Methods for Improving English Vocabulary Learning Efficiency Based on Deep Learning Technology

Jiangliu Yu*

Anhui International Studies University, Hefei, 231201, China

*Corresponding author's e-mail: 820064428@qq.com

Abstract. To enhance the efficiency of English vocabulary learning for non-native speakers, this study employs deep learning technology to optimize the vocabulary learning process by constructing encoder-decoder models. The analysis shows that this method significantly improves learners' speed and quality in mastering specialized and low-frequency vocabulary, achieving a more personalized learning path, and has important implications for the modernization of language education.

Keywords: Deep learning technology; English vocabulary learning; personalized teaching

1 Introduction

With the advancement of globalization, the importance of English vocabulary learning for non-native speakers is increasingly prominent. However, traditional learning methods often fail to meet the demands of modern fast-paced learning due to inefficiency. Deep learning technology, as an advanced data processing method, provides new approaches for achieving personalized and efficient language learning. By applying this technology, this study aims to improve the process of English vocabulary learning, particularly in enhancing the mastery of specialized and low-frequency vocabulary, bringing innovative strategies to language education and validating their practicality.

2 Prospects of Applying Deep Learning Technology in English Vocabulary Learning

The prospects of applying deep learning technology in English vocabulary learning are promising, primarily because it can provide personalized learning solutions based on learners' specific circumstances [1]. By analyzing learners' learning history and behavior patterns, this technology can dynamically adjust teaching content and difficulty, achieving precise vocabulary teaching alignment. It can also effectively utilize big data analysis to filter the most suitable vocabulary from extensive corpora for learners' current levels, optimizing learning paths and improving learning efficiency.

C. Lin et al. (eds.), *Proceedings of the 2024 9th International Conference on Modern Management, Education and Social Sciences (MMET 2024)*, Advances in Social Science, Education and Humanities Research 880,
https://doi.org/10.2991/978-2-38476-309-2_30

3 English Vocabulary Learning Models Based on Deep Learning

3.1 Encoder Model

The encoder model utilizes neural networks, especially recurrent neural networks (RNNs) or their variants such as long short-term memory networks (LSTM) and gated recurrent units (GRU), to efficiently encode the input vocabulary data [2]. The encoding process involves converting each word into a vector representation in a high-dimensional space, capturing the complex relationships and semantic information between words. Specifically, the encoder operates through the following equations:

$$\begin{cases} h_t = f(W_h \cdot [h_{t-1}, x_t] + b_h) \\ c_t = g(W_c \cdot [h_{t-1}, x_t] + b_c) \\ y_t = \sigma(W_y \cdot c_t + b_y) \end{cases} \tag{1}$$

The input word vector at time step x_t is t, the hidden state is h_t, the cell state c_t captures and maintains long-term dependency information, and the output vector y_t represents the learned word features. W_h, W_c, W_y and b_h, b_c, b_y are the weight matrices and bias vectors, while f, g and σ are the nonlinear activation functions, such as sigmoid or tanh. The implementation of this model provides a solid theoretical foundation for personalized learning strategies and aids in designing more efficient teaching methods [3].

3.2 Decoder Model

The decoder model is primarily responsible for transforming the feature vectors extracted by the encoder into specific word outputs [4]. This process is achieved by using a network structure similar to the encoder, such as long short-term memory networks (LSTM). The focus is on predicting the appropriate use of vocabulary based on the context vector output by the encoder, and it can be described by the following equations:

$$\begin{cases} s_t = LSTM(s_{t-1}, y_{t-1}, c) \\ p_t = soft\max(W_p \cdot s_t + b_p) \\ y_t = \arg\max(p_t) \end{cases} \tag{2}$$

Here, s_t is the decoder's hidden state at time t, dependent on the previous state s_{t-1}, the previous output y_{t-1}, and the context vector c received from the encoder. This structure design ensures that each output step reflects the information of the entire input sequence, enhancing sensitivity to context [5]. p_t is the probability distribution over the vocabulary for each word, transformed from the decoder's output via the softmax function, indicating the next word. The weight matrix W_p and bias vector b_p play key roles in this computation. Finally, y_t is the output word at time step t, selected by performing an $\arg\max$ operation on the probability distribution p_t to choose the word with the highest probability. This method not only optimizes the vocabulary selection process but also aligns the learning method more closely with the natural language acquisition process, generating language output after fully understanding the input information.

3.3 Vocabulary Learning Model Architecture

The overall architecture is designed to achieve efficient vocabulary mastery and application. This architecture integrates the encoder and decoder models, forming a closed-loop learning system. Through this system, the model not only learns the vocabulary itself but also its usage context, enhancing the naturalness and practicality of language learning, as shown in Table 1.

Table 1. Vocabulary Learning Model Architecture Components

Component	Function Description	Key Technologies
Encoder	Extracts feature representations of input vocabulary	Recurrent Neural Networks (RNN), LSTM
Attention Mechanism	Focuses on important information in the input sequence	Attention weight matrix
Decoder	Generates vocabulary output based on context vector	LSTM, softmax
Optimizer	Adjusts model parameters to minimize error	Gradient Descent
Output Layer	Outputs predicted vocabulary	Activation function, output encoding

In this architecture, the addition of the attention mechanism significantly enhances the model's performance, allowing it to more accurately identify key information when processing input vocabulary. The calculation of the attention mechanism can be expressed through the following equations:

$$\begin{cases} a_t = soft\max(W_a \cdot \tanh(W_s s_{t-1} + W_h h)) \\ c_t = \sum_i a_{t,i} h_i \end{cases} \tag{3}$$

Here, a_t is the attention weight at time step t, h is the set of all hidden states from the encoder, and c_t is the context vector after weighted averaging, which depends on all past hidden states and their corresponding attention weights. W_a, W_s, and W_h are the weight matrices.

The optimizer uses gradient descent to adjust the network parameters, minimizing the difference between the predicted output and the actual output, thereby improving the model's learning efficiency and accuracy. The formulas are as follows:

$$\Theta_{new} = \Theta_{old} - \eta \cdot \nabla_{\Theta} J(\Theta) \tag{4}$$

Here, Θ represents the model parameters, η is the learning rate, $J(\Theta)$ is the loss function, and $\nabla_{\Theta} J(\Theta)$ represents the gradient of the loss function. This method is suitable not only for basic vocabulary learning but also for the development of advanced language skills, providing English learners with a comprehensive and in-depth learning pathway.

4 Methods for English Vocabulary Learning Based on Deep Learning

4.1 Word Vector Generation

The core of word vector generation methods is to learn the vector representations of vocabulary through training algorithms, which can capture the similarities and differences between words [6]. Word2Vec offers two main training architectures: Continuous Bag of Words (CBOW) and Skip-gram. CBOW predicts the target word based on the context, while Skip-gram predicts the context based on the target word. Both methods train the model through the following optimization objective:

$$\max imize \sum_{(w,c) \in D} \log p(c|w) \tag{5}$$

Here, D is the set of vocabulary and their contexts in the training dataset, w is the target word, and c is the context word, while $p(c|w)$ is defined through the softmax function as follows:

$$p(c|w) = \frac{\exp(\vec{v}_c^T \vec{v}_w)}{\sum_{c' \in C} \exp(\vec{v}_{c'}^T \vec{v}_w)} \tag{6}$$

Here, $\vec{v}_c$ and $\vec{v}_w$ are the word vectors for the context word and the target word, respectively, and C is the set of all context words. To further enhance the quality and applicability of word vectors, subword information (such as in the FastText method) can be introduced. This method considers the internal structure of words (such as prefixes, suffixes, and roots), enhancing the expressive capability of word vectors through the following approach:

$$\vec{v}_w = \sum_{s \in S_w} \vec{v}_s \tag{7}$$

Here, S_w is the set of subwords into which the word w is decomposed, and $\vec{v}_s$ is the vector representation of a subword. This representation not only captures the overall semantics of the vocabulary but also refines it to the constituent elements of the words, allowing the model to generalize better to new or low-frequency words.

4.2 Vocabulary Alignment and Classification

Vocabulary alignment typically uses a supervised learning model, with training data consisting of a series of word pairs, each labeled as synonyms or similar words. A classifier based on cosine similarity can be used, and the model can be optimized through the following loss function:

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \cos(\vec{v}_{wi}, \vec{v}_{ci}))^2 \tag{8}$$

Here, N is the number of training samples, y_i is the sample label (1 for synonym, 0 for non-synonym), $\vec{v}_{wi}$ and $\vec{v}_{ci}$ are the vector representations of the two words in the pair, and $\cos$ represents the cosine similarity between the vectors. The subsequent vocabulary classification step involves categorizing the aligned word vectors based on their semantic properties. In this process, clustering algorithms such as K-means can be used to divide the word vectors into different categories based on their distances. The specific clustering process can be represented as follows:

$$\min imizej = \sum_{j=1}^{k}\sum_{i \in S_i}\left\|\vec{v}_i - \vec{\mu}_j\right\|^2 \tag{9}$$

Here, k is the predetermined number of categories, S_j is the set of words assigned to the j-th cluster, $\vec{\mu}_j$ is the centroid vector of the j-th cluster, and $\|\cdot\|$ represents the Euclidean distance [7]. Through this method of vocabulary alignment and classification, large volumes of vocabulary data can be processed and utilized more accurately, providing high-quality input for deep learning models and achieving higher accuracy and efficiency in practical language applications.

4.3 Personalized Vocabulary Learning Optimization

Personalized vocabulary learning optimization customizes exclusive learning paths and content by analyzing learners' study history, preferences, and progress. The system first uses advanced data analysis techniques to assess the learner's vocabulary proficiency, identifying their weaknesses and strengths [8]. Then, based on this information, it adjusts the learning materials and difficulty levels, such as increasing the practice frequency in weaker areas or providing vocabulary content related to their interests to enhance the relevance and attractiveness of learning. A feedback loop is also employed, continuously optimizing the learning plan based on the learner's feedback and study results.

5 Experimental Validation

5.1 Experimental Data and Parameter Settings

To ensure the scientific validity and effectiveness of the experiment, a training set containing 10,000 words was selected, along with validation and test sets, each containing 2,000 words. These datasets ensure that the model can be effectively trained and evaluated on a large amount of vocabulary data. Additionally, a total of 140 learners participated in the experiment, with 100 in the training set, 20 in the validation set, and 20 in the test set, covering learners of different language levels and learning backgrounds, as shown in Figure 1.

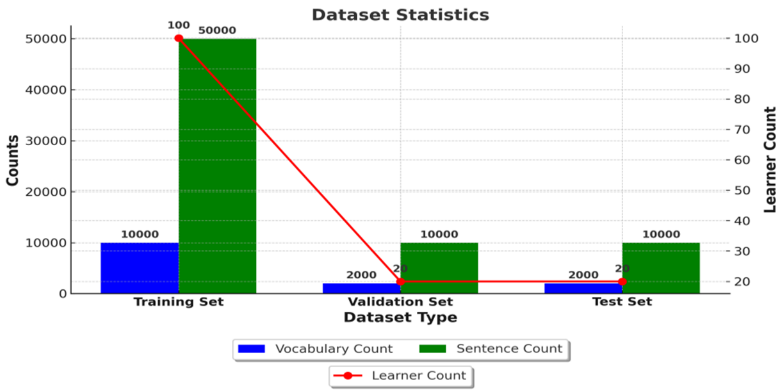

Fig. 1. Statistics of Experimental Datasets

During model training, an Adam optimizer with a learning rate of 0.01 was used, the batch size was set to 32, and the number of training epochs was 50. These parameters were adjusted through multiple trials to ensure that the model converges within a short time while maintaining high accuracy and stability [9]. The number of hidden layer units was set to 256, which showed good capability in capturing vocabulary features during the experiment, as shown in Table 2.

Table 2. Experimental Parameter Settings

Parameter	Value
Learning Rate	0.01
Batch Size	32
Training Epochs	50
Hidden Layer Units	256
Optimizer	Adam
Parameter	Value

To ensure the quality and consistency of the data, detailed data preprocessing and feature extraction were performed. During the data cleaning stage, noise data, duplicate words, and irrelevant content were removed to ensure the purity and representativeness of the data, as shown in Table 3. Subsequently, tokenization was performed to split sentences into individual words, and the Word2Vec model was used to generate word vector representations [10]. These word vectors were standardized to the same range during the feature normalization stage, ensuring convenience and consistency in subsequent processing.

Table 3. Data Preprocessing and Feature Extraction

Step	Description
Data Cleaning	Removing noise data, duplicate words, and irrelevant content
Tokenization	Splitting sentences into individual words
Word Vector Generation	Using the Word2Vec model to generate word vector representations
Feature Normalization	Standardizing word vectors to the same range for subsequent processing

5.2 Analysis of Vocabulary Learning Efficiency

In the deep learning-based English vocabulary learning method, detailed analysis of vocabulary learning efficiency is particularly important. A series of experiments were designed to validate the model's effectiveness, and the experimental results were thoroughly analyzed to ensure comprehensive and accurate analysis. Learners using the deep learning-based method were able to master an average of 25 new words per hour, while those using traditional memorization methods could only master 10 new words per hour. This indicates a significant advantage in vocabulary learning speed for the deep learning-based method. Additionally, the standard deviation showed that the deep learning method's effects were more stable among different learners, as shown in Table 4.

Table 4. Comparison of Vocabulary Mastery Speed between Different Learning Methods

Learning Method	Average Words Mastered per Hour	Standard Deviation
Traditional Memorization	10	2.1
Deep Learning-Based	25	3.5

Figure 2 shows the vocabulary mastery rates of learners at different levels using the deep learning-based method. Beginner learners achieved an 85% mastery rate for basic vocabulary and a 60% mastery rate for advanced vocabulary; intermediate learners achieved a 90% mastery rate for basic vocabulary and a 75% mastery rate for advanced vocabulary; advanced learners achieved a 95% mastery rate for basic vocabulary and an 85% mastery rate for advanced vocabulary. These data indicate that the deep learning method is highly effective for learners at all levels, especially in mastering advanced vocabulary.

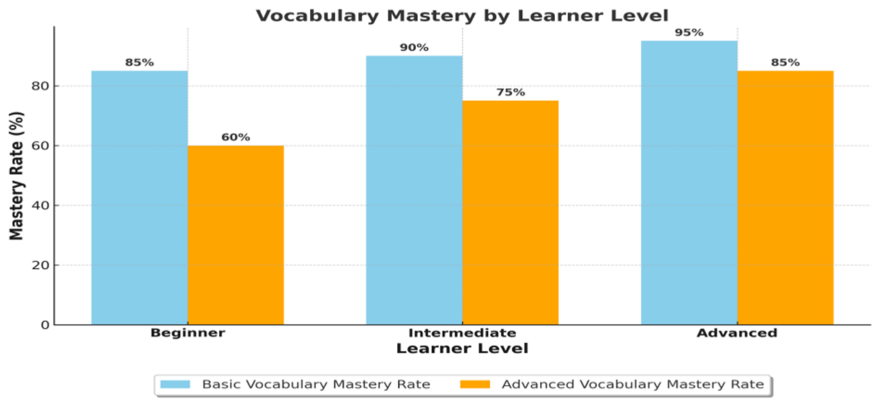

Fig. 2. Vocabulary Mastery Rates for Learners of Different Levels

Figure 3 shows that the vocabulary learning model achieved an accuracy and recall rate of 92% and 90%, respectively, on the training set, 89% and 87%, respectively, on the validation set, and 88% and 86%, respectively, on the test set. These data indicate that the model performs consistently across different datasets and has good generalization ability.

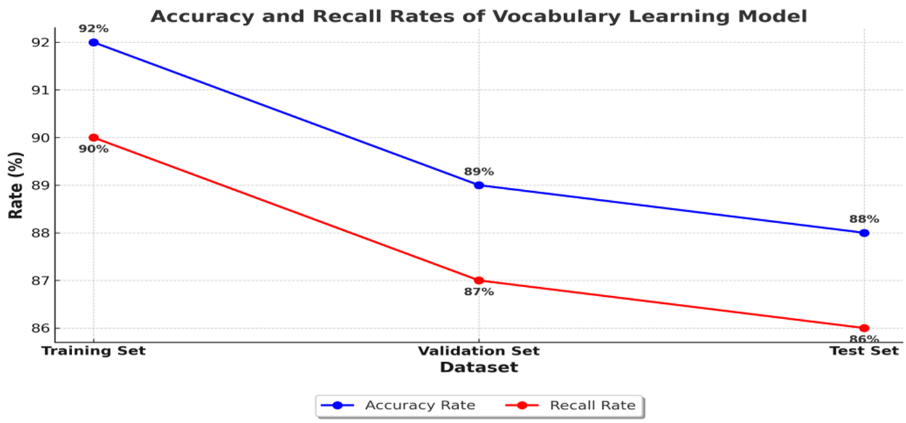

Fig. 3. Accuracy and Recall Rates of the Vocabulary Learning Model

The mastery rate for common vocabulary reached 95%, for specialized vocabulary 80%, and for low-frequency vocabulary 70%. These data show that the deep learning method performs excellently not only in learning common vocabulary but also significantly improves the mastery of specialized and low-frequency vocabulary, meeting various learning needs, as shown in Figure 4.

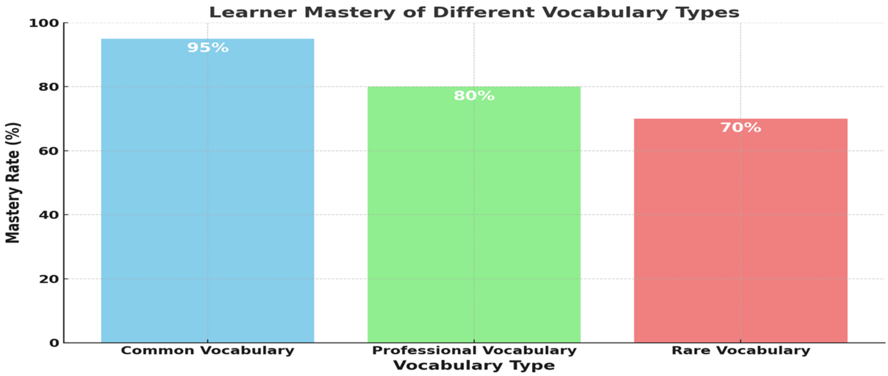

Fig. 4. Learners' Mastery of Different Types of Vocabulary

5.3 Evaluation of Learning Effectiveness for Different Vocabulary Categories

Evaluating the learning effectiveness of different vocabulary categories is a crucial step in verifying the method's effectiveness. By deeply analyzing the experimental data and assessing the learning outcomes of various vocabulary categories, the superiority of the deep learning model in practical applications can be further revealed. Learners at different levels showed significant improvement in mastering common vocabulary. Beginner learners' initial mastery rate was 50%, which increased to 85% after learning; intermediate learners improved from 60% to 90%; and advanced learners from 70% to 95%, as shown in Figure 5.

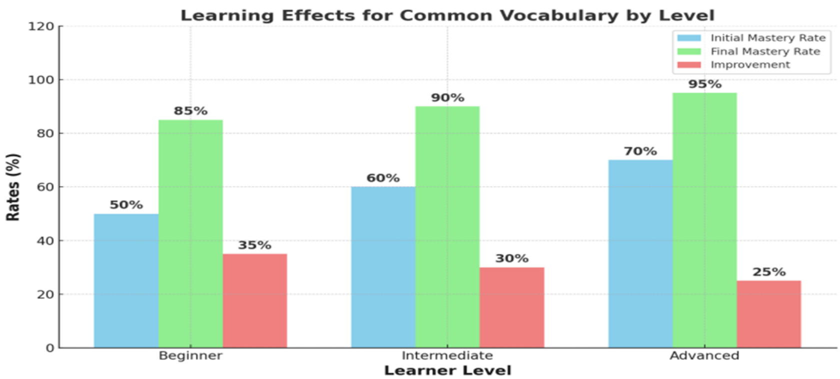

Fig. 5. Learning Effectiveness for Common Vocabulary

In terms of learning effectiveness for specialized vocabulary, as shown in Figure 6, learners also exhibited significant improvements. Beginner learners increased their mastery rate from 30% to 60%; intermediate learners from 40% to 75%; and advanced learners from 50% to 85%. This demonstrates the strong capability of the deep learning method in handling complex and specialized vocabulary.

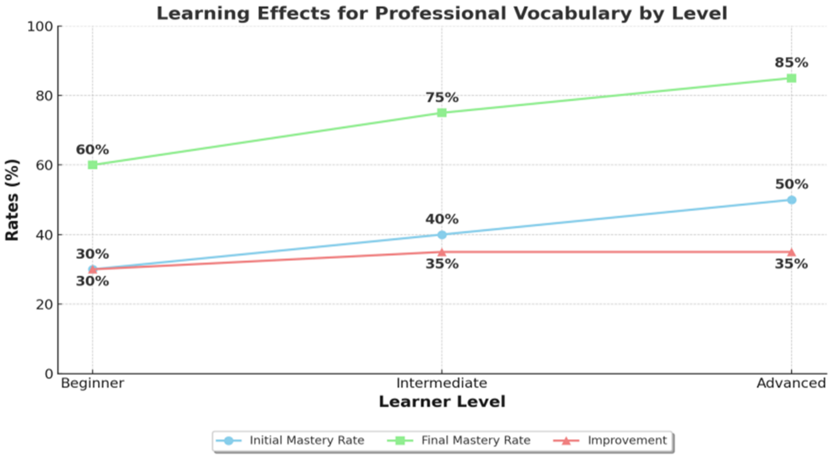

Fig. 6. Learning Effectiveness for Specialized Vocabulary

Figure 7 shows the learning effectiveness for low-frequency vocabulary. Beginner learners' initial mastery rate was 20%, which increased to 50%; intermediate learners improved from 25% to 65%; and advanced learners from 30% to 70%. Although low-frequency vocabulary is more challenging to master, the deep learning method still significantly improved learning outcomes.

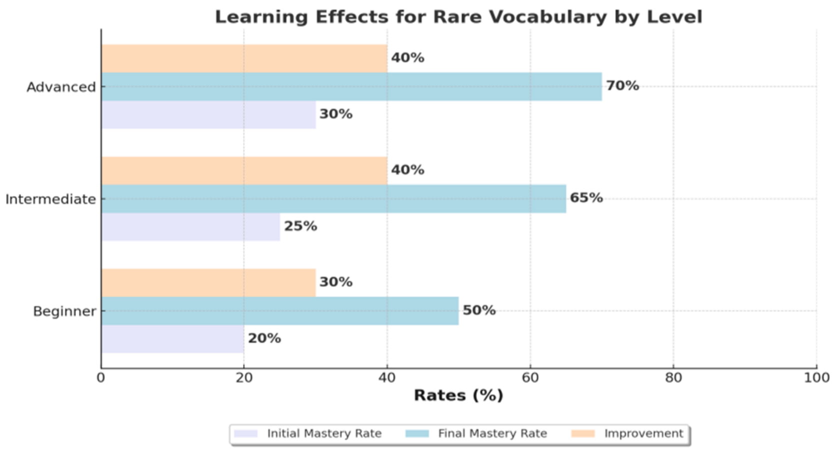

Fig. 7. Learning Effectiveness for Low-Frequency Vocabulary

Figure 8 summarizes the comparison of learning efficiency for different vocabulary categories. The average learning time for common vocabulary is 2.0 minutes per word, with an average improvement of 30%. For specialized vocabulary, the average learning time is 3.5 minutes per word, with an average improvement of 33%. For low-frequency vocabulary, the average learning time is 4.0 minutes per word, with an average improvement of 33%. These data indicate that although low-frequency and specialized vocabulary require more learning time, the deep learning method can still achieve significant learning outcomes across all types of vocabulary.

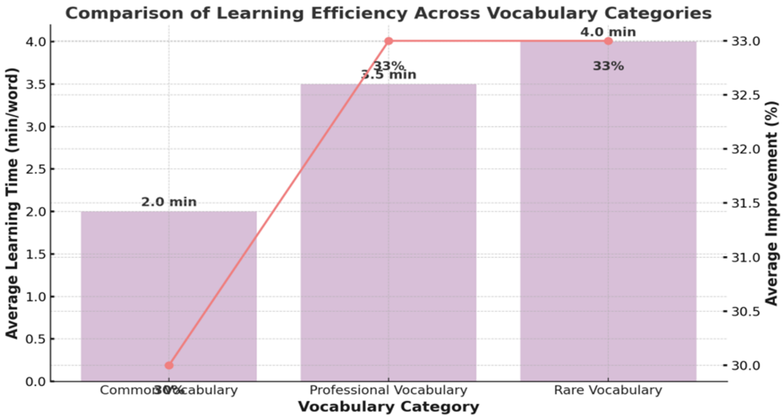

Fig. 8. Comparison of Learning Efficiency for Different Vocabulary Categories

6 Conclusion

The application of deep learning technology in English vocabulary learning significantly enhances learning efficiency and vocabulary mastery quality, especially in the learning of specialized and low-frequency vocabulary. In the future, further optimization of models and algorithms could achieve more refined personalized learning paths, driving profound changes in the field of language education.

References

1. Shuo F ,Lixia Y ,Fen L .Scoring method of English composition integrating deep learning in higher vocational colleges[J].Scientific Reports,2024,14(1):7287-7287.
2. Suri S S .Teachers' Strategies to Enhance Deeper Learning Skills in English Language Classes[J].International Journal of Linguistics, Literature and Translation, 2024, 7(3):118-125.
3. Qin M .Study of an Automatic Marking Algorithm for Subjective Questions in College English Exams Based on Deep Learning[J].Automatic Control and Computer Sciences, 2023, 57(6):638-645.

4. Tran Q T ,Nguyen T L V .Insights Into Vietnamese Bilingual Students' Quizlet-Based English Vocabulary Learning[J].Journal of Language Teaching and Research, 2023,14(3):682-688.
5. Xiuli L .Investigating the effectiveness of biofeedback techniques in improving English vocabulary learning in ESL learners[J].Journal of Commercial Biotechnology, 2023, 28(1):54-61.
6. Ling Y .Intelligent Recommendation System for English Vocabulary Learning – Based on Crowdsensing[J].Applied Mathematics and Nonlinear Sciences,2022,8(2):885-896.
7. Fan L, Si F, Yanjun W , et al. Chinese University Students' Experience of WeChat-Based English-Language Vocabulary Learning[J].Education Sciences,2021,11(9):554-554.
8. Yujuan C .THE IMPACT OF NEGATIVE TRANSFER OF MOTHER TONGUE ON JUNIOR MIDDLE SCHOOL ENGLISH VOCABULARY LEARNING[J]. Socio-Technical Dimensions For Future Vision,2020,2(2):104-108.
9. Jiang Y. Application of the Mind Map in Learning English Vocabulary[J]. Open Access Library Journal, 2020, 07(06):1-4.
10. Ajisoko P .The Use of Duolingo Apps to Improve English Vocabulary Learning [J]. International Journal of Emerging Technologies in Learning (iJET), 2020, 15(07):149-155.