was found that within a reasonable range, an appropriate increase in Batch size can make the descent direction more accurate. Therefore, this study further investigated the performance comparison of ERNIE under different Batch size, hoping to shorten the training time by adjusting the Batch size. Due to the GPU (Graphics Processing Unit) memory limitation, we could only adjust the Batch size to 64 for comparison with the previous results. It was obvious that the pre-training time was significantly reduced by increasing the Batch size. On the contrary, we also discovered that as the Batch size increases, the test accuracy of ERNIE decreased by at least 1.5% for the splitting ratios of 7:2:1 and 6:3:1. ERNIE with the splitting ratio of 8:1:1 in different Batch sizes exhibited the same confusion matrixes coincidentally, so the final test accuracy and evaluation metrics are all the same. Both showed the same excellent performance, while the ERNIE model with Batch size of 64 takes only one third of the time of the ERNIE model with Batch size of 32.

NLP technology has been successfully applied in the medical field for electronic health records, identifying post-operative complications from doctors' notes after inpatient surgery, classifying patients by identifying syndromes, and many other fields [16–19]. The emergence of Transformer has propelled NLP into the golden age, and BERT and ERNIE are both products of the transformation of Transformer. Transformer is built entirely on a self-attentive mechanism, which not only allows parallel operations but also captures long-range feature dependencies [20]. In the classification task of Chinese eligibility criteria sentences, ERNIE narrowly beat BERT's results in both micro and macro F_1 score [21], and it exhibited higher accuracy under the same conditions which was consistent with the results of this study. The results of the final hidden layer computation we can perform downstream tasks by changing the pre-trained model parameters which is also called fine-tuning. Batch size determines the time required to complete each Epoch and the smoothness of the gradient between each Iteration in the deep learning training process. The choice of Batch can determine the descent direction. In this study, the training time

was shortened by adjusting the Batch size, and ERNIE achieved our expected results with excellent efficiency when the splitting ratio was 8:1:1.

In our dataset with the study population, we found that there were some patients who did not care about diabetes, and they did not follow the advice of their doctors. The expenditure caused by diabetes has a serious impact on China's health budget [22]. Diabetes prevention and self-management of existing diabetes is critical to reducing healthcare expenditures [23]. Currently, health education relies heavily on healthcare professionals [24], but with limited manpower and time, it cannot accommodate the demands of educating the increasing number of patients with diabetes. A recent study [25] showed that knowledge of diabetes is an important influence on self-management in terms of blood glucose monitoring and an important influence on quality of life for patients with diabetes. Therefore, our study was necessary and needed to be carried out.

It is worth noting that our study has several limitations. First, due to the age of most participants, there may be semantic bias caused by unclear expressions. Second, low sample size caused the training data set to be too small, and the small training set was greatly affected the effectiveness of the pre-trained model. So, we expanded the number, but probably changed the original meaning of some sentences in the process and affected the learning effect of the model. In the future, we will further collate the vulnerability data of Cities Changing Diabetes (Tianjin) and expand it into the dataset to further improve the pre-training model. It is expected that the dataset will be further expanded vertically and deeply through cooperation with other cities in this project. Finally, due to the limitations of the research device GPU, we only explored the performance comparison of the ERNIE model with Batch size of 64 and Batch size of 32.

Conclusions

In the classification task of data on the vulnerability of Cities Changing Diabetes, we trained BERT and ERNIE, and evaluated the effectiveness of the two models by test accuracy, precision, recall, F_1 score and completion time calculated by confusion matrix to explore the advantages and shortcomings of the two models. From the point of view that we pursue a model with high accuracy and fast processing speed, which means that we can obtain the highest accuracy in the shortest time. In this study, the accuracies of prediction were similar, and the time used for BERT was shorter for the same hyperparameters, but ERINE had a better application when the Batch size was increased to 64. Therefore, it could be selected according to the actual situation in the application process.

Abbreviations

BERT	Bidirectional Encoder Representation from Transformers
GPU	Graphics Processing Unit
ERNIE	Enhanced Language Representation with Informative Entities
MLM	Masked Language Model
NLP	Natural Language Processing
NSP	Next Sentence Prediction
ReLu	Rectified Linear Unit
TCA	Thematic Analysis

Declarations

Ethics approval and consent to participate

Our research data has been ethically reviewed. The consent we received from the study participants was written. The study was approved by the Ethics Committees of Tianjin Medical University.

Consent for publication

Not applicable.

Availability of data and materials

The data cannot be shared publicly due to ethical reasons. The data contain potentially identifiable information of study participants. For example, participants stated their living location during the conversation, and thus could easily be located. Data are available upon request to Chen Jiageng(chenjiageng@tmu.edu.cn), Tianjin Medical University.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

SW and FS contributed to develop the main idea, data curation, experiment and write the manuscript. QQ contributed to data curation and formal analysis. Dr. J.C contributed to methodology, experiment, formal analysis and revised the manuscript. Prof. J.M and Dr. Y.L supervised the data analysis and coordinated the overall study. All authors reviewed the manuscript.

Acknowledgement

Not applicable.

Author details

¹School of Public Health, Tianjin Medical University, Tianjin 300070, China

References

1. Li HF, Cai L, Golden AR. Short-term trends in economic burden and catastrophic costs of type 2 diabetes mellitus in rural southwest China. *J Diabetes Res*. 2019;2019:9626413.
2. Chen C, Wang L, Chi HL, Chen W, Park M. Comparative efficacy of social media delivered health education on glycemic control: A meta-analysis. *Int J Nurs Sci*. 2020;7(3):359–368.
3. Jia W, Tong N. Diabetes prevention and continuing health-care reform in China. *Lancet Diabetes Endocrinol*. 2015;3(11):840–842.
4. Wang J, Ma Q, Li Y, Li P, Wang M, Wang T, et al. Research progress on Traditional Chinese Medicine syndromes of diabetes mellitus. *Biomed Pharmacother*. 2020;121:109565.
5. Wu Z, Jin T, Weng J. A thorough analysis of diabetes research in China from 1995 to 2015: current scenario and future scope. *Sci China Life Sci*. 2019;62(1):46–62.
6. Yamada M, Saito Y, Imaoka H, Saiko M, Yamada S, Kondo H, et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci Rep*. 2019;9(1):14465.
7. Beard EJ, Sivaraman G, Vázquez-Mayagoitia Á, Vishwanath V, Cole JM. Comparative dataset of experimental and computational attributes of UV/vis absorption spectra. *Sci Data*. 2019;6(1):307.
8. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Minneapolis: ACL; 2019. p. 4171–86.
9. Sun Y, Wang S, Li Y, Feng S, Chen X, Zhang H, et al. ERNIE: Enhanced representation through knowledge integration. *CoRR*. 2019;abs/1904.09223.
10. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Quoc V. Le. XLNet: Generalized autoregressive pretraining for language understanding. *CoRR*. 2019;abs/1906.08237.
11. Lai S, Xu L, Liu K, Zhao J. Recurrent convolutional neural networks for text classification. *Proc Conf AAAI Artif Intell*. 2015;2267–2273.
12. Sun Y, Wang S, Li Y, Feng S, Tian H, Wu H, et al. ERNIE 2.0: A continual pre-training framework for language understanding. *CoRR*. 2019;abs/1907.12412.
13. Jason W, Kai Z. EDA: Easy data augmentation techniques for boosting performance on text classification tasks. *CoRR*. 2019;abs/1901.11196.

14. Kaliyar, R.K., Goswami, A. & Narang, P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimed Tools Appl* 80. 2021;11765–11788.
15. Wei X, Peng Y, Zhang T. Research on medical entity extraction model based on the BERT multi-layer network. *China Digital Medicine*. 2021;16(05):36–40.
16. Chary M, Parikh S, Manini AF, Boyer EW, Radeos M. A review of natural language processing in medical education. *West J Emerg Med*. 2019;20(1):78–86.
17. Kaufman DR, Sheehan B, Stetson P, Bhatt AR, Field AI, Patel C, et al. Natural language processing-enabled and conventional data capture methods for input to electronic health records: A comparative usability study. *JMIR Med Inform*. 2016;4(4):e35.
18. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. *J Biomed Inform*. 2018;77:34–49.
19. Zheng L, Wang Y, Hao S, Shin AY, Jin B, Ngo AD, et al. Web-based real-time case finding for the population health management of patients with diabetes mellitus: A prospective validation of the natural language processing-based algorithm with statewide electronic medical records. *JMIR Med Inform*. 2016;4(4):e37.
20. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, et al. Attention is all you need. *CoRR*. 2017;abs/1706.03762.
21. Zong H, Yang J, Zhang Z, Li Z, Zhang X. Semantic categorization of Chinese eligibility criteria in clinical trials using machine learning methods. *BMC Med Inform Decis Mak*. 2021;21(1):128.
22. Hu S, Su X, Deng X, Wang Y. Exploring the appropriate price of semaglutide for type 2 diabetes patients based on cost-utility analysis in China. *Front Pharmacol*. 2021;12:701446.
23. Buis LR, Hirzel L, Turske SA, Des Jardins TR, Yarandi H, Bondurant P. Use of a text message program to raise type 2 diabetes risk awareness and promote health behavior change (part II): Assessment of participants' perceptions on efficacy. *J Med Internet Res*. 2013;15(12):e282.
24. Zhao X, Yu X, Zhang X. The role of peer support education model in management of glucose and lipid levels in patients with type 2 diabetes mellitus in Chinese adults. *J Diabetes Res*. 2019;2019:5634030.
25. Kueh YC, Morris T, Ismail AA. The effect of diabetes knowledge and attitudes on self-management and quality of life among people with type 2 diabetes. *Psychol Health Med*. 2017;22(2):138–144.