WILEY | Hindawi

Retraction

Retracted: An Intelligent Diagnosis System for English Writing Based on Data Feature Extraction and Fusion

Complexity

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. He, "An Intelligent Diagnosis System for English Writing Based on Data Feature Extraction and Fusion," *Complexity*, vol. 2021, Article ID 4960893, 7 pages, 2021.

Research Article

An Intelligent Diagnosis System for English Writing Based on Data Feature Extraction and Fusion

Yizhou He (ID)

School of Foreign Languages, Xinyu University, Xinyu 338004, Jiangxi, China

Correspondence should be addressed to Yizhou He; 06122721@cumt.edu.cn

English writing is conducive to the online communication and communication of language; the current diagnosis system of English writing is difficult to accurately find and diagnose the wrong words, which leads to a low diagnosis rate of wrong words in English writing system. To solve this problem, this paper designs an intelligent diagnosis system for English writing based on data feature extraction and fusion. First of all, B/S architecture is introduced on the basis of the conventional intelligent diagnosis system structure of English writing, which makes up for the problem that the C/S mode is prone to diagnostic errors. Secondly, the features of English lexical data are extracted and fused to provide better input for the diagnostic model, which effectively solves the problems of complex vocabulary and feature redundancy in English writing. The simulation results show that the proposed intelligent diagnosis system for English writing has higher diagnostic accuracy and faster query speed.

1. Introduction

With the globalization of economy and trade, the communication between people is becoming more and more close and frequent. Therefore, the establishment of English writing system is very necessary to help improve the speed of information exchange and promote the smooth degree of communication. As English writing system is common both in school and in all kinds of English training classes, the performance of different writing systems is also uneven. Therefore, it is important to improve the model performance of the existing English writing system. Error diagnosis is one of the key techniques to improve the efficiency of English writing system. Therefore, how to establish an English writing system with better diagnostic performance [1] is a problem worth studying.

After investigation and analysis, it is found that the most common English writing system is based on hardware. However, with the increase of music data, the efficiency of vocabulary diagnosis in English writing system is low, the error rate is high, and it is easy to fall into local minimum problem [2]. Therefore, the existing framework of English writing system is optimized, and the method of data feature extraction and fusion of English writing intelligent diagnosis

system is proposed, which accelerates the speed of the system, thus improving the recognition effect of words and the accuracy of diagnosis. This kind of new intelligent diagnosis system for English writing first needs to extract the features of English words, which can describe the important information contained in the words [3] and effectively identify whether the words are correct or not.

In order to improve the diagnosis effect of English vocabulary, an intelligent diagnosis system for English writing based on feature extraction and data fusion is proposed. The simulation results are compared with the existing English vocabulary diagnosis models, which show that the proposed intelligent diagnosis system for English writing has obvious advantages in running speed and high diagnosis accuracy. This article first features extraction experiments on the characteristics of the English vocabulary. The four extracted features were used in combination with the feature fusion strategy based on Fisher criterion, and experiments were designed for analysis and comparison.

English writing error intelligent diagnosis system is the use of corpus and cloud computing technology on English writing error information intelligent diagnosis system. Its high-tech research results, such as computing, artificial intelligence technology and statistical natural language

processing technology, will be used in the intelligent diagnosis system of English writing errors. However, at present, the most common intelligent diagnosis models of English writing errors are hidden Markov model, neural network model, and hidden Markov model. The hidden Markov model belongs to a kind of technology of linear classification. The idea is that the relationship between lexical types and feature sets is a simple linear relationship. However, this is not the case. Because of the nonlinear relationship between parts of speech and feature sets, the hidden Markov model is not applicable and the vocabulary diagnosis results are not ideal [5]. It belongs to the nonlinear neural network diagnosis method and can effectively fit the relationship between parts of speech and feature sets. Nevertheless, neural network has some limitations. For example, when the number of vocabulary samples is large or small, the English vocabulary diagnosis model is prone to overfitting or underfitting problems [6–8], which makes the vocabulary diagnosis system difficult to reach the convergence state and requires too long time for vocabulary diagnosis.

The research on English vocabulary diagnosis system mainly focuses on the following two categories: first, English vocabulary diagnosis system based on hardware design; second, English vocabulary diagnosis system based on software design of the hardware aspects: English Writing Intelligent Error Diagnosis System can use a variety of architectures, and quickly return to the English Writing Error Evaluation Server, using C/S mode to achieve diagnosis. It also provides web client for college English writing error information query and subsequent information correction function. Its main advantages are as follows: (1) the network resource utilization rate is high, which can make full use of the operation functions of the equipment at both ends and store the various complex operations of the diagnosis system and the relevant data of English writing error diagnosis in a scientific and reasonable way at both ends of the server according to different requirements. (2) The network resource utilization rate is high, which can make full use of the operating functions of the devices at both ends and make the various complex operations in the operation of the diagnosis system [9]. In terms of software design, the working ideas of the existing intelligent diagnosis system for English writing errors are as follows: the scoring standard of each English writing error test point is set up in advance; open the documents after English writing and standard documents [10] and extract the standard attribute value containing the same writing error point and compare it with the user's English writing document. If it is the same, it will pass the diagnosis; if it is different, it will not pass the diagnosis and need to return to modify [11]. This kind of method has the advantage of being more intelligent and can automatically diagnose and identify English words without too much human intervention, which improves the teaching efficiency. The common lexical diagnosis algorithms in this kind of software method system have already been proposed [12–16]. Among many lexical diagnosis algorithms, finding the one that is most consistent with the current lexical features can significantly improve the accuracy of English vocabulary diagnosis.

Although many diagnostic methods can be used to diagnose English vocabulary, the current diagnostic effect of these methods is not very ideal [17–19], and there are also defects such as high error rate of vocabulary diagnosis and imperfect diagnostic system. This is mainly because at the level of music feature extraction, the feature classification ability extracted by the classification method is nonideal, and the information of the vocabulary cannot be fully used for diagnosis or classification. Hence, the music classification and retrieval function can be better improved by finding an efficient feature extraction and classification method. The improved quantum evolutionary algorithm based on allocation optimization scheme coevolutionary strategy and enhanced particle swarm optimization is mostly used to improve the parallel writing effect [20]. Fuzzy SVM algorithm [21] and keyword recognition [22] are also the first issues to be considered in English writing, which can enhance the accuracy of English writing.

Therefore, this paper proposes an intelligent diagnosis system for English writing based on data feature extraction and fusion, which can combine the advantages of hardware and software at the same time. Firstly, an intelligent diagnosis system of English vocabulary based on B/S structure is proposed on hardware, which ensures the timely updating of database and provides sufficient model input for software algorithm. In the aspect of software design, the feature extraction method is used to extract and fuse the key features contained in the vocabulary, which avoids the interference of irrelevant and redundant words and greatly improves the accuracy of the diagnosis system and the running speed of the model.

2. Materials and Methods

2.1. The Overall Structure Design of the System. The diagnostic system constructed in this paper is mainly based on the B/S structure design, the overall functional structure of the system is designed to achieve a diagnostic function for English writing errors under specific conditions, and the diagnostic results are presented to users intuitively through the software functions [23]. Users of the system can understand the functions covered by the software. The overall structure of English writing error diagnosis system is shown in Figure 1. The system mainly includes hardware part and software part. The hardware design includes system user login, user management, user rights management, and data acquisition module; the software includes data preprocessing, feature extraction, feature fusion, and vocabulary diagnosis.

2.2. Hardware Design. English writing error intelligent diagnosis system can use a variety of systems architecture English writing error assessment and analysis in the form of centralized organized examination, considering the large amount of data of operation types of questions, it is necessary to return to the English Writing Error Evaluation Server in time and use C/S mode to realize diagnosis. In
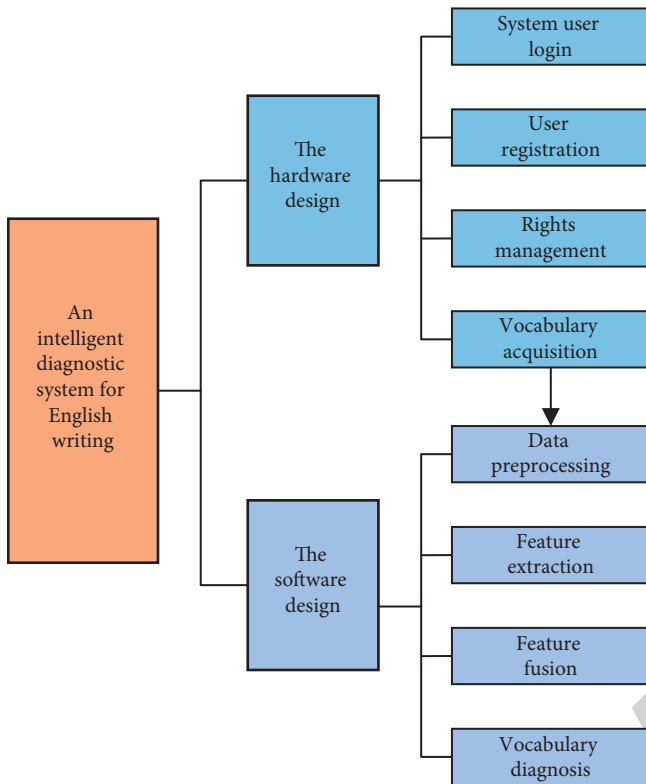
FIGURE 1: Functional structure of diagnosis for English writing system.

addition, the C/S mode of providing Web is mainly divided into client and server, and the system adopts C/S mode [24].

As can be seen from Figure 2, the intelligent diagnosis process of English writing errors focuses on the server side of the system, forming a three-tier architecture of intelligent diagnosis browser, intelligent diagnosis web server, and intelligent diagnosis data baser, among which intelligent diagnosis web server is mainly responsible for dealing with the intelligent error diagnosis in English writing of logic and is also equipped with a dedicated server port to manage the diagnostic system database as well as the logic of data processing; it is mainly used for reducing system client computer work load, can reduce the overall economic costs of users at the same time, and can reduce the work load [25]. One point that needs special attention is that because the intelligent diagnostic program and data of English writing errors are mainly loaded on the relevant servers, if the system server conflicts and causes the operation to terminate, the consequences cannot be predicted, and the vocabulary existing in the English writing errors that need to be diagnosed shall be diagnosed in a comprehensive way.

2.3. English Vocabulary Diagnosis Based on Features Extraction and Fusion. After the requirements of vocabulary storage and update are met by hardware design, it is necessary to extract the features of a large number of vocabulary data stored and realize intelligent diagnosis of English vocabulary. English vocabulary type can be

divided by its characteristics, and in the actual modeling process, English vocabulary diagnosis is a typical classification problem, so effective feature extraction is necessary; it will determine the performance of vocabulary diagnosis. At present, the features of English vocabulary are usually modeled and analyzed, and the feature extraction of information is limited, which cannot fully describe the types of vocabulary. Therefore, to extract various features of vocabulary for diagnosis and classification, the first step is to obtain data of English vocabulary. Because English vocabulary is discrete, English vocabulary enters the frame. Meanwhile, to better extract the lexical characteristics, lexical signals need to be boosted, as given in Figure 3. This framework includes four parts: data preprocessing module, feature extraction, multifeature fusion, and naive Bayesian model (NBM) pattern recognition module.

Data preprocessing module: according to the data and the corresponding calculation, English vocabulary words label the size of the coefficient of Pearson relationship between them and find the vocabulary rich information of the sample interval. If all the original vocabulary data are used as the input of the model, it will increase, which will lead to complex calculation of the model and contain a lot of interference information, which is not conducive to the improvement of the model performance. The time window method is used to find the data samples with the most abundant information and extract the more representative data features and lexical patterns according to the Pearson relation coefficient between the calculated data matrix and the corresponding labels.

Feature extraction: extracting features of English words including time domain and frequency domain. In order to make full use of the nonlinearity and complexity of vocabulary data, statistics and the number of keywords appearing in each example will automatically display a single letter, leaving only the letter string as the eigenvalue, and the statistical eigenvalue and the number of occurrences in the data will be displayed according to the eigenvalue of each order. Compared with dictionary feature extraction, Dictionary is a typical data input word, such as inputting three dictionaries, each dictionary with two eigenvalues, and finally three samples of two eigenvalues, which will be position variables, corresponding values and text feature extraction are input text to eliminate a character, and then corresponding to each eigenvalue appears several times, such as dictionary feature extraction.

Multifeature fusion module: a feature fusion method based on Bayesian theory is adopted. According to the Bayesian theory of English lexical features, the unknown samples are composed of the real value features of the dimensions. According to the Bayesian decision theory with the minimum error rate, if the samples are divided into the second class, this class is the feature class with the highest posterior probability under the condition of known samples, and the result of multifeature fusion can be obtained.

Lexical classification module: using NBM for classification and identification. Bayesian classification is a very simple classification algorithm; which is called
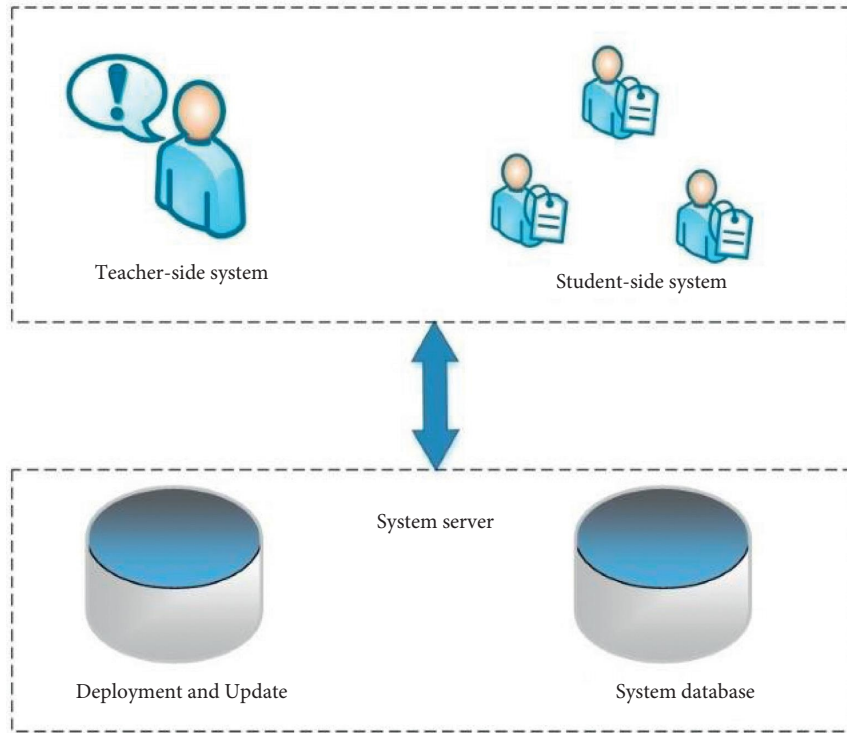
FIGURE 2: Functional structure of system.

Tianbayesian classification, because this method is very simple. The basic idea of simple Bayesian ideology is classification, which solves the probability of each category, which category is the largest. The Bayesian algorithm calculates $P(B|A)$ by three probabilities of the occurrence of $P(A|B)$, $P(A)$, and $P(B)$, assuming that we now know $P(A|B)$, $P(A)$ (the probability of occurrence of event A), and $P(B)$ (the probability of occurrence of event B). How to calculate $P(B|A)$? With the previously known probabilities and $P(A|B)$ probabilities, probability $P(B|A)$ is the probability of occurrence of B under the condition that event A occurs, so $P(B|A)$ is expressed as the intersection of event A and event B divided by $P(A)$,

$$P(B|A) = \frac{P(B \cap A)}{P(A)}. \tag{1}$$

The formula can be converted to the following form:

$$P(B \cap A) = P(B \cap A) \times P(A). \tag{2}$$

To this point, we only need to prove that $P(A\ B) = P(A, B)$ can prove the known $P(A|B)$ which can be calculated under the condition of getting the probability $P(B|A)$. We will put the probability tree into the probability table, listed $P(A|B)$, respectively, $P(B|A)$, $P(A)$, and probability $P(B)$. Through calculation, we can prove that $P(A|B) * P(B)$ and $P(B|A) * P(A)$; the last obtained result is the value of the probability of the same area, so

$$P(B \cap A) = P(A \cap B). \tag{3}$$

Further, the known $P(B|A)$, $P(A)$, and $P(B)$ are the probabilities of $P(A|B)$:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{4}$$

3. Results and Discussion

In order to verify the stability of the system designed in this paper in diagnosis, the system in this paper is applied to test the diagnosis and comprehensive diagnosis of English writing vocabulary in different network environments. The test results are shown in Figure 4.

To analyse Figure 4 in the different network environment, this article system for English writing words diagnosis and comprehensive diagnosis contrast curve change is not big, the basic linear growth; this article system at a constant speed in a different environment to be able to conduct English writing words diagnosis experiments proved that this system has good stability. To further demonstrate the accuracy of the proposed method in English vocabulary diagnosis, the platform of simulation experiment in this paper is firstly introduced, as shown in Table 1.

In addition, to better test the diagnostic effect of multifeature fusion and algorithm on English vocabulary, different types of vocabulary datasets were selected as experimental subjects. A total of 10 kinds of phonetic English vocabulary were collected, the first six categories are content words and the last four categories are function words. The specific number of English vocabulary samples is given in Table 2.

The pure hardware method of the model in this paper and the conventional NBM model are respectively to train the 10 kinds of English vocabulary in Table 1 and then
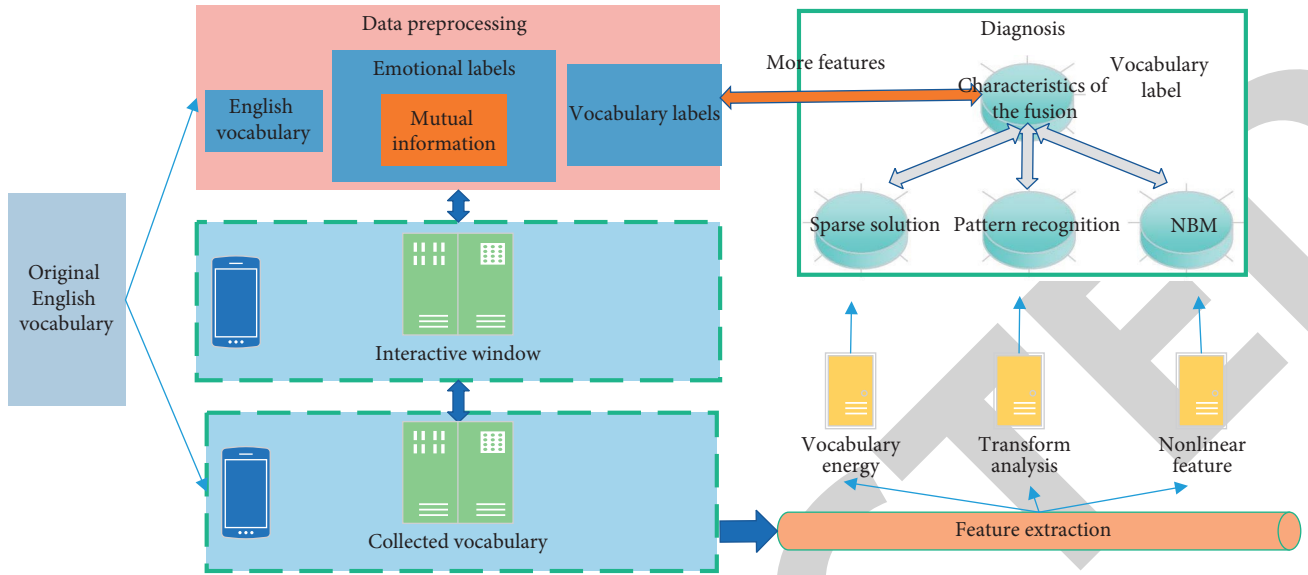
Figure 3: English writing diagnosis based on data feature extraction and fusion.
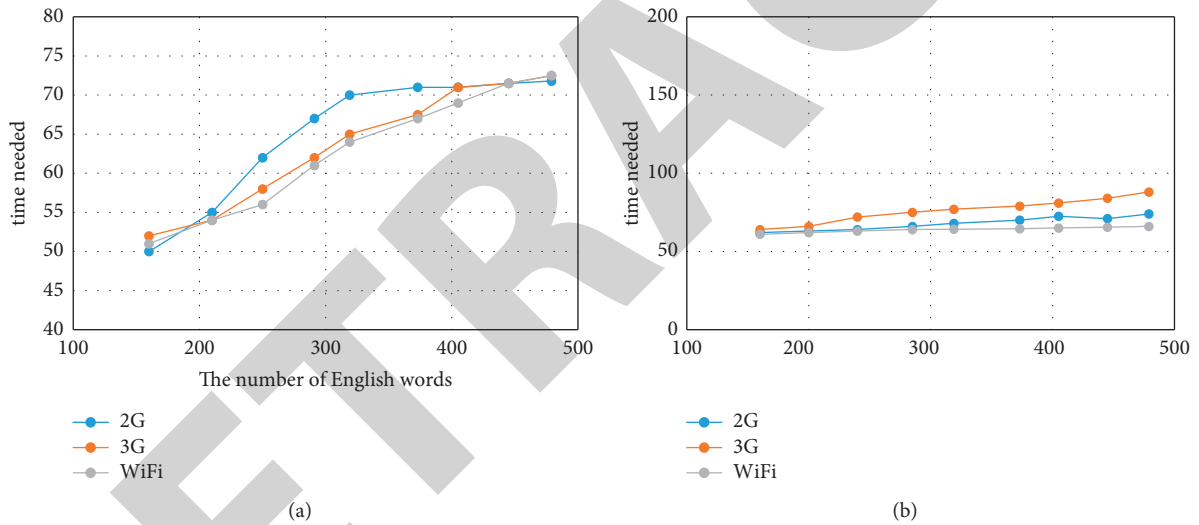

(a)

(b)

Figure 4: Time consumption comparison for different forms of diagnosis of English writing vocabulary.

Table 1: English vocabulary diagnostic experimental setting.

Main parameters	Design
CPU	6 Core 2.80 GHz
Memory size	64 GB
Programming language	MATLAB 2016
System	Windows 8
Sound card	Sound Blaster Z

Table 2: Sample distribution of 10 types of English vocabulary.

Serial number	Name of the music	Sample size
1	Noun	250
2	Pronoun	320
3	Adjective	210
4	Adverb	265
5	Verb	98

TABLE 2: Continued.

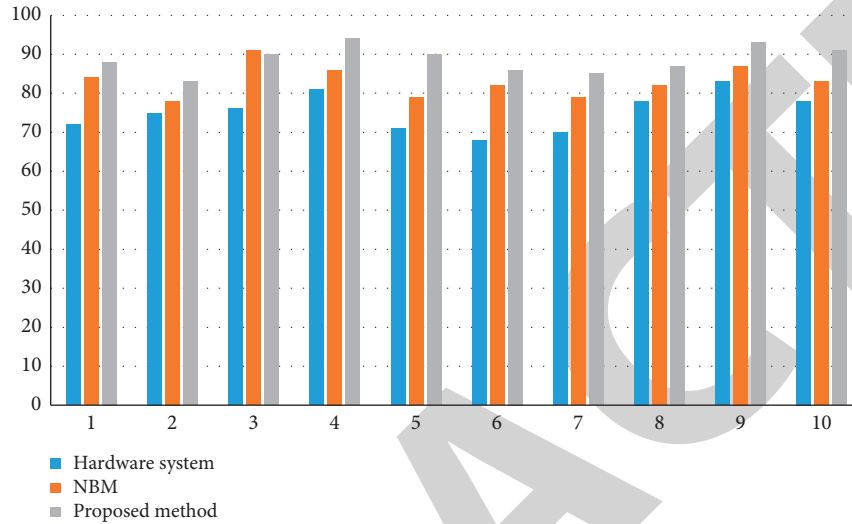| Serial number	Name of the music	Sample size
6	Numeral	220
7	Article	88
8	Preposition	30
9	Conjunction	266
10	Interjection	310

FIGURE 5: Comparison of accuracy of three music classification models.

calculate the diagnostic accuracy of their 10 categories of vocabulary. The specific results are shown in Figure 5, where the definition of accuracy is shown in the following formula:

$$ACC = \frac{\text{number of samples correctly diagnosed}}{\text{total number of samples}}. \quad (5)$$

As can be seen from Figure 5, the classification accuracy of the method based on hardware system for 10 kinds of music is relatively low, no more than 85%, which is far from the actual demand. The lowest accuracy of hardware system is only about 68%, and the highest accuracy is only about 82%. In this way, hardware system can no longer be applied to actual writing in recognition accuracy. This is mainly because the implementation of hardware system method is linear in nature, while English data types often have nonlinear characteristics. In this way, the optimal vocabulary diagnosis model cannot be accurately established, resulting in a low diagnosis rate and a high error rate of vocabulary diagnosis. However, the classification accuracy of NBM method for 10 kinds of vocabulary is more than 85%, which can meet the requirements of practical application of English vocabulary diagnosis. The lowest accuracy rate of NBM method is only about 78%, the highest accuracy rate is only about 90%, the average accuracy rate is about 80%, and the overall accuracy rate is relatively low. This is because NBM belongs to a nonlinear music classification model. It can describe the linear and nonlinear relationships between English words and establishes a good vocabulary diagnosis model. The highest accuracy rate of the proposed method is about 95%, the lowest accuracy rate is about 85%, and the average accuracy rate is 90%, which is more obvious than the other two algorithms. However, due to the feature of using a single classifier, the results of vocabulary diagnosis should be enhanced. The classification performance of the model in this paper is more than 90% for 10 kinds of music, which is much higher than the accuracy of the comparison algorithm. This is mainly because the model proposed can not only effectively describe the nonlinear characteristics among words and overcome the limitations of linear classifier but also can effectively extract the features of lexical data, improve the optimal input for the model, and effectively improve the accuracy of the model and reduce the speed of the model.

4. Conclusions

English vocabulary diagnosis is an important technology to improve the level of language retrieval. To solve the problems of low accuracy and slow speed, this paper proposes an intelligent diagnosis system for English writing based on data feature extraction and fusion, and compared with other English vocabulary diagnosis models, the results show that the proposed model in classification is realized by using multiple feature fusion machine learning in the process of fully considering the time domain and frequency domain characteristics of the characteristics of 10 kinds of English vocabulary and showed better overall diagnosis performance. The proposed methods also have a very broad application prospect in big data feature extraction.

Data Availability

Conflicts of Interest

The author declares no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] Z. Wang, X. Zhao, J. Cheng et al., "Design and implementation of mobile intelligent terminal controlled LAPS detection system," *Acta Scientiarum Naturalium Universiatis Nankaiensis*, vol. 50, no. 6, pp. 11–15, 2017.

[2] X. Xu, Z. Tao, W. Ming, Q. An, and M. Chen, "Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion," *Measurement*, vol. 165, pp. 108086–108097, 2020.

[3] J. Lu, H. Wang, J. Zhou et al., "Low-rank adaptive graph embedding for unsupervised feature extraction," *Pattern Recognition*, vol. 113, Article ID 107758, 2021.

[4] Y. Chen, X. Wang, and X. Du, "Diagnostic evaluation model of English learning based on machine learning," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2169–2179, 2021.

[5] A. A. S. Mohamedahmed, "Pronunciations of the past tense inflectional morpheme-ed sounds among north border university students: diagnosis and remedy," *European Journal of English Language and Literature Studies*, vol. 9, no. 2, pp. 23–32, 2021.

[6] K. Choo, A. Mezzacapo, and G. Carleo, "Fermionic neural-network states for ab-initio electronic structure," *Nature Communications*, vol. 11, no. 1, pp. 2368–2377, 2020.

[7] X. Xu, D. Cao, Y. Zhou et al., "Application of neural network algorithm in fault diagnosis of mechanical intelligence," *Mechanical Systems and Signal Processing*, vol. 141, Article ID 106625, 2020.

[8] A. Kumar, G. Vashishtha, C. P. Gandhi, Y. Zhou, A. Glowacz, and J. Xiang, "Novel convolutional neural network (NCNN) for the diagnosis of bearing defects in rotary machinery," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[9] Z. Wang, S. L. Lukowski, S. A. Hart et al., "Is math anxiety always bad for math learning? The role of math motivation," *Psychological Science*, vol. 26, no. 12, pp. 1863–1876, 2015.

[10] C. P. Chandrika and J. S. Kallimani, "Polarity identification for handwritten text in multilingual documents using open source optical character recognition tools," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 9-10, pp. 4045–4049, 2020.

[11] D. Gupta, J. J. P. C. Rodrigues, S. Sundaram et al., "Usability feature extraction using modified crow search algorithm: a novel approach," *Neural Computing & Applications*, vol. 32, pp. 1–11, 2018.

[12] H. Guo and W. Wang, "An active learning-based SVM multi-class classification model," *Pattern Recognition*, vol. 48, no. 5, pp. 1577–1597, 2015.

[13] J. Liang, Y. Zhang, J.-H. Zhong, and H. Yang, "A novel multi-segment feature fusion based fault classification approach for rotating machinery," *Mechanical Systems and Signal Processing*, vol. 122, no. 5, pp. 19–41, 2019.

[14] X. Li, Z. Tang, W. Chen, and L. Wang, "Multimodal and multi-model deep fusion for fine classification of regional complex landscape areas using Zi Yuan-3 imagery," *Remote Sensing*, vol. 11, no. 22, pp. 2716–2723, 2019.

[15] T. Wang, J. Dong, T. Xie, D. Diallo, and M. Benbouzid, "A self-learning fault diagnosis strategy based on multi-model fusion," *Information*, vol. 10, no. 3, pp. 116–136, 2019.

[16] Y. Zhou, M. Xiao-Dong, and Z. Feng-An, "Scene classification of remote sensing image based on deep network and multi-scale features fusion," *Optik*, vol. 171, pp. 287–293, 2018.

[17] X. Zhao and Z. Feng, "Research on repetition rate of English vocabulary based on large-scale corpus," *Foreign languages and their teaching*, vol. 69, no. 4, pp. 87–95, 2016.

[18] P. Jinyu, "Fuzzy semantic selection technology in English semantic machine translation," *Modern electronics technique*, vol. 40, no. 22, pp. 42–44, 2017.

[19] X. Wang, "Dynamic impact of formative assessment on college students' English writing level," *Journal of PLA University of Foreign Languages*, vol. 39, no. 4, pp. 102–110, 2016.

[20] D. Wu, "A novel gate resource allocation method using improved PSO-based QEA," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–9, 2020.

[21] G. Chen and S. Li, "Research on location fusion of spatial geological disaster based on fuzzy SVM," *Computer Communications*, vol. 153, no. March, pp. 538–544, 2020.

[22] G. Chen, J. Qi, C. Tang, Y. Wang, Y. Wu, and X. Shi, "Analysis and research of key genes in gene expression network based on complex network," *Complexity*, vol. 2020, Article ID 8816633, 12 pages, 2020.

[23] S. C. Freeman, C. R. Kerby, A. Patel, N. J Cooper, T. Quinn, and A. J Sutton, "Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: MetaDTA," *BMC Medical Research Methodology*, vol. 19, no. 1, pp. 81–11, 2019.

[24] L. Yingfeng, "Bibliography," *Pirates, Traitors, and Apostates*, vol. 41, no. 22, pp. 159–174, 2018.

[25] A. K. Tiwari, A. Dutta, A. Tewari, and R. Mohan, "Development and Simulation of a Novel Approach for Dynamic Workload Allocation Between Fog and Cloud Servers," Springer, Singapore, pp. 595–602, 2021.