Language Related Research
E-ISSN: 2383-0816
https://lrr.modares.ac.ir
https://doi.org/10.29252/LRR.14.5.11
http://dorl.net/dor/20.1001.1.23223081.1401.0.0.322.4

Tarbiat Modares University

Vol. 14, No. 5
pp. 285-307
November & December 2023

Received: 27 February 2022
Received in revised form: 23 April 2022
Accepted: 9 May 2022

Optimization of College English Dynamic Multimodal Model Teaching Based on Deep Learning

Yanli Zhao[1&2]* *& Nur Ainil Binti Sulaiman[1]*

Abstract

Since 2010, deep learning has been further developed, and the concept of multi-modality has penetrated into all walks of life. However, it has not been fully researched and applied in college English teaching, so this study modeled and practiced the multimodal teaching method of college English under the deep learning mode and its application. The definitions of modality and medium are first introduced, and then the definition of multimodality in this study is clarified. Then the classification of multimodal transport is expounded. The random forest algorithm is chosen as the main algorithm of this research, and a dynamic multimodal model is established. After that, there was a collaboration with a university and sophomore students were selected for practice. After processing and analyzing the collected data, it was found that in the data sample of 268 students, the number of students who did not study independently accounted for 24%, which indicates that most college students lack interest in learning English. Preliminary tests were also conducted on students' English proficiency throughout the year, and the results showed that the students' English proficiency was at a pass level and the overall English proficiency was weak. Reassessment of students' English proficiency showed that the actual teaching effect of each English proficiency was greater than 85%, and the effectiveness of English teaching in the selected universities was significantly improved. The average score improved by 8 points, indicating that multimodal teaching is scientifically effective

After a semester of multimodal teaching, the English teaching effectiveness of the university selected in this article has significantly improved. The research results indicate that the development of deep computer learning has introduced multimodal concepts into the teaching field, which is very suitable for assisting language learning based on its own advantages.

Keywords: deep learning, multimodal theory, random forest algorithm, dynamic multimodal model

[1] Faculty of Education, Universiti Kebangsaan Malaysia, 43600 UKM Bangi Selangor, Malaysia;
[2] Foreign Languages Department, Zhengzhou Business University, Zhengzhou Gongyi, 451200, China;
* Corresponding Author: *e-mail: fao@zbu.edu.cn;* ORCID: https://orcid.org/0000-0002-7611-2207

1. Introduction

After 2010, computer deep learning neural networks have led to their rapid development in various artificial intelligence fields such as vision, hearing, language processing, and recommender systems. With the joint efforts of relevant practitioners and researchers, many results have been achieved and huge economic benefits have been generated. Machine learning enables computers to solve problems by learning from data. Mater and Coote (2019) aim to explain the concept of deep learning to chemists of any background and provide an overview of various applications presented in the literature (Gao et al., 2022; Mater & Coote, 2019; Wang, 2023; Wang et al., 2023a). Janiesch et al. (2021) summarized the fundamental principles of machine learning and deep learning to gain a broader understanding of the system foundation of current intelligent systems. (Janiesch et al., 2021). Bartlett et al. (2021) paid special attention to the linear state of neural networks. The network can be approximated through a linear model. In this case, it proves the success of gradient flow, considers the benign overfitting of the two-layer network, and gives an accurate asymptotic analysis (Bartlett et al., 2021).

Diao and Hu (2021) analyzed the complexity of deep learning and the application of multimodal target recognition in English education system. In the process of abundant teaching resources, multi-mode teaching has gradually become of practical significance (Diao & Hu, 2021). Delecraz et al (2021) proposed a method to parse the syntactic appendage of the parse tree intermediary according to the visual and lexical features. Visual features are derived from the nature and location of the objects detected in the image, which are aligned with the text phrase in the caption. The reorder uses this information to reorder the shift to reduce the syntax tree generated by the parser (Delecraz et al., 2021). Deep learning methods have made significant breakthroughs and exhibit considerable performance in various applications with useful security tools. Dargan et al. (2020) introduced the main differences and future challenges between deep learning, classical machine learning, and traditional learning methods (Dargan et al., 2020). Deep learning is a subcategory of machine learning that follows human learning instincts and produces accurate results through examples.

Bashar (2019) introduced a survey of deep learning neural network architectures used in various applications to achieve accurate classification through automatic feature extraction (Bashar, 2019). College English teaching requires certain flexibility and activity in the classroom, so Marufuzzaman and EkşioğLu (2017)

[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]

proposed a multimodal classroom teaching model to improve the overall teaching effect. It is expected to build and improve the curriculum model and improve the English level of college students through the improvement of teaching intuition (Marufuzzaman & Ekşioğlu, 2017). Through the analysis of college English teaching practice, Ramtohul et al (2021) found that multimodal classroom teaching mode can effectively alleviate the disadvantages of traditional teaching (Ramtohul et al., 2021).

In college English teaching, it is not the application of multimedia and multiple symbolic modalities that can achieve effective multimodal teaching modes (Wang & Hemchua, 2022). At present, some teachers believe that as long as multimedia is used in English teaching, a multimodal teaching mode can be achieved. In this regard, teaching practice has had a high demand on EFL teachers' literacy in this multicultural education context (Fu & Wang, 2022). In classroom teaching, PPT courseware is relatively relied on, and the teaching activities carried out are also based on the teaching PPT. Its teaching methods are relatively single, lacking sufficient communication and interaction with students. Students in a relatively monotonous and tedious English learning environment are prone to develop fatigue and boredom, leading to unsatisfactory teaching outcomes. To leverage the role and advantages of multimodal college English teaching models, teachers need to continuously accumulate teaching experience in their teaching. Reasonably design college English teaching, select appropriate modes, and achieve its teaching effectiveness. This study aims to investigate these issues using all available resources and has the potential to achieve some results. How to mobilize greater motivation and desire for self-directed learning through multiple sensory stimuli, how to immerse students in learning even outside of class, and how to better utilize deep learning technologies all make this topic of great research value.

2. Review of the Literature

Li et al (2022) proposed a multimodal learning emotion analysis method based on DNN. Combining video and voice to detect students' learning emotions in real-time (Li et al., 2022). Neuroengineering is an emerging interdisciplinary discipline aimed at utilizing engineering techniques to study the function and manipulation of the nervous system. Wang et al (2021) generated a map of existing research results and their relationships, and provided meaningful suggestions and assistance for future researchers (Wang et al., 2021). Dong (2020) carries out research in English-based

[DOR: 20.1001.1.23223081.1401.0.0.322.4]
[DOI: 10.29252/LRR.14.5.11]

language, and builds an intelligent English recognition and prediction system in combination with support vector machine to obtain EEG signals related to Chinese speech. In addition, this paper uses wavelet packet decomposition and common space mode to extract the features of the collected EEG signals, and uses support vector machine model to classify and compare the signal features. The research results show that the recognition rate of the traditional algorithm is significantly lower than that of the algorithm model, and the algorithm in this study basically meets the requirements of the actual use of the system (Dong, 2020). And when it is applied to English teaching the definition will be clearer and more specific, referring to a new way of teaching in which multimedia teaching is used in the process of English teaching to actively engage all the organs that can receive signals, such as students' vision, hearing, taste and touch, to deepen language memory.

Pan and Zhang (2020) aim to apply multimodal teaching methods to high school English reading teaching and attempt to understand whether multimodal teaching can stimulate students' interest in English reading and improve their reading proficiency (Pan & Zhang, 2020). Wang et al. (2022) proposed an educational text classification method based on learning depth. It improves the efficiency of traditional educational text classification methods and the recall rate of research objectives. The deep learning algorithm is introduced to perform preprocessing and feature learning, and the learning features are input into the deep learning Softmax classifier based on the learning results (Wang, 2022). The analysis of multimodal arguments in advertising is a key and problematic research field. Macagno and Pinto (2020) aims to combine tools developed in social semiotics, pragmatics and argumentation theory. It proposes and explains a method for reconstructing and analyzing the "Dual mode" argument in advertising (Macagno et al., 2021). Chen and Jiang (2018) discussed the cognitive mechanisms behind this multimodal humor and their understanding (Chen & Jiang, 2018). Nemati et al. (2019) proposed a hybrid multimodal data fusion method that utilizes latent spatial linear mapping to fuse audio and visual modalities. Then, the evidence fusion method based on Dempster Shafer (DS) theory is used to fuse the features projected into the cross-modal space with the text modality (Nemati et al., 2019). Oghyanous et al. (2022) examined the content, structural validity, and internal consistency of the Academic Emotional Regulation Questionnaire (AERQ) in the context of Iranian English as a Foreign Language (EFL) (Oghyanous et al. 2022).

The above research is still only a small representative part of the multimodal theory based on deep learning models that has flourished in ELT in the past decade or so. As theoretical research has progressed, practical proof has followed, and more and more examples have shown its scientific validity and feasibility in English teaching. Abdel (2021) investigated a common phenomenon known as multimodal recycling. Because cartoonists often recycle their early development of painting and composition ideas. Assuming two basic types of recycling: the same narrative and narrative transformation. This type of research will contribute to the academic study of political comics and cognitive research based on images and multimodal stimuli (Abdel, 2021).

3. Types of Multimodality

The Dodo modality can be easily classified according to its definition in the introduction. Because it is a response of the human senses, the academically accepted classifications are visual modality, auditory modality, olfactory modality, gustatory modality, and tactile modality. Stöckl (2019) investigated a common phenomenon known as multimodal recycling, as cartoonists often recycle their early-developed painting and composition ideas (Stöckl, 2019). Studies have shown that adults obtain approximately 80% to 83% of their information visually. Therefore, text, pictures, and layout in English language teaching together form the most dominant information transfer modality that conveys specific meanings to students and further their understanding. While this was the most important way of teaching in the previous traditional way, more advanced multimodal concepts would enrich the visual modality by adding underlining or highlighting nouns and verbs that have actual meaning, or by boldly summarizing sentences in a text. Like visual modality, auditory modality is also a very important input method. Studies have shown that when blind people lose their vision, the primary way of receiving information shifts to the auditory system. In English language teaching, listening has always been in an awkward position because it accounts for a low percentage of the test paper, so it has not received much attention from teachers and students. In the multimodal concept, auditory is one of the most important aspects to be developed, because the visual modality has matured, and auditory information reception is second only to visual. Moreover, auditory stimulation can be applied to more after-school scenarios and has a broad research prospect. The olfactory, gustatory and tactile modalities have a narrower but deeper scope than the first two

[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]

in language teaching. Adding sensory stimuli to certain words and sentences related to smell, taste, and touch can greatly improve students' comprehension and retention of them.

As the application and research on multimodality in language teaching has intensified, some experts have classified it according to the different scenarios. Lestari (2022) used a multimodal approach to study digital based learning management to transform textbooks (Lestari, 2022). Cultural scenarios give students exposure to different ways of thinking and values in different languages. Life situations are the largest of these categories, and everyday conversations can be effective in improving students' practical application of the language, which will be used frequently in the workplace after graduation. Entertainment scenarios are relatively relaxed and lively, creating a relaxed and enjoyable classroom atmosphere that is fun and educational. Evaluation scenarios and problem scenarios can promote interaction between students and teachers, identify problems and provide feedback at any time, which is more conducive to improving English performance in various aspects, like speaking and writing (Wang et al., 2023b).

4. Random Forest Algorithm to Build Multimodal Models

4.1 Selection of Algorithm

To address the above concepts and problems, this paper establishes a multimodal dynamic model based on a widely used deep learning algorithm, the random forest algorithm. By integrating the model into the teaching information management system and collecting information such as basic information of students and English teaching related information, the English teaching data of teachers and students are analyzed (Wang, 2017). The relationship rules between different items of English learning were obtained using the random forest algorithm, and the students' weaknesses were quickly and accurately calculated so that teachers could target their teaching and after-school homework tasks. The final results obtained, over a period of time, showed the superior characteristics of the multimodal model, enabling students to learn English to reach the strategic goal of sustainability.

In order to adapt to the new requirements in the multimodal concept, the algorithm is constantly evolving and updating. The random forest algorithm has been widely used in the first line of data processing since its invention, and it is chosen for this study for the following three reasons: firstly, school English

[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]

teaching data is a sparse data set, which fits well with the random forest algorithm; secondly, the principle of the random forest algorithm is simple and easy to implement the research requirements; finally, the student English reading data collected in English classrooms, English listening data, English reading data and Finally, the English reading data, English listening data, English reading data, and test score data collected in the English classroom have strong correlations and complex intrinsic relationships, which are suitable for using the random forest algorithm to establish dynamic multimodal modeling to conduct the study.

In data mining, random forest algorithm is a method to find association rules between data samples, which can find the hidden frequent data sets in big data. The classic example of Random Forest algorithm in research is finding the intrinsic association between diapers and beer, and placing them together on the supermarket shelf to increase the sales of both products, which is a demonstration of how Random Forest algorithm is beneficial to decision making. The random forest algorithm iterates layer by layer, first finding frequent data set 1, then frequent data set 2 through data set 1, and so on in successive iterations. The random forest algorithm uses the implication expression from Y to Z to reflect the association rules of two disjoint item sets Y and Z. It uses support and confidence rates to measure the strength of the association between Y and Z.

$$S(Y,Z) = P(Y,Z) = \frac{num(yz)}{num(allsamples)} \tag{1}$$

$$C(YZ) = P(y|Z) = \frac{P(yz)}{P(z)} \tag{2}$$

S Theand C in equations (1) and (2) represent the Support rating and Confidence rating, respectively, and num represents the total set.

4.2 Student ELT Data Collection and Pre-Processing

In the traditional teaching data collection in schools, the main component is the students' achievement statistics. This traditional collection only requires teachers to enter data into Office software tables, and then add, subtract and rank them through simple excel self-contained formulas and other operations. The multimodal need for student English learning data is more three-dimensional and multi-faceted, with significant informational characteristics. The new era of data collection with the

[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]

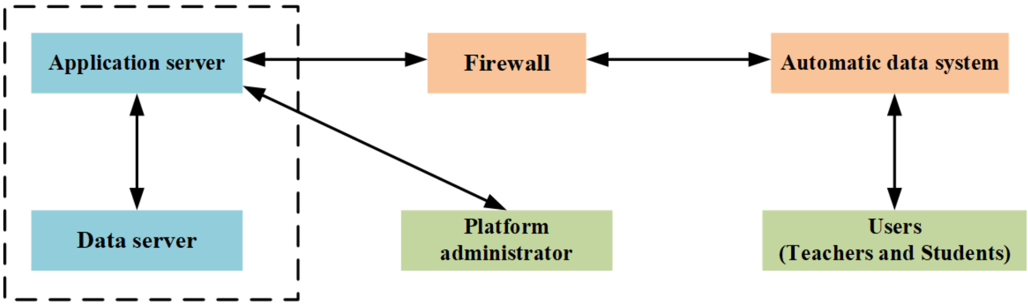
[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]

support of big data is more focused on the use of multiple sensors and teaching management platforms, and the use of existing resources to achieve the goal of efficient learning with multi-sensory stimulation. First of all, the experimental data to be collected in this study include students' basic information, English classroom situation, after-school homework situation, midterm and final English exam results, and English competition results. Among them, students' basic information such as student name, student number, gender and required elective courses will be entered or assigned at the time of enrollment and can be obtained through the school's teaching management platform. Students' English reading preferences, hobbies and personalities can be obtained through questionnaires. What happens in English classes is recorded by English teachers and learning class representatives in class and recorded in the Teaching Management Platform database after class. After-class English exercises and assignments are collected by the students themselves through the "Independent Multimodal Learning English Data System". English competition results are entered into the TLM database by the lead teacher, or reported to the teacher if the students are competing individually.

It can be seen that the original data collection of students' English school situation mainly relies on the teaching information management platform and the autonomous data system. The teaching information management system is more intelligent than the traditional excel sheet statistics. It uses a CS hybrid architecture with modules for data storage, analysis and output integrated within the system, capable of handling relatively complex and large amounts of student data. Similarly, the architecture of the autonomous data system is also completed by the CS hybrid structure, so the two can be organically combined together to perform the functions of data collection and daily use by teachers and students. The common advantage of the two systems is that the system administrator can directly operate the server without relaying, giving instructions and downloading the required information. The structure layout of teaching information management platform and autonomous data system is shown in *Figure 1*.

Figure 1
Structure Diagram of Teaching Information Management Platform and Autonomous Data System



As shown in the structure diagram above for the teaching information management platform and the autonomous data system, it can be seen that the application server and the data server can interact with each other on their own. The application data server can be used directly by the platform administrator or filtered by the firewall and used by teachers and students through the autonomous data system.

The Random Forest algorithm differs from older data processing approaches in that it cannot perform a brute force search, and therefore requires a pruning operation on the collected data on students' English learning. The operation is based on the two main theorems of the algorithm. The first theorem of the random forest algorithm: If a set is a frequent itemset, then all its subsets are frequent item sets. The second theorem of this algorithm: If a set is not a frequent itemset, then all its supersets are not frequent item sets. As shown in Figure 2 below, the original data is first subdivided into several subsets, and then the subsets are merged into a valid factor ensemble. Then we find out the subset AB that is not a frequent itemset, then according to the first and second theorem of the random forest algorithm, we can know that the set containing the subset AB is not a frequent itemset, i.e., the subset within the dashed line in the figure. Finally, the pruning operation is completed by eliminating this part. Considering that the scope of this study is students' English learning situation data, the pruning is usually the information of transferring and withdrawing students, invalid classroom after-school practice data, invalid midterm and final exam results and invalid English competition results, etc.

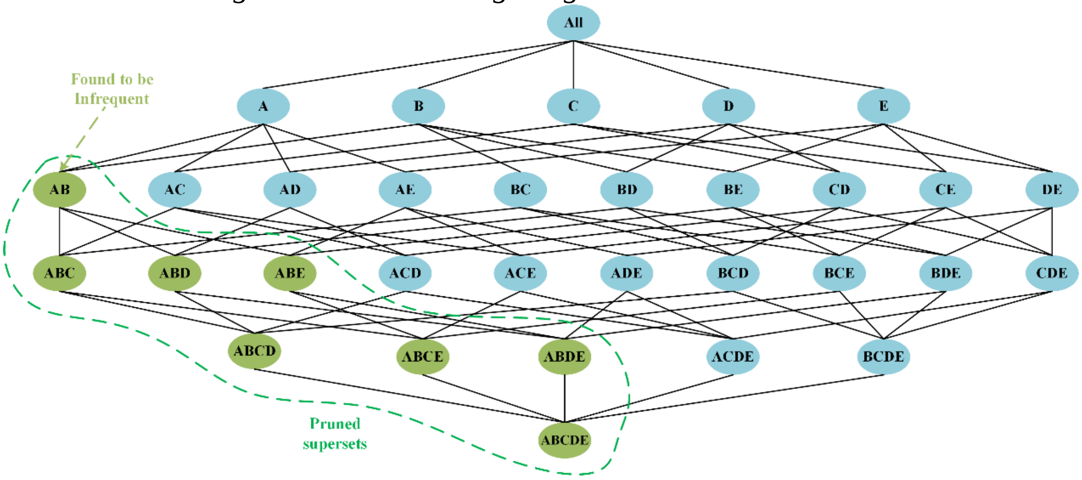
[DOR: 20.1001.1.23223081.1401.0.0.322.4]
[DOI: 10.29252/LRR.14.5.11]

[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]

Figure 2

Random Forest Algorithm Tree Pruning Diagram



After pruning the raw data, it is also necessary to normalize the data obtained from the Teaching Information Management Application Server because it is heterogeneous. First, the data are initially categorized and arranged using the matrix formula (3).

$$D_T = \begin{bmatrix} d(y_1, y_1) & \cdots & d(y_1, y_T) \\ \cdots & \cdots & \cdots \\ d(y_T, y_1) & \cdots & d(y_T, y_T) \end{bmatrix} \tag{3}$$

Where T denotes the total number of all data, and then the data are averaged using Equation (4).

$$\overline{R} = \frac{\sum_{i=1}^{T} R_i}{T} (i = 1, 2, \cdots, T) \tag{4}$$

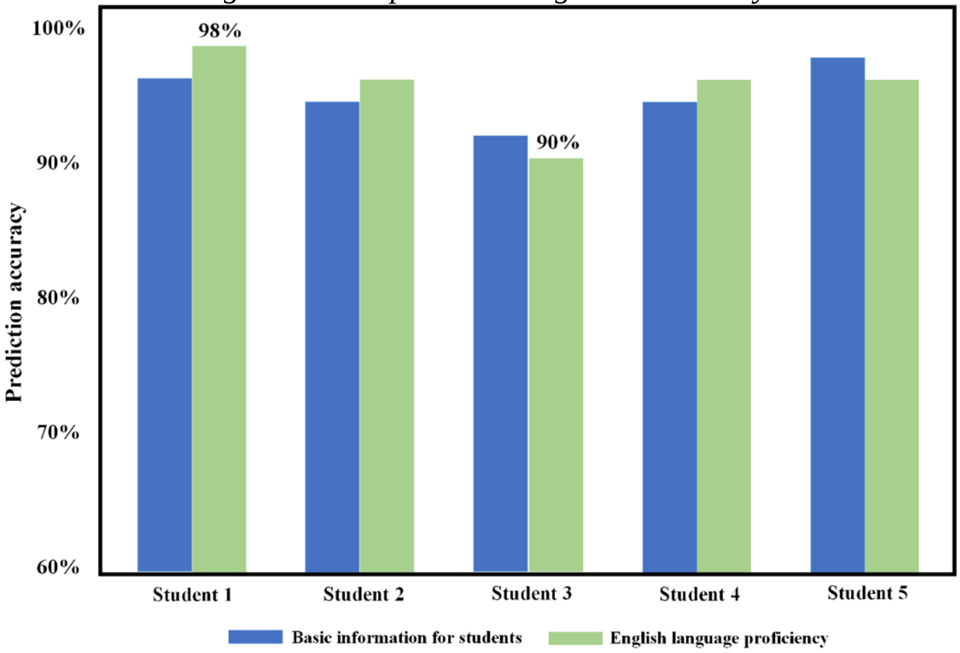
The above equation $\overline{R}$ means the average of R_i for the total number of T. Finally, the data is normalized by weighting the error on the classification, which requires the use of the following formula.

$$R = \overline{R} + d_{\min}(y_i, y_j) \tag{5}$$

After pruning operation and normalization, the data are unified and the structure is easier to perform multimodal mining and modeling afterwards. The random forest algorithm can also compensate for the missing necessary data using machine learning methods, and the method of supplementing data is divided into two main steps. The first step is to compare the missing students' data with all other students and select a number of key factors, such as students' reading comprehension scores and spoken English expressions. No less than 50 samples of similar data were found. The second step is to make predictions on the resulting 50 samples to scientifically impute the missing data. This study used data from all sophomore students at a university for hypothetical missing complements. The key information of five students was first hidden, and the data was compared with the real data using the above method of imputation, and the results are shown in the following figure.

Figure 3

Random Forest Algorithm Completes Missing Data Accuracy Test Result Graph



The accuracy obtained by dividing the inferred data by the real data using the above method is shown in *Figure 3* above, which shows that the inferred accuracy of the basic personal information and English proficiency data of the five selected sophomore students are in the lowest 90% and the highest 98%. This result is in line with the expectation and indicates that the random forest algorithm is scientific and feasible to compensate for the missing data.

[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]

5. Random Forest Algorithm for Dynamic Multimodal Modeling and Example Analysis

5.1 Establishment of Dynamic Multimodal Model

In order to capture the extent of multimodal proficiency improvement of sophomore students with different levels of English mastery, data mining was performed using the random forest algorithm. In the construction of the model, considering different projects in English classes and different competitions in English competitions, the data locus state prediction vector is.

$$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n) \neq 0 \tag{6}$$

The data related to the English proficiency of the diversified sophomore students are transferred to the application server, and when scheduling the data, the characteristic vector is used to refer to.

$$y^{(k)} = \left[y_1^{(k)}, y_2^{(k)}, \cdots, y_{N_{k-1}}^{(k)} \right]^T \tag{7}$$

$$s^{(k)} = \left[s_1^{(k)}, s_2^{(k)}, \cdots, s_{N_k}^{(k)} \right]^T \tag{8}$$

$$z^{(k)} = \left[z_1^{(k)}, z_2^{(k)}, \cdots, z_{N_k}^{(k)} \right]^T \tag{9}$$

Where $y^{(k)}$ and $z^{(k)}$ denote linear horizontal and vertical inputs to the system and $s^{(k)}$ denotes reversible invariant outputs.

The English test scores under the guidance of the multimodal concept are used as the time-frequency characteristics, and thus this is used as a medium to reflect the essential characteristics of this performance data. The ensemble of all data is N discrete dynamically distributed points $P = \{p_1, p_2, \cdots, p_N\}$, and the gain index time mean and frequency mean of students' physical training in this experiment can be calculated as the following equations (10) and (11), respectively

$$t_m = \frac{1}{E} \int_{-\infty}^{+\infty} t \left| y(t) \right|^2 dt \tag{10}$$

$$v_m = \frac{1}{E} \int_{-\infty}^{+\infty} v \left| Y(v) \right|^2 dv \tag{11}$$

After the above series of processing, the system of students' English proficiency

[DOR: 20.1001.1.23223081.1401.0.0.322.4]
[DOI: 10.29252/LRR.14.5.11]

index using a linear or nearly linear dynamic multimodal model was obtained. Information storage tree structure fractal In order to analyze the dynamic model of the promotion relationship of the college English program on the English proficiency gain index of the sophomore group of students, the degree of proficiency gain of the English improvement program for different students was grasped and adjusted regularly based on the examination and data collection and analysis uploaded by teachers and students together for a whole academic year. Based on this model, a dynamic fractal of the index information storage tree structure was used to calculate the variance matrix of this storage structure using principal component analysis C:

$$C = \frac{z+1}{N}\left[Y-\overline{Y_i}\right]\left[Y-\overline{Y_i}\right]^T \tag{12}$$

A linear dynamic system that fits multiple factors influencing college English proficiency is developed to achieve fitting of diverse data in a dynamic model, which can be expressed by the following equation.

$$R_\beta Y = U\left\{E \in U/R \middle| c(E,Y) \le \beta\right\} \tag{13}$$

$$R_\beta Y = U\left\{E \in U/R \middle| c(E,Y) \le 1-\beta\right\} \tag{14}$$

$$bnr_\beta(Z) = R_\beta Z - R_\beta Z_1 \tag{15}$$

It is necessary to merge the frequent subsets into the parent set with universal influence, and the number of master sets is selected according to the degree of influence accumulated by the variance, and only when the variance accumulates to a certain contribution, the subsets corresponding to it can be merged into the parent set with universal influence. Through the above dynamic fractal design, this experiment realizes the multimodal modeling of students' English proficiency data based on deep learning random forest algorithm.

5.2 Example Analysis of Students' English Proficiency Data

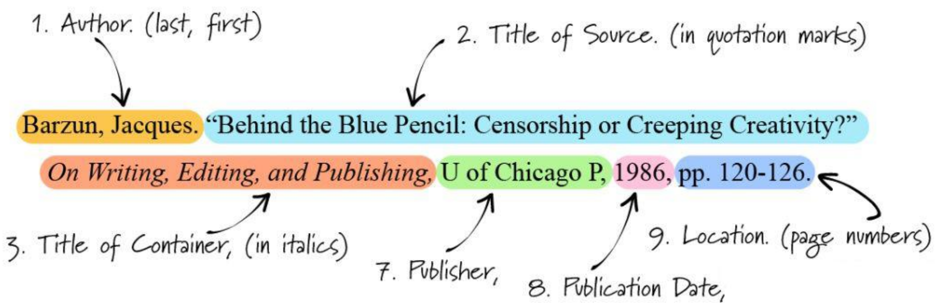
Attention should also be paid to the use of non-verbal symbols other than words in the construction of multimodal concepts and in actual communication. In English reading, non-native learners are expected to master grammar, which is a certain pattern of linguistic expression. To fully engage visual modality, one can mark the components of a long and difficult sentence with different colors to help sophomores understand better while deepening their memory. Suitable annotations

[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]

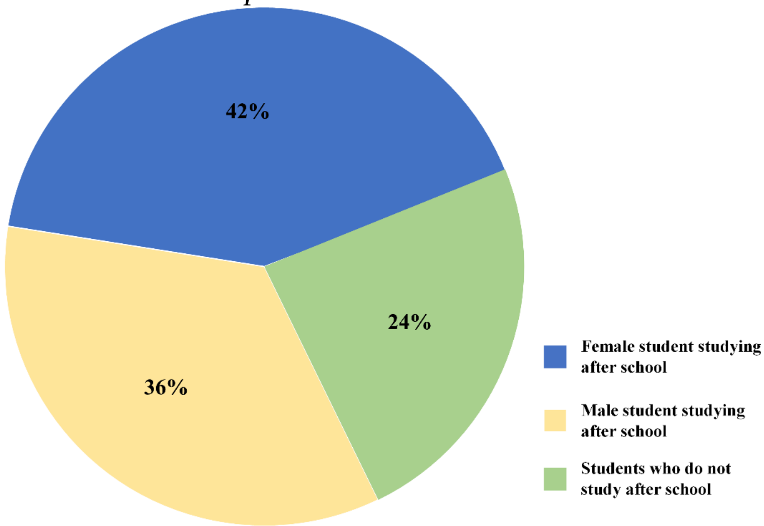
can also be added to complement the audio recording of the textbook and the teacher's colorful explanations to achieve a multifaceted motivation. The sentence markers and annotations can follow the pattern shown below.

Figure 4
Example Sentence Markup and Annotation Display Diagram



The English student population is a complex system, and this study conducted data collection at a university. The data was also pre-processed to remove errors and missing data and normalized. The data were made available to be applied to a dynamic model built by the random forest algorithm of deep learning, and after processing and analysis, several results were obtained in this paper.

Figure 5
The Number of Students Who Actively Study English After School and Their Ratio of Male to Female Population

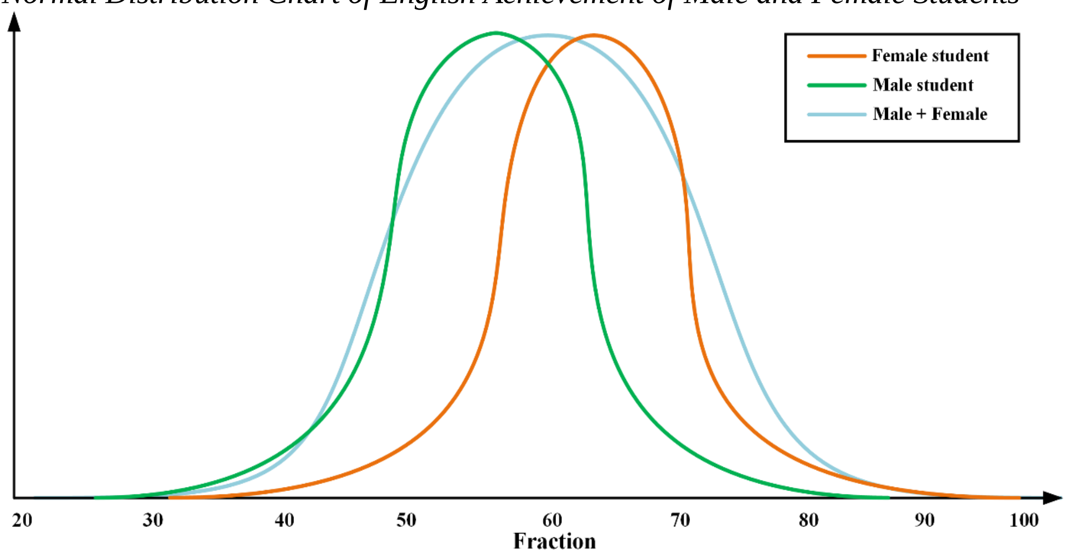


The analysis showed that 76% of the 268 sophomore students in the university sample selected for this study studied English independently after school, while 24% did not. The number of female students who studied independently was 6 percentage points higher than the number of male students. From this result, it can be seen that a larger percentage of the selected targets lack interest in learning English and need a change in teaching mode to improve it.

The 268 students were also tested on all aspects of their English proficiency and the results are shown below.

Figure 6

Normal Distribution Chart of English Achievement of Male and Female Students

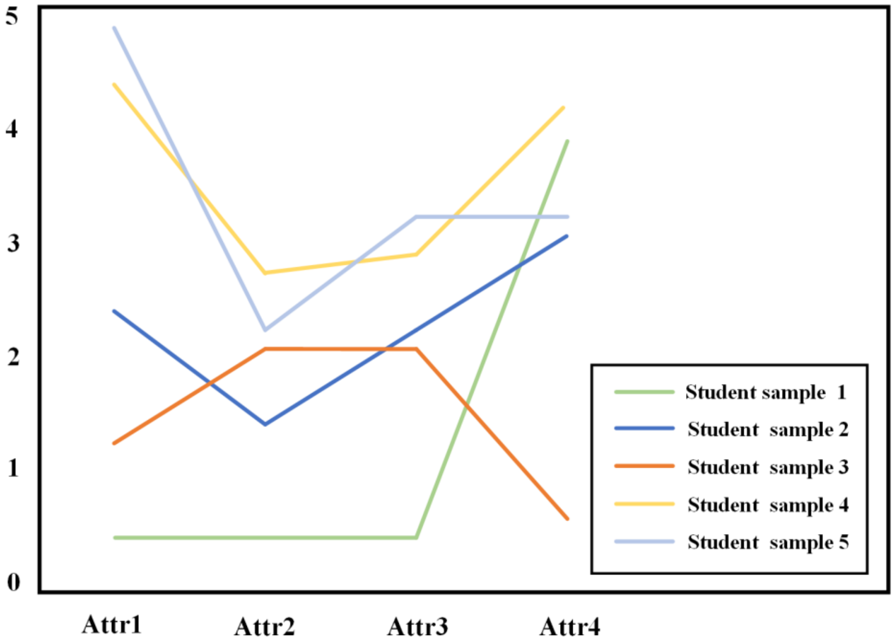


The results of the data analysis showed that the students of the whole year were in a passing condition in English, and the performance of female students was overall stronger than that of male students. The reason for this analysis may be that girls in this age group are more likely to concentrate. The overall English general level is weaker and there is a lot of room for improvement.

Since the random forest algorithm also requires the analysis of the attribute features of the selected sophomore students, the many attribute features are automatically filtered for invalid features. So the analysis was performed using the predictive model, and since quite a few of the relevant data attributes were consistent, the results obtained after filtering and analyzing the database for the established teaching effectiveness model are shown in the figure below.

Figure 7

Distribution Map of Various Characteristics of Students In English Courses



From Figure 7 above, it can be seen that the samples of students' common feature attributes generated after the random forest algorithm calculation are five cases. This indicates that the five cases of instructional data all share the four common dynamic attributes of Attr1, Attr2, Attr3, and Attr4. This indicates that although there may be many attributes in the data collected in the study, not all of them are of real concern to educators or teachers, and only these four types of features need to be focused on to improve students' performance in a targeted manner.

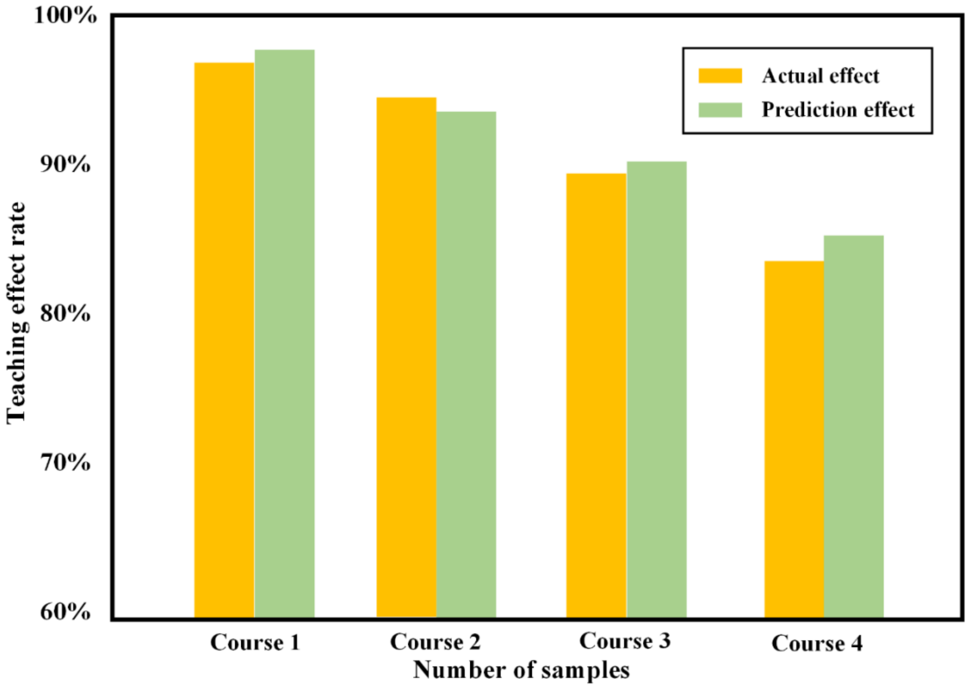
The model constructed by the random forest algorithm can also predict the performance of students who have been studying for a period of time to improve. A comparison of the predicted results with the actual teaching results gives the following results.

[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]

Figure 8

Comparison Between the Predicted Results of the English Teaching Effect Prediction Model and the Actual Results



It is obvious from the figure that the actual teaching effect of English reading, writing, speaking and listening courses is greater than 85%, which indicates that after one semester of multimodal teaching, the selected college English teaching effect has improved considerably. It can also be seen that the maximum difference between the actual teaching effect and the predicted effect is 4%, and the prediction difference in the prediction model of the machine deep learning algorithm is allowed to range from 0 to 6%, and the 4% result is in the region of accurate prediction in the range.

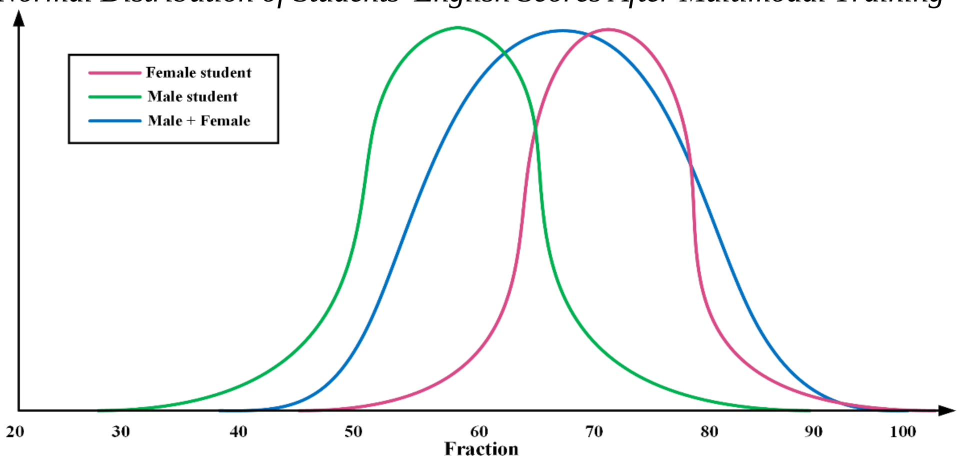
After a semester of teaching guided by multimodal theory, the English proficiency of these 268 students was again tested in a comprehensive manner, and the results obtained are shown below.

[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]

Figure 9

Normal Distribution of Students' English Scores After Multimodal Training



From the normal distribution, it is obvious that the English proficiency of all 258 sophomore students has improved significantly, and the average score has increased from 61 to 69. The reason for this is that female students have a more solid foundation and it is easier to improve their English skills after using the multimodal approach. Compared with the traditional computer algorithm of data in Euclidean space, the multimodal technology adopted in this paper can process data in non-Euclidean space, which greatly improves its application range.

6. Conclusion

The development of deep computer learning has introduced the concept of multimodality into the field of teaching, which is very suitable for assisting language learning based on its own advantages. This article introduces the definitions of two concepts related to multimodality, modality and medium. The definition of multimodality was clarified from the perspectives of modality and media. Then, the classification of multimodality was elaborated. The random forest algorithm is selected as the main algorithm of this study. After normalizing and diversifying the original data, a dynamic multi-modal model is established. Afterwards, we collaborated with a university and selected all sophomore students for internships. The teaching management platform collaborates with an independent multimodal English learning system to effectively collect raw data. The processing and analysis of collected data shows that out of 268 sophomore students' data samples, 76% of students independently learn English after class, and

[DOR: 20.1001.1.23223081.1401.0.0.322.4]
[DOI: 10.29252/LRR.14.5.11]

24% of students do not independently learn English. It can be seen that a large portion of the selected targets lack interest in learning English and need to change the teaching mode for improvement. Preliminary tests were also conducted on students' English proficiency throughout the year. The results show that the students' English proficiency is at a passing level, and the overall performance of girls is stronger than that of boys. The overall English proficiency is weak, and there is still a lot of room for improvement. Then, guided by the concept of multimodality, a one-year English reinforcement program was conducted. The reassessment of students' English proficiency shows that the actual teaching effectiveness of English reading, writing, speaking, and listening courses is greater than 85%, indicating that after a semester of multimodal teaching, the English teaching effectiveness of the selected university has significantly improved. In addition, both male and female students have made progress, with female students making more significant progress, and after using multimodal teaching, it is easier to improve all English skills.

However, there are still some shortcomings in the mining efficiency of random forest algorithm. This algorithm may not necessarily achieve good results in classifying small and low dimensional datasets, so in the future multidimensional research methods (Derakhshan et al., 2023). Although the execution speed is relatively fast, it is much slower than a single decision tree. At the same time, some trees with very small differences may appear, drowning out some correct decisions. When there are many decision trees in the random forest, the space and time required for training will be large. Therefore, in the future, with the development of technology, new teaching models will inevitably emerge, and timely application in English teaching is a necessary condition to keep up with the development of the times. For example, intelligent education systems based on artificial intelligence technology have begun to emerge, providing students with personalized multimodal English reading and learning experiences. In addition, virtual reality technology is also expected to provide a richer media experience for multimodal English reading teaching, allowing students to better understand and remember English knowledge in simulated real scenarios. In short, multimodal English reading teaching has great potential and prospects, and will be more widely applied and developed in the future.

[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]

References

Abdel, R. A. (2021). The multimodal recycling machine: Toward a cognitive-pragmatic theory of the text/image production. *Journal of Graphic Novels and Comics*, *12*(3), 207–239. https://doi.org/10.1080/21504857.2019.1651358

Bartlett, P. L., Montanari, A., & Rakhlin, A. (2021). Deep learning: A statistical viewpoint. *Acta numerica*, *30*, 87–201. https://doi.org/10.1017/S0962492921000027

Bashar, A. (2019). Survey on evolving deep learning neural network architectures. *Journal of Artificial Intelligence*, *1*(02), 73–82. https://doi.org/10.36548/jaicn.2019.2.003

Chen, Q., & Jiang, G. (2018). Why are you amused: Unveiling multimodal humor from the prototype theoretical perspective. *The European Journal of Humour Research*, *6*(1), 62–84. https://doi.org/10.7592/EJHR2018.6.1.chen

Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020). A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering*, *27*, 1071–1092. https://doi.org/10.1007/s11831-019-09344-w

Delecraz, S., Becerra-Bonache, L., Favre, B., Nasr, A., & Bechet, F. (2021). Multimodal machine learning for natural language processing: disambiguating prepositional phrase attachments with images. *Neural Processing Letters*, *53*(5), 3095–3121. https://doi.org/10.1007/s11063-020-10314-8

Derakhshan, A., Wang, Y. L., Wang, Y. X., & Ortega-Martín, J. L. (2023). Towards innovative research approaches to investigating the role of emotional variables in promoting language teachers' and learners' mental health. *International Journal of Mental Health Promotion*, 25 (7), 823–832. https://doi.org/10.32604/ijmhp.2023.029877

Diao, L., & Hu, P. (2021). Deep learning and multimodal target recognition of complex and ambiguous words in automated English learning system. *Journal of Intelligent & Fuzzy Systems*, *40*(4), 7147–7158. https://doi.org/10.3233/JIFS-189543

Dong, S. (2020). Intelligent English teaching prediction system based on SVM and heterogeneous multimodal target recognition. *Journal of Intelligent & Fuzzy*

Systems, *38*(6), 7145–7154. https://doi.org/10.3233/JIFS-179792

Fu, J., & Wang, Y. (2022). Inspecting EFL teachers' academic literacy development in multilingual contexts: A global vision. *Heliyon*, e12143(8),1–6. http://doi.org/10.1016/j.heliyon.2022.e12143.

Gao, Y., Zeng, G. Wang, Y., Klan, A. & Wang, X. (2022). Exploring educational planning, teacher beliefs, and teacher practices during the pandemic: A study of science and technology-based universities in China. *Frontiers in Psychology*, 13, 903244. https://doi.org/10.3389/fpsyg.2022.903244.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets, 31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Lestari, A. S. (2022). Digital based learning management with a multimodal approarch. *Al-Tanzim: Jurnal Manajemen Pendidikan Islam*, *6*(2), 465–476. https://doi.org/10.33650/al-tanzim.v6i2.3235

Li, M., Liu, M., Jiang, Z., Zhao, Z., Zhang, J., Ge, M., ... & Wang, Y. (2022). Multimodal emotion recognition and state analysis of classroom video and audio based on deep neural network. *Journal of Interconnection Networks, 22*(Supp04), 2146011. https://doi.org/10.1142/S0219265921460117

Macagno, F., & Pinto, R. B. W. S. (2020). Reconstructing multimodal arguments in advertisements: Combining pragmatics and argumentation theory. In *Argumentation through languages and cultures* (pp. 141–176). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-19321-7_7

Marufuzzaman, M., & Ekşioğlu, S. D. (2017). Designing a reliable and dynamic multimodal transportation network for biofuel supply chains. *Transportation Science, 51*(2), 494–517. https://doi.org/10.1287/trsc.2015.0632

Mater, A. C., & Coote, M. L. (2019). Deep learning in chemistry. *Journal of chemical information and modeling*, *59*(6), 2545–2559. https://doi.org/10.1021/acs.jcim.9b00266

Nemati, S., Rohani, R., Basiri, M. E., Abdar, M., Yen, N. Y., & Makarenkov, V. (2019). A hybrid latent space data fusion method for multimodal emotion recognition. *IEEE Access, 7,* 172948–172964. https://doi.org/10.1109/ACCESS.2019.2955637

Oghyanous, P. A., Karimi, M. N., & Hashemi, M. R. (2022). Validation of the Persian adaptation of academic emotion regulation questionnaire in the EFL context. *Language Related Research*, *13*(5). https://doi.org/10.52547/LRR.13.5.5

Pan, X., & Zhang, Z. (2020). An empirical study of application of multimodal approach to teaching reading in EFL in senior high school. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(2), 98–111. https:/doi.org/10.3991/ijet.v15i02.11267

Ramtohul, P., Engelbert, M., Malclès, A., Gigon, E., Miserocchi, E., Modorati, G., ... & Freund, K. B. (2021). Bacillary layer detachment: Multimodal imaging and histologic evidence of a novel optical coherence tomography terminology: Literature review and proposed theory. *Retina*, *41*(11), 2193-2207. https://doi.org/10.1097/IAE.0000000000003217

Stöckl, H. (2019). Linguistic multimodality–multimodal linguistics: A state-of-the-art sketch. Multimodality: disciplinary thoughts and the challenge of diversity. *de Gruyter, Berlin/Boston*, 41–68. https://doi.org/10.1080/21504857.2019.1651358

Wang, J., Cheng, R., & Liao, P. C. (2021). Trends of multimodal neural engineering study: a bibliometric review. *Archives of Computational Methods in Engineering*, *28*(7), 4487–4501. https://doi.org/10.1007/s11831-021-09557-y

Wang, Y. (2022). Research on the method of educational text classification based on deep learning. *International Journal of Continuing Engineering Education and Life Long Learning*, *32*(3), 313–326. https://doi.org/10.1504/IJCEELL.2022.124032

Wang, Y. (2017). Construction elements and path of practical education model in universities. *EURASIA Journal of Mathematics, Science and Technology*, 13(10): 6775–6782. https://doi.org/10.12973/ejmste/78525.

Wang, Y., & Hemchua, S. (2022). Can we learn about culture by EFL textbook images? A semiotic approach perspective. *Language Related Research*, *13*(3), 479–499. https://doi.org/10.29252/LRR.13.3.18.

Wang, Y. (2023). Probing into the boredom of online instruction among Chinese English language teachers during the Covid-19 pandemic. *Current Psychology*, *43*(1),1–15.https://doi.org/10.1007/s12144-022-04223-3.

Wang, Y., Derakhshan, A., Pan, Z., & Ghiasvand, F. (2023a). Chinese EFL Teachers' Writing Assessment Feedback Literacy: A Scale Development and Validation Study. *Assessing Writing*, 56, 1–16. https://doi.org/10.1016/ j.asw.2023.100726.

Wang, Y., Pan, Z. W., & Wang, M. Z. (2023b). The moderating effect of participation in online learning on EFL teachers' teaching ability. *Heliyon*, 9(3)e13890, 1–12. https:// doi.org/10.1016/j.heliyon.2023.e13890.

About the Authors

Yanli Zhao was born in Hebei, China, in 1984. From 2002 to 2006, she studied in Zhengzhou Shengda University of Finance and Economics and received her Bachelor's degree in 2006. From 2006-2010, she worked in Foreign Languages Department in Zhengzhou Business University. From 2010 to 2014, she studied in Zhengzhou University and received her Master's degree in 2014. Currently, she works in Zhengzhou Business University. She also studied her Phd program majored in Tesol in Universiti Kebangsaan Malaysia from March, 2022. She has published 10 papers. Her research interests include Linguistics and Teaching English as a Second Language.

Nur Ainil Binti Sulaiman was born in Malaysia. She studied in University Putra Malaysia and got Bachelor's degree in 2010. From 2010-2013, she studied in Universiti Kebangsaan Malaysia and was awarded Master degree. In 2021, she received her Doctor degree in Universiti Kebangsaan Malaysia. Currently, she works in Universiti Kebangsaan Malaysia. She has published more than 20 papers. Her interests are included Second Language Acquisition, vocabulary learning and sociolinguistics. Email: e-mail: nurainil@ukm.edu.my

[DOR: 20.1001.1.23223081.1401.0.0.322.4]

[DOI: 10.29252/LRR.14.5.11]