Hindawi

Retraction

Retracted: The Evaluation Model of College English Diagnostic Exercises Based on Machine Learning

Journal of Function Spaces

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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] Q. Wang and H. Wu, "The Evaluation Model of College English Diagnostic Exercises Based on Machine Learning," *Journal of Function Spaces*, vol. 2022, Article ID 9185827, 12 pages, 2022.

Hindawi

Research Article

The Evaluation Model of College English Diagnostic Exercises Based on Machine Learning

Qi Wang [ID] [1] and Hao Wu [ID] [2]

[1]*Department of Public, West Anhui Health Vocational College, Lu'an, Anhui 237005, China*
[2]*Electromechanical Department, Anhui Chuzhou Technician College, Chuzhou, Anhui 239000, China*

Correspondence should be addressed to Qi Wang; wangqi9828@ustc.edu

Online learning is an important way for college students to learn English independently. The evaluation information provided by the previous online teaching platform is more summative evaluation, which cannot make students have a more intuitive and comprehensive understanding of their English learning status and lack of personalized guiding suggestions. Therefore, this paper combines data mining technology with machine learning to build an English diagnostic exercise model that can analyze students' learning status, the correlation between knowledge points and question types, and predict English achievement, so as to provide students with more comprehensive analysis data information. The experimental results show that the evaluation model of college English diagnostic practice based on machine learning has the classification results of learning state with finer granularity, effectively analyzes the association rules of knowledge points and question types, and has high prediction performance. It can help students fully understand their English learning status, provide students with personalized analysis data and effective guiding suggestions, and enhance students' English application ability, improving CET-4.

1. Introduction

Since the reform and opening up, China has made quite gratifying achievements in English teaching. However, there are still some problems and deficiencies in our teaching mode and teaching methods. China does not pay enough attention to English teaching, especially the cultivation of English application ability of higher vocational students. There are phenomena such as "unable to understand and speak clearly," and the students' English level is very low. The application ability of English in daily communication is seriously insufficient, and the basic function of English as a tool of interpersonal communication cannot be fully played. At present, the English teaching of higher vocational students in China focuses on exam-oriented education. Students spend a lot of time in memorizing words and reciting grammar but have little training in English listening and speaking. For a long time, it has developed into "deaf English" and "dumb English." English proficiency and application ability is one of the necessary basic conditions for

many enterprises to recruit fresh college students. CET-4 and CET-6 are also one of the important examinations that college students need to pass before graduation. The results of CET-4 and CET-6 can largely reflect the English foundation and application ability of college students, as well as the effect and level of college English teaching [1]. College teaching and middle school teaching have different emphases. English teaching and examination pay more attention to the comprehensive application ability after mastering basic knowledge and skills, which is like a mountain in front of the door for students with poor English foundation [2]. English self-study and practice in spare time is particularly important for college students. The rise of online education and learning has opened up a new way of self-study for college students [3]. Compared with traditional college English teaching, online teaching and learning methods have many choices and wide resources. Students can arrange learning time and content according to their actual situation to meet students' personalized English practice requirements [4]. In addition, online teaching teachers can realize one-to-one or

one to a few students' teaching and have relatively sufficient time to solve and correct the problems and deviations existing in students' English learning [5]. E-learning has relatively high requirements for learners. Students need to clarify their learning objectives, fully understand their current learning status, and clearly plan their learning plans and contents [6]. Although many e-learning platforms can provide students with English performance tests, the whole e-learning framework process is different from the traditional teaching mode in essence. In most cases, only the status of obtaining scores, question type answers and the analysis of question gains and losses [7, 8]. Based on this, students can only understand their own gains and losses in a certain type of English questions but cannot master their own deficiencies in knowledge points and English skills and lack of targeted and guiding data information. The application of big data technology and machine learning technology can carry out data mining for the massive information data of online English learning system, enhance the analysis of English learners' personalized information and question type relevance information, provide English diagnostic evaluation and guiding suggestions, reduce detours in students' learning process, reduce learning effect deviation, and improve learning efficiency and effect [9].

The innovative contribution of this paper is to establish a diagnostic evaluation model of English in online English learning system by combining data mining, association rules, and random forest model. The model includes student learning state evaluation module, English information correlation analysis module, and English performance prediction module. The corresponding performance experiments, teaching application comparative experiments, and experimental results are analyzed. The evaluation model of college English diagnosis practice based on machine learning has a finer granularity of learning state classification results, effectively analyzes the association rules of knowledge points and problem types, and has high prediction performance. It can help students fully understand their English learning situation, provide students with personalized analysis data and effective guidance and suggestions, improve students' English application ability, and improve college English level.

2. Development and Research Status of Diagnostic Exercise Evaluation Model

The educational process is a circular process. Both educators and learners need to obtain evaluation information in the teaching process as the evaluation standard for the current teaching and learning effect [10]. Educational evaluation is phased and summative. Suggestions that can help students find and solve problems in the process of students' learning play the role of phased evaluation, and the evaluation of students' final achievements and achievements is summative evaluation [11]. There are differences between the two kinds of evaluation for students' learning guidance. The combination of the two can provide students with systematic evaluation information and help students comprehensively understand their own learning situation. Diagnostic evalu-

ation is to provide students with learning status analysis, learning effect feedback, problem analysis, supervision, and early warning of students' learning status and assist students in formulating effective learning plans [12, 13].

The development of network education has changed people's learning methods and ideas. Many learners who are not satisfied with the current situation continue to improve their abilities and expand their knowledge base through network teaching, which also makes the network education platform pay more and more attention to the data information services provided for learners. Diagnostic practice system has become the focus and hotspot of the research of network autonomous learning platform [14]. Some scholars proposed an adaptive learning system based on diagnostic evaluation, hoping that the system can evaluate the learning ability level and learning status of different learners [15]. Based on this, some scholars have carried out system optimization and established a learning system that can recommend exercises according to learners' current learning state. Learners can more truly show the learning effect through exercises with appropriate difficulty [16, 17]. Some companies have introduced the concept of diagnostic evaluation into teaching to form a new teaching model. Through the application experiment of teaching model, it is found that students are easier to show the advantages and disadvantages of their learning and mastery in this model. Educators provide personalized guidance according to students' specific conditions and strengthen students' weak links [18]. Other scholars have studied the computer adaptive test system according to the English learning situation of college students to provide diagnostic evaluation for college students [19]. With the diversified development of college students' needs for English evaluation, some scholars have studied the English learning system facing mobile devices to break the restrictions of learning conditions and realize personalized learning [20]. Other scholars have proposed that English learning is a gradual process, and leapfrog learning has poor effect on most students. Therefore, they classify English reading according to different difficulty levels and provide reading content with corresponding difficulty according to the specific situation of students [21]. At present, the research on English learning system with diagnostic evaluation function is still in the development stage, and the realization of many functions is not ideal. Personalized data analysis and targeted test paper generation module need to be further improved.

3. Construction of College English Diagnostic Practice Evaluation Model Based on Machine Learning

3.1. Overall Framework of Diagnostic Exercise Model. In the current college English evaluation system, there is often a phenomenon that students are only evaluated by their examination results. The evaluation of English teachers' teaching mainly focuses on the completion of the syllabus and students' examination results. This is not conducive to the improvement of students' comprehensive ability and
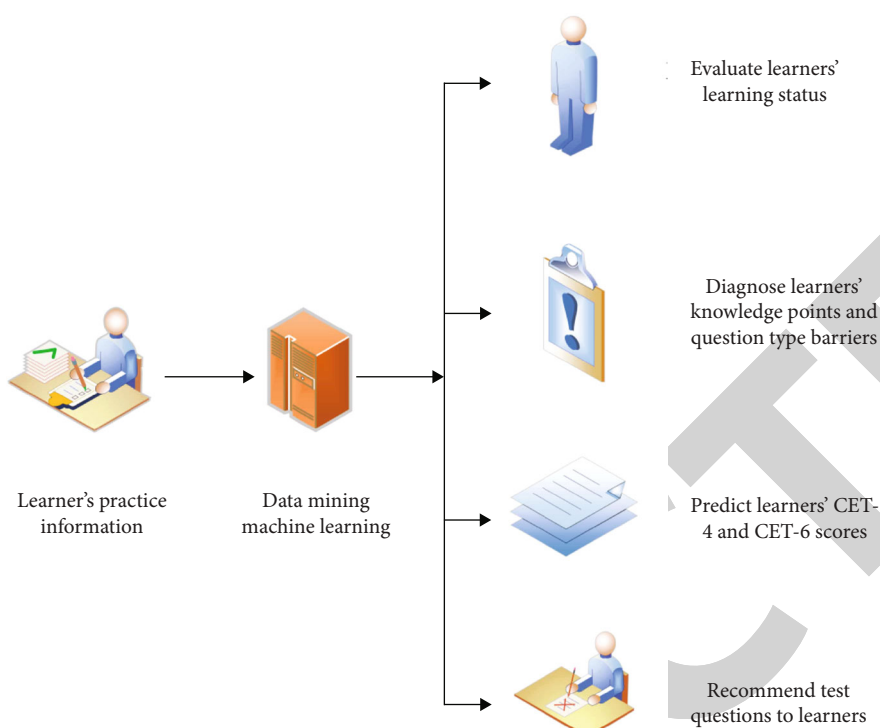
FIGURE 1: Block diagram of college English diagnostic practice evaluation model based on machine learning.

teachers' innovative teaching methods. In view of this phenomenon, we should build a diversified diagnostic system of college English, which is of great significance for reforming this unscientific and unfair teaching method. CET-4 and CET-6 are examinations to evaluate college students' English ability from many aspects, paying attention to students' basic English and comprehensive application ability [22]. However, different students have individual differences in learning ability, understanding ability, reading ability, and application ability; that is, each college student has different problems in English learning status, mastery level of knowledge points, exercise practice status, and so on. Therefore, the evaluation model of college English diagnostic practice should provide targeted, effective, and diversified English practice modes and guidance according to the actual situation of college students. However, the early diagnostic evaluation system still maintains the traditional teaching evaluation mode and ignores the actual state of college students, students can obtain less learning evaluation information and single guidance suggestions, and the direction of learning improvement is not clear [23]. Therefore, this paper introduces the evaluation factors of students' learning status into the model to help students understand their actual situation on the basis of evaluating the current learning ability of college students. The model also analyzes the correlation between English knowledge points and question types to predict students' scores in college English tests. According to students' practice data and feedback data information, focus on adjusting students' English learning arrangements, and pay attention to students' personalized factors. In addition, predict and compare the students' performance,

dynamically show the changes of students' state, analyze the negative factors, and achieve the purpose of supervision and early warning. Figure 1 shows the block diagram of college English diagnostic practice evaluation model based on machine learning.

As can be seen from the above figure, the workflow of the model mainly includes four stages. The first stage is to extract, process, analyze, and calculate the data related to college students' English tests and exercises. The second stage takes students as the core of data information analysis, makes a relatively fair and overall learning state judgment from the test score and students' learning stability, and sends early warning information to students with poor stability and slow overall learning process. In the third stage, analyze the knowledge points and question types of students' possible problems through data correlation and relevant algorithms, effectively find the shortcomings of students, and practice and improve their corresponding abilities. The fourth stage is to predict the score of CET-4 and CET-6 according to the existing student test and practice data. Based on this, students can comprehensively and intuitively understand their current mastery of CET-4 and CET-6.

3.2. Student Learning State Evaluation Module. Generally, the S-P table analysis method will be selected to evaluate students' learning status. The two-dimensional indicators for analyzing students' learning status are the scores of English knowledge points and question types and the classification status evaluation of students' attention coefficient. Let the probability that the n student gets a score be m and
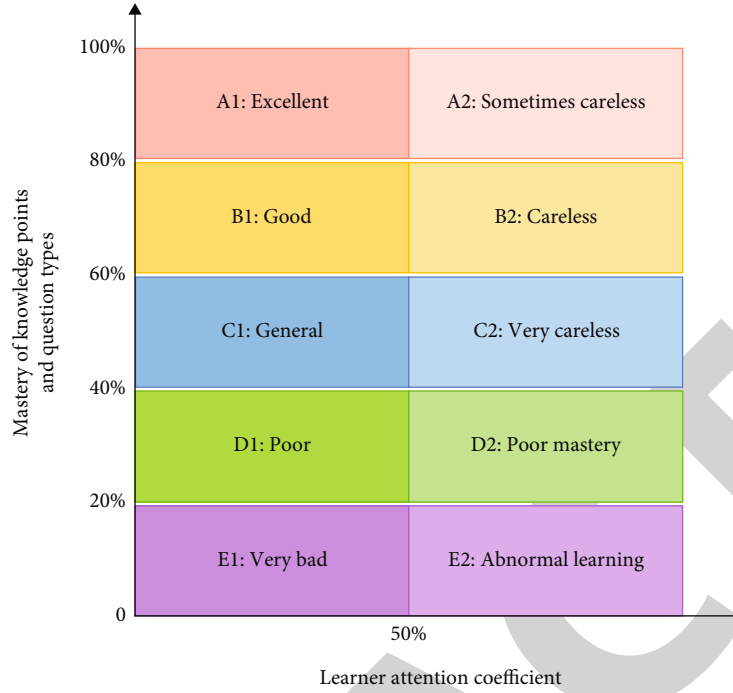
FIGURE 2: Results of students' English learning status score differentiation.

write y_{nm}, and the calculation of attention coefficient is as shown in

$$\mathrm{CS}_n = 1 - \frac{\sum_{m=1}^{a}(y_{nm})(y_m) - (y_n)(\eta)}{\sum_{m=1}^{y_n} y_m - (y_n)(\eta)}, \tag{1}$$

where the total score obtained by students is y_n, the number of correct answers to knowledge points is y_m, the average value of correct answers is η, and the attention coefficient is CS_n.

The mastery status of knowledge points and question types is calculated according to

$$S_{\mathrm{master}} = \frac{N_k}{N_t}, \tag{2}$$

where students' mastery status is recorded as S_{master}, N_k represents the number of knowledge point question types whose score rate is higher than the average, and N_t represents all knowledge point question types.

There are three evaluation results of S-P table analysis method, but such evaluation granularity is relatively large. In practical application, the classification of students' status is a little rough and inaccurate. Therefore, this paper refines the classification of students' learning state on its basis, as shown in Figure 2. The scoring results of refined learning state are divided into five types, each of which is divided into two cases. The first category of the level is when the students' learning state and attention coefficient remain stable at the current level. The second category is when the students' learning state is unstable, but the attention coefficient

remains at the current level. In this way, students' learning state can be divided into ten levels to increase the accuracy and personalized effect of evaluation.

The model stores the information of students' knowledge points and question type scores as matrix data and processes the data into a data form that can be used by the evaluation algorithm. Set the threshold value as the average score rate of knowledge points and question types. The data not less than this value is 1; otherwise, it is 0. Judge according to formula (3),

$$\mathrm{Line}(x_{nm}) = \begin{cases} 0, x < \dfrac{\sum_{n=1}^{a} x_{nm}}{a}, \\[4mm] 1, x \geq \dfrac{\sum_{n=1}^{a} x_{nm}}{a}. \end{cases} \tag{3}$$

3.3. Knowledge Point Type Relevance Analysis Module.
CET-4 and CET-6 test the comprehensive application of students' English knowledge and ability. There is correlation between English words, sentences, grammar, and other knowledge points, as well as between the test questions, which cannot be analyzed according to the independent situation. Therefore, this paper integrates data mining technology and association analysis technology into the analysis module for association analysis. Firstly, the massive original data is processed into effective analysis data; that is, the data due to invalid operation, lack of information, and abnormal errors are excluded, and the repetitive English practice results are screened out. The difficulty of English exercises in the model is different. The average score rate of a certain question type is taken as the judgment standard; that is, the
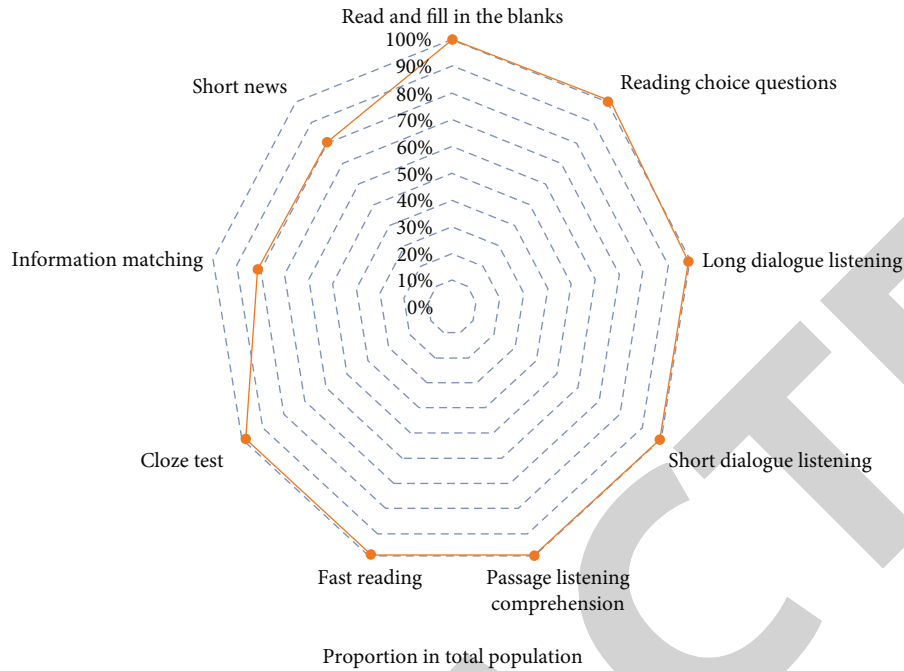
FIGURE 3: Comparison of the number of learners of each English test type before the update of English test type.

students' scores are divided into two types according to the standard, as shown in

$$r_{\text{correct}(nm)} = \begin{cases} F, r_{\text{correct}(nm)} < \overline{r_{\text{correct}(nm)}}, \\ T, r_{\text{correct}(nm)} \geq \overline{r_{\text{correct}(nm)}}, \end{cases} \quad (4)$$

where the probability of students' question type score is recorded as $r_{\text{correct}(nm)}$ and its average value is expressed as $\overline{r_{\text{correct}(nm)}}$.

The form and content of CET-4 and CET-6 in colleges and universities are not invariable. It is adjusted according to the actual needs and the overall quality of modern college students. Figure 3 shows the comparison of the number of learners of each English test type before the update of the English test type. The number in the figure shows that before the adjustment of the test type, some students were relatively weak in information matching and short news practice. The real English questions before adjustment are still of great help to students' practice. Therefore, it is necessary to deal with the English question type data in layers and analyze the data relevance, respectively.

There are two rule evaluation indicators for data association analysis. One is support, as described in

$$S(X \longrightarrow Y) = \frac{P(X \cup Y)}{P(I)}. \quad (5)$$

Another confidence is shown in

$$C(X \longrightarrow Y) = P(Y|X) = \frac{P(X \cup Y)}{P(X)}. \quad (6)$$

3.4. Students' English Score Prediction Module of CET-4 and CET-6. Hidden data may lead to the deviation of mining results, so data cleaning is particularly important. After the completion of data cleaning, a series of processes such as data integration, transformation, and specification are carried out at the same time. This process is data preprocessing. Statistically speaking, the hidden data may produce biased estimates, so that the sample data cannot well represent the population, and most of the data in the implementation contain missing values, so how to deal with missing values is very important. The processing includes two steps, namely, the identification of hidden data and the processing of hidden values. Commonly used methods include deletion method, replacement method, and interpolation method. After the preprocessing of the original data, the module needs to further mine the hidden features between the data, that is, the score of students' English questions and their gender. Students of different genders have different thinking performance, which has a great impact on their grades. The scoring rate is calculated as shown in

$$r_{\text{score}} = \frac{N_r}{N_t}. \quad (7)$$

The number of correct questions answered by students is N_r.

The prediction model of this paper selects the random forest model and combines it with multiple linear regression model to optimize the performance. Decision tree is the basic unit of random forest model. Its construction process includes decision tree generation and pruning. The information gain rate of decision tree is the factor

to be considered when selecting attribute division, and its formula is shown in

$$r_{\text{Gain}}(A) = \frac{\text{Gain}(A)}{S(A)}, \tag{8}$$

where the information gain is represented by $\text{Gain}(A)$ and the split information value is represented by $S(A)$, as shown in

$$\text{Gain}(A) = S_{\text{Info}}(D) - S_{\text{Info}(A)}(D), \tag{9}$$

$$S_A(D) = -\sum_{m=1}^{U} \frac{|D_m|}{|D|} \times \log_2\left(\frac{|D_m|}{|D|}\right). \tag{10}$$

Random forest model adds the algorithm of random selection of attributes to the training of unit decision tree and completes the construction of unit decision tree with bagging integration. It shows the advantages of relatively easy implementation conditions and strong performance in practical application. The best classification index of random forest is selected according to the principle of Gini coefficient. The calculation formula is shown in

$$\text{Gini}(t_n) = 1 - \sum_{m=1}^{2} [P(Y_m|t_n)]^2, \tag{11}$$

where the probability of dividing node t_n is expressed as $P(Y_m|t_n)$.

Bagging algorithm is derived from autonomous sampling method, which is used to enhance the generalization ability of unit decision tree and improve the prediction accuracy of random forest model. The random forest model completes the prediction and estimation through out of bag data, and its accuracy prediction is shown in

$$Q_n(x,y) = \frac{\sum_{n=1}^{p} I(h_n(x) = y, (x,y) \in O_n(x))}{\sum_{n=1}^{p} I(h_n(x), (x,y) \in O_n(x))}, \tag{12}$$

where the denominator represents the number of unselected samples and the numerator represents the number of correctly divided unselected samples.

The prediction results are shown in

$$Y_f = \arg\max_y \sum_{p=1}^{\text{ntree}} \delta(h(x,\theta_p) = y), \tag{13}$$

where ntree is the number of decision trees and $\delta(\bullet)$ is the indicative function.

The model generalization error formula and interval function are shown in

$$\text{PE}^* = P_{x,y}(\text{mg}(x,y) < 0), \tag{14}$$

$$\text{mg}(x,y) = \text{av}_p \delta(h_p(x) = y) - \max_{m \neq y} \text{av}_p \delta(h_p(x) = m). \tag{15}$$

In the formula, the average value of the correctly predicted sample is recorded as $\text{av}_p \delta(h_p(x) = y)$, the maximum value is recorded as $\max_{m \neq y} \text{av}_p \delta(h_p(x) = m)$, and the generalization error is represented as PE^*.

The expression of multiple linear regression equation is shown in

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p + v. \tag{16}$$

In the formula, the target variable is y, the variable is $x_1, x_2, \cdots, x_p$, the parameter to be solved is $\alpha_m, m = 0, 1, 2, \cdots, p$, and the random error is v. The parameters of the demand solution are estimated through the observation matrix vectors of the target variable and the variable, i.e., y_n and $x_{1n}, x_{2n}, \cdots, x_{pn}$. If the estimated parameter value is recorded as $\widehat{\alpha}_m, m = 0, 1, 2, \cdots, p$, the multivariate linear sample regression equation is shown in

$$\widehat{y}_n = \widehat{\alpha}_0 + \widehat{\alpha}_1 x_{1n} + \widehat{\alpha}_2 x_{2n} + \cdots + \widehat{\alpha}_p x_{pn}, \tag{17}$$

where $\widehat{y}_n$ is the estimated value.

The difference equation between the estimated value and the actual value of the target variable is shown in

$$e_n = y_n - \widehat{y}_n = y_n - (\widehat{\alpha}_0 + \widehat{\alpha}_1 x_{1n} + \widehat{\alpha}_2 x_{2n} + \cdots + \widehat{\alpha}_p x_{pn}). \tag{18}$$

RMSE and RMSLE are the verification indicators of the prediction module, and their calculation formulas are shown in

$$\text{RMSE} = \sqrt{\frac{\sum_{n=1}^{a}(x_{\text{obs},n} - x_{\text{mdl},n})}{i}}, \tag{19}$$

$$\text{RMSLE} = \sqrt{\frac{1}{i}\sum_{n=1}^{a}(\log(x_{\text{obs},n}+1) - \log(x_{\text{mdl},n}+1)^2}. \tag{20}$$

In the formula, the predicted value represents $x_{\text{obs},n}$ and the actual value represents $x_{\text{mdl},n}$.

4. Experimental Results of College English Diagnostic Practice Evaluation Model Based on Machine Learning

There are some differences between machine learning and deep learning. First, the application scenarios are different. The application of machine learning in fingerprint recognition, feature object detection, and other fields has basically met the requirements of commercialization. Deep learning is mainly used in word recognition, face technology, semantic analysis, intelligent monitoring, and other
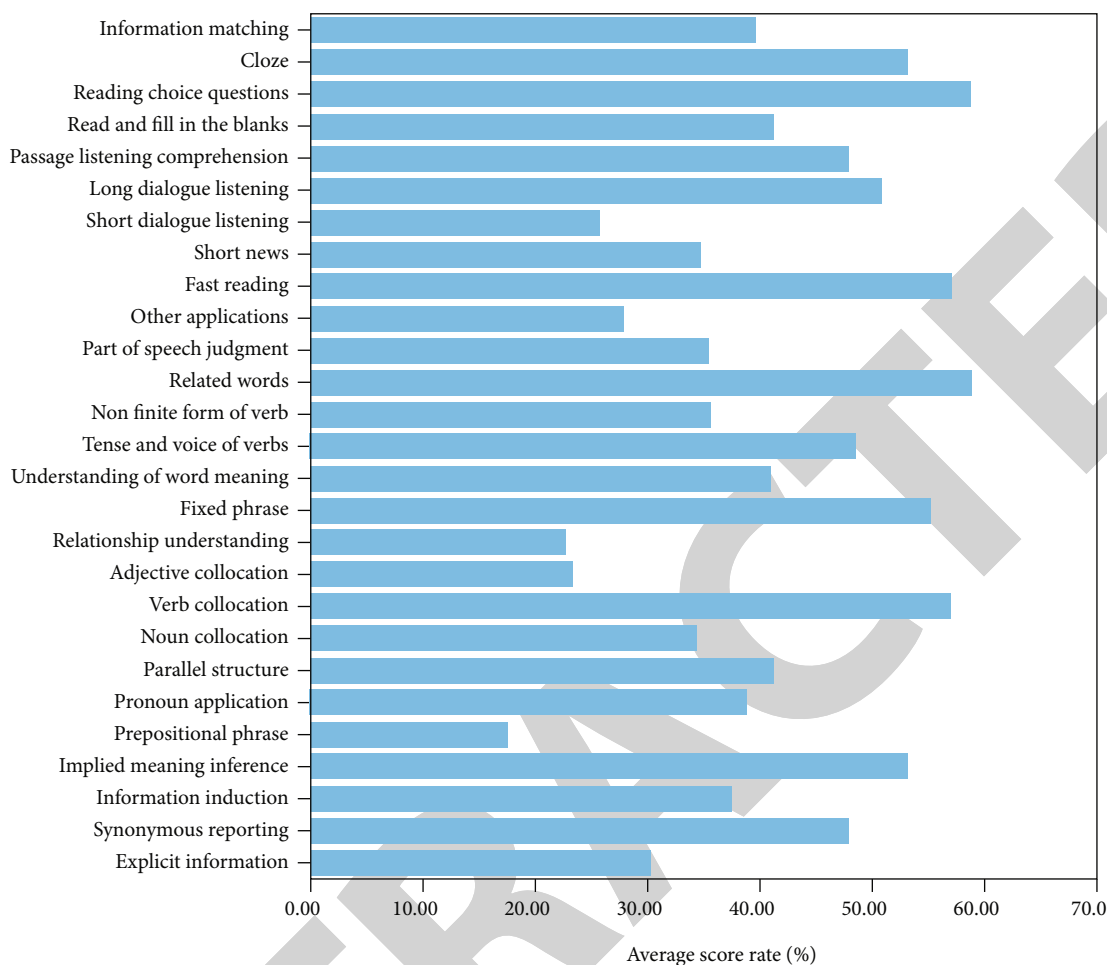
Figure 4: The average score of all students participating in the experiment before the experiment.

fields. At present, it is also rapidly deployed in intelligent hardware, education, medical, and other industries. Secondly, the required amount of data machine learning can adapt to a variety of data, especially in scenarios with a small amount of data. If the amount of data increases rapidly, the effect of deep learning will be more prominent, because deep learning algorithms require a lot of data to be perfectly understood. The execution time is different. The execution time refers to the amount of time required by the training algorithm. Generally speaking, deep learning algorithm requires a lot of time to train. This is because the algorithm contains many parameters, so it takes longer time to train them than usual. Relatively speaking, machine learning algorithms take less time to execute.

The application of the evaluation model of college English diagnostic practice based on machine learning was aimed at evaluating students' current English ability, providing targeted and personalized suggestions for students, and improving their English test scores. Therefore, this paper selects two classes of unified major in a university for a comparative experiment of English performance. Class one is an experimental class; that is, the model of this paper is introduced into English teaching. The second class is the control

class, whose English scores before the experiment are similar to those of the experimental class, and both classes are CET-4-nonassessed classes. Before the comparative experiment, the performance of the three modules of the college English diagnostic practice evaluation model based on machine learning is tested to ensure the stability and reliability of the comparative experiment. Take the exercises or homework that students usually do as the training data and the examination organized by the teacher as the verification data, and the final exam we take is called the test data. These three parts of module data are all existing data, in which the training data is only used for training. The goal of students (learning algorithms) is to continuously improve performance on this part of data. The teacher will use verification data to monitor students' learning and then adjust students' learning methods. When the teacher thinks that the student can no longer improve, he uses the test data to simulate the unknown data to conduct the final performance test on the student.

As shown in Figure 4, it is the average result of the score rate of English knowledge points and question types of all students participating in the experiment before the experiment. The data results show that the distribution range of the score rate of students participating in the experiment is

	Fast reading	Short news	Short dialogue	Long dialogue	Passage listening	Banked cloze	Cloze	Information matching	Reading choice
No.1	F	F	F	F	F	F	F	F	F
No.2	F	F	F	T	T	T	F	T	F
No.3	T	T	F	F	F	F	F	F	T
No.4	F	T	F	F	T	F	T	T	F
No.5	F	F	T	T	F	F	T	F	T
No.6	F	F	F	F	F	T	F	T	F
No.7	F	F	F	F	F	F	F	F	F
No.8	T	F	T	F	F	F	T	T	F
No.9	F	T	T	T	T	T	F	T	F
No.10	F	F	T	F	T	T	F	T	F

FIGURE 5: The result of binary Boolean of the score rate of ten students' CET-4 questions.

relatively concentrated, mainly between 30% and 59%. If students are classified according to the traditional S-P table analysis method, the results obtained are too concentrated, which is not conducive to the further analysis of students' learning state. The refined classification table can more refine the classification of students and better show the differences of student groups.

Ten students are randomly selected from the experimental class, and the binary Boolean result of the score rate of level 4 question type is shown in Figure 5. The results showed that before the comparative experiment, most of the ten students did not reach the average value, and even some students scored lower than the average value.

Further analyze the relevance of the judgment data in Figure 5, and analyze the relevance of the question types with high and low scoring rate under the same setting conditions. The analysis results are shown in Figure 6. Two layers are included in T and F analyses, and the minimum reliability is 90%, excluding the data with low probability.

Select the one with the highest confidence and support of association rules in the four analysis layers of T and F, as shown in Figure 7. The support degree of association rules reflects the probability that the front and back association items, i.e., item a and item b exist at the same time, and the confidence degree indicates the possibility that item a also exists within the existence time of item b. It can be seen from this that in T analysis, when the three question types of information matching, rapid reading and reading comprehension all reach the average standard at the same time; the probability is 41.89%. When the first two question types are not less than the average value, the probability of reading comprehension score rate not less than the average value is

96.52%. By further analyzing the relationship between reading comprehension and information matching, the probability that the scores of the two types are not less than the average at the same time is 46.85%. When the score of information matching question type is greater than or equal to the average, the possibility of reading comprehension also reaching this condition is 96.04%. In the F analysis, the first level analyzes the correlation among short passage, long dialogue, and short dialogue listening questions. The probability that the scores of the three questions are lower than the average score standard is 45.97%. When the scores of the first two types of questions are lower than the average standard at the same time, the probability of short dialogue listening comprehension is also lower than the average standard, up to 99.57%. In the second layer, the probability that the scores of reading selection, word filling, and information matching cannot meet the average standard at the same time is 32.89%. When the scores of the first two types do not meet the average standard, the probability that the scores of information matching cannot meet the condition is 95.81%. To sum up, among the CET-4 and CET-6 question types, the question types belonging to the same category have high internal relevance. When students have low accuracy of some question types in the same category, the probability of low accuracy of other question types in this category is very high. In addition, the relevance of the same type of questions is high, which reflects that no matter how the CET-4 and CET-6 question types are updated, students can be familiar with the changes of new questions and enhance their English application ability through the practice of changing the former question types.
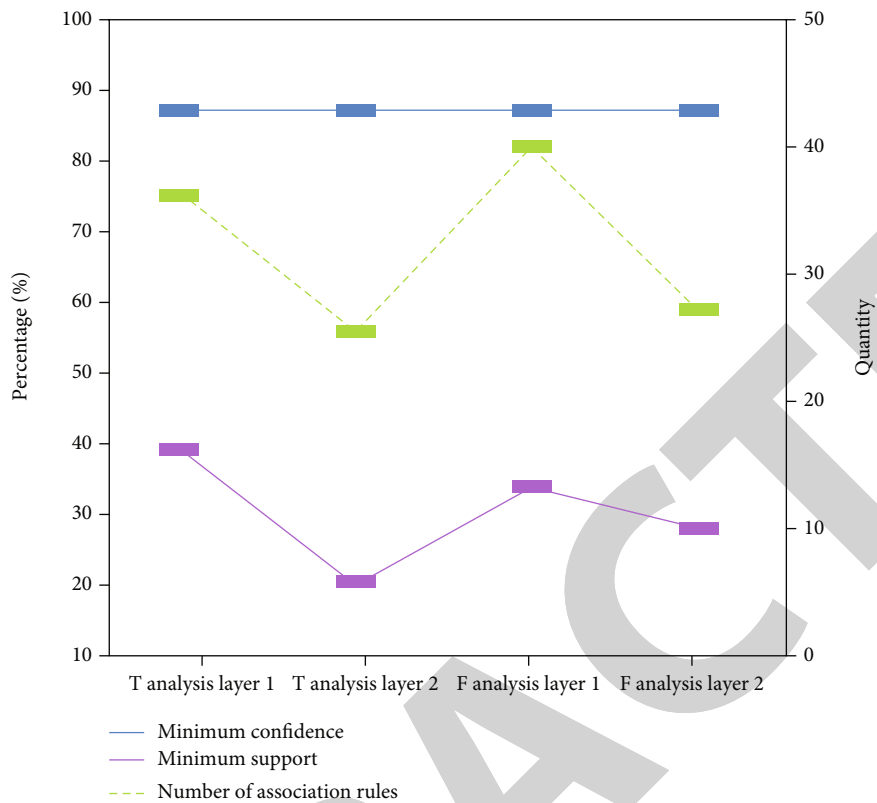
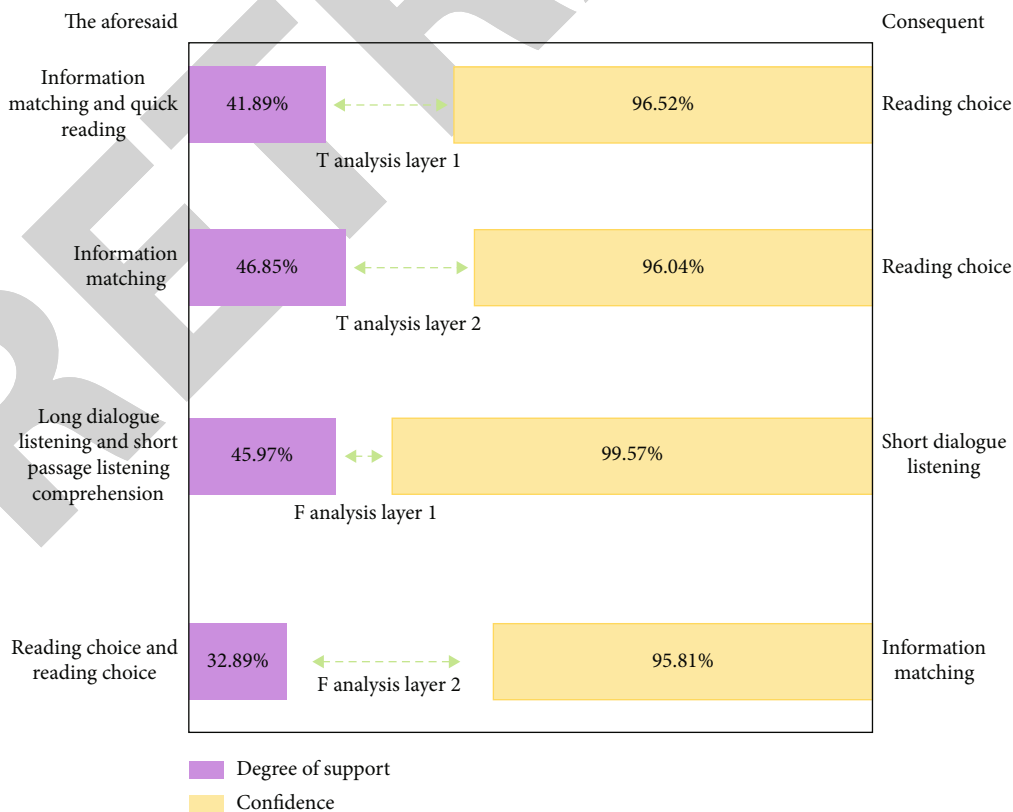Figure 6: Model association rule analysis output.

Figure 7: The association rule results of question types with the highest confidence and support among all analysis levels.
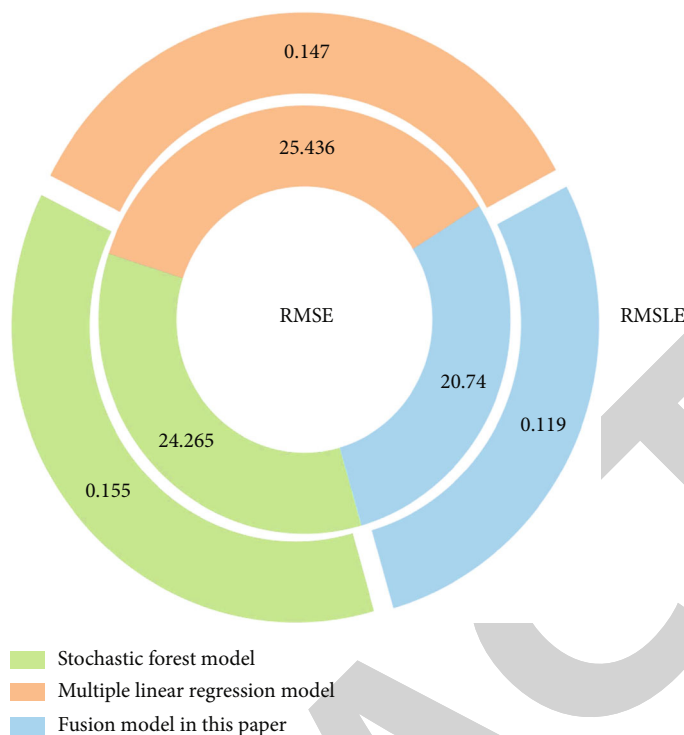
FIGURE 8: Comparison results of RMSE and RMSLE evaluation indexes of the three models.

Figure 8 shows the comparison results of RMSE and RMSLE evaluation indicators of the three models. The comparison results show that the prediction module is only established based on random forest model or multiple linear regression model, and there is little difference between the evaluation index results of the two. After the fusion of the two models, the evaluation index values of the model change obviously, which shows that the model in this paper has a great improvement in the two indexes and enhances the prediction accuracy of the model.

According to the above experimental results, the evaluation model of college English diagnostic practice based on machine learning can divide students' current English learning state in a fine granularity, better realize the correlation analysis between knowledge points and question types, and provide more accurate English performance prediction results for the comparative experiment. In the comparative experiment, the experimental class and the control class will be tested on CET-4 before the experiment, that is, before school and after the experiment, that is, at the end of the semester. In addition to the zero CET-4 test in the experimental class, the college English diagnostic practice evaluation model based on machine learning also predicts the students' English test scores before the second test. The results are shown in Figure 9. The results in the figure show that before the experiment, the English scores of the two classes are basically the same, and most of them are between 280 and 390, which shows that the application ability of CET-4 of the students in the two classes is weak, and most of them still fail to meet the passing standard. After the experiment, the results of the students in the experimental class have greatly improved the results of the students in

the control class, the number of students with low scores has been significantly reduced, and the number of students with up to standard scores is also higher than that in the control class. In addition, there is a close relationship between the actual score of CET-4 in the experimental class and the predicted score of college English diagnostic practice evaluation model based on machine learning, and there is a relatively large error between the actual score and the predicted score of only a few students. This shows that the model can effectively help students analyze the current English learning state, provide effective suggestions and help for students according to the analysis data, enhance students' weak English application ability, systematically practice English question types, and quickly improve students' learning efficiency and adaptability to different question types. The results of the prediction model fit the actual results, more intuitively reflect the students' English learning progress and overall level, promote the students to check and make up the deficiencies, strengthen the students' self-awareness, and achieve the purpose of improving their English performance.

Figure 10 shows the evaluation results of the experimental class students on the evaluation model of college English diagnostic exercises based on machine learning. The evaluation content includes the students' satisfaction with model diagnostics, learning state analysis, knowledge point diagnosis, and question type diagnosis, with a full score of five. The evaluation results show that the average performance scores of the model are more than 3.5 points, indicating that most students are satisfied with the application performance of the model, which indirectly reflects that the diagnostic effect of the model can meet the needs of most students and
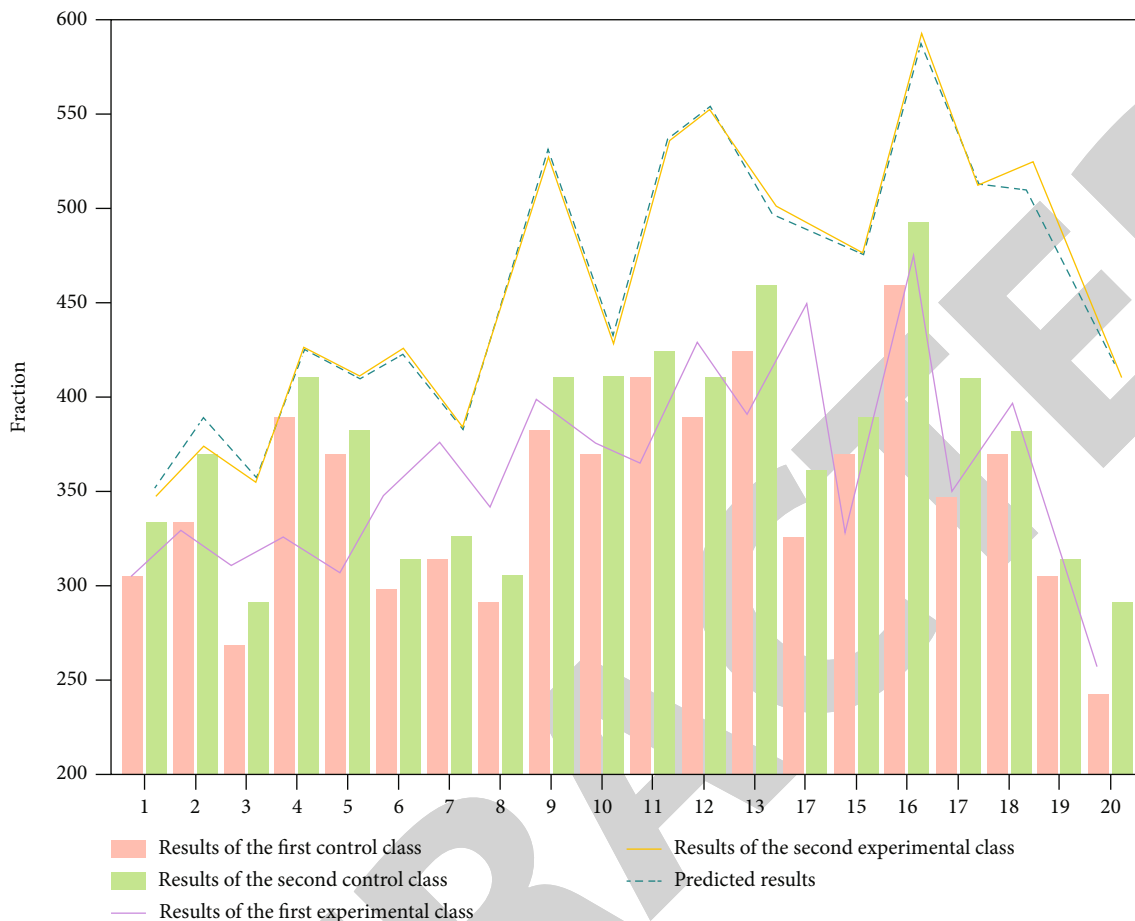
FIGURE 9: The application of test scores before and after the experiment in the experimental class and the control class and the student model in the experimental class predict the results of English scores.
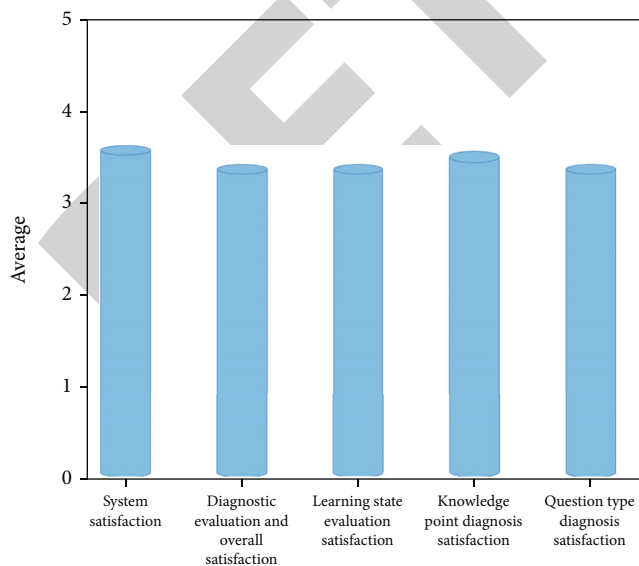

FIGURE 10: The evaluation results of students in the experimental class on the evaluation model of college English diagnostic exercises based on machine learning.

achieve the purpose of the experiment in practical application.

5. Conclusion

This paper combines data mining, association rules, and random forest model to build an English diagnostic exercise model to analyze students' English learning status, knowledge mastery, problem type correlation, and English performance prediction. The experimental results show that the evaluation model of college English diagnosis practice based on machine learning can refine the classification of students' learning status and show the differences of different students' English learning status. The evaluation index of machine learning based college English diagnostic practice evaluation model performs well and has high accuracy. The results of teaching comparison experiments show that the model can help students understand their English learning as a whole, better plan their learning plans, and improve their learning efficiency. Correlation analysis found that the problem types and knowledge points belonging to the same category have strong correlation, and the gain and loss points have a greater impact. The research provides students with intuitive learning state feedback information through English score prediction close to the actual value and realizes

the learning state early warning function. In addition, the evaluation results of the experimental class students on the model show that the model has good application performance, and its diagnostic evaluation can meet the needs of most students' autonomous learning.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] A. Z. A. Abu, "Telegram app in learning English: EFL students' perceptions," *English Language Teaching*, vol. 13, no. 1, pp. 51–62, 2019.

[2] Z. Xu and Y. Shi, "Application of constructivist theory in flipped classroom-take college English teaching as a case study," *Theory and Practice in Language Studies*, vol. 8, no. 7, pp. 880–887, 2018.

[3] Y. Zhang, "The realization of intelligent algorithm of knowledge point association analysis in English diagnostic practice system," *Complexity*, vol. 2021, 10 pages, 2021.

[4] R. Hanifa, "Factors generating anxiety when learning EFL speaking skills," *Studies in English Language and Education*, vol. 5, no. 2, pp. 230–239, 2018.

[5] Q. Meng, "Psycholinguistic analysis of the influence of affective factors on college English learning," *Summary of Educational Theory*, vol. 4, no. 1, p. 45, 2021.

[6] N. Baharin, N. Kamarudin, and U. K. A. Manaf, "Integrating STEM education approach in enhancing higher order thinking skills," *International Journal of Academic Research in Business and Social Sciences*, vol. 8, no. 7, pp. 810–822, 2018.

[7] P. R. Wu, "Cultural reflection on the reform and development of teaching evaluation thought in China," *Shangqiu Normal College Journal*, vol. 37, no. 1, pp. 97–102, 2021.

[8] X. Liu, P. K. Singh, and P. A. Pavlovich, "Accent labeling algorithm based on morphological rules and machine learning in English conversion system," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 881–892, 2021.

[9] W. Chang and L. Xibo, "Application of diagnostic assessment in English Reading teaching in senior high school," *Southern Journal*, vol. 4, pp. 89–91, 2019.

[10] T. Xi, "Design of English diagnostic practice sentence repetition recognition system based on matching tree and edge computing," *Wireless Communications and Mobile Computing*, vol. 2021, 8 pages, 2021.

[11] P. Dan, S. Xia, M. Xin, Z. Dan, and R. Baohua, "Research on teaching practice of breaking through students' thinking obstacles in learning organic chemistry," *Chemistry Education*, vol. 40, no. 5, pp. 30–36, 2019.

[12] A. Cano and J. D. Leonard, "Interpretable multiview early warning system adapted to underrepresented student populations," *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 198–211, 2019.

[13] S. Lee and J. Y. Chung, "The machine learning-based dropout early warning system for improving the performance of dropout prediction," *Applied Sciences*, vol. 9, no. 15, p. 3093, 2019.

[14] 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.

[15] Y. Huo, "Analysis of intelligent evaluation algorithm based on English diagnostic system," *Cluster Computing*, vol. 22, no. S6, pp. 13821–13826, 2019.

[16] Z. Liu, T. Li, and H. Diao, "The influence of multi-assessment model on the motivation of adults' English listening and speaking learning," *Theory and Practice in Language Studies*, vol. 10, no. 9, p. 1086, 2020.

[17] G. Akçapınar, A. Altun, and P. Aşkar, "Using learning analytics to develop early-warning system for at-risk students," *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, pp. 1–20, 2019.

[18] J. Liu, "Study on the influence of blended teaching mode on English learners' motivation under the background of information technology," *Journal of Physics: Conference Series*, vol. 1738, no. 1, article 012096, 2021.

[19] X. Qin, "Evaluation of English intercultural communication ability based on machine learning and fuzzy mathematics," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 6, pp. 1–13, 2020.

[20] B. Hoffman, "The influence of peer assessment training on assessment knowledge and reflective writing skill," *Journal of Applied Research in Higher Education*, vol. 11, no. 4, pp. 863–875, 2019.

[21] H. Liu and G. Liu, "An analysis of the listening comprehension in CET-6–illustrated by the CET-6 test on June, 2018," *Summary of Educational Theory*, vol. 3, no. 3, p. 5, 2020.

[22] S. Wang, "Simulation of English translation text filtering based on machine learning and embedded system," *Microprocessors and Microsystems*, vol. 83, no. 3, article 103982, 2021.

[23] X. Mai, "The application of formative evaluation strategies in college English classroom from the perspective of teacher questioning," *Open Journal of Modern Linguistics*, vol. 11, no. 6, pp. 873–882, 2021.