International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-8

https://doi.org/10.70917/ijcisim-2026-0379
Article

Design and Practice of English Learning Support Systems in Intelligent Housing Environments: Exploring the Integration of Technology and Education

Peilan Peng [1,*]

[1] ZHIYUAN (Liberal Study) College, Beijing Institute of Petrochemical Technology University, Beijing, 102617, China
* Correspondence author: lansunny1115@126.com

Abstract: In order to enhance the effectiveness of classroom teaching, this paper designs and builds a set of English learning support system by combining an intelligent housing environment based on multimedia technology with English teaching. The system architecture mainly includes four functional layers. The data layer relies on the collaboration of multiple servers to realize the transmission and preservation of data. The algorithm layer adopts deep learning algorithms to complete semantic understanding and classification of utterances. The intermediate logic layer accomplishes speech recognition and intelligent dialog through the AI module, the script module is used to process the business logic in the speech data, and the image module recognizes and simulates residential scenes, characters, actions and expressions. The performance layer is equivalent to the client, supporting operations such as system login, system settings and logging out. The experimental results show that the system takes less than 30s to respond to user requests such as questions, corrections, pre-class pre-study and post-class consolidation, and the accuracy of the system in intent recognition and slot value classification of English semantics is more than 95%. This shows that the system's stable performance and high semantic recognition accuracy in the intelligent housing environment can effectively improve students' English learning efficiency in daily life scenarios.

Keywords: Multimedia Technology; Intelligent Housing; English Teaching; Deep Learning; Living Scene

1. Introduction

Intelligent remodeling of the housing environment is essentially a systematic upgrade of the housing environment through multimedia technology based on scientific discourse and design thinking [1]. During the transformation process, a variety of intelligent devices such as networks, computers, cameras, smoke alarms, sensor lighting systems, temperature and humidity sensing systems, projectors, speakers, etc., are set up in the housing environment [2]. This kind of renovation section improves the comfort of housing through remote control, but also simultaneously strengthens the security of housing and improves the spatial performance of housing [3].

With the accelerated development of economic globalization, English as an international common language has become an important bridge between countries, which also promotes the community to increase the degree of attention to English learning [4]. The traditional English classroom relies on the blackboard for writing and explaining, but with the popularization of the Internet as well as multimedia technology, the English teaching paradigm has begun to gradually shift to multimedia technology-supported situational teaching. This paradigm can not only stimulate students' enthusiasm for English learning, but also improve their motivation for learning [5]. English learning support system is a system solution to empower education with technology, providing personalized and efficient English learning support for learners of different ages and levels [6].

Based on this, this study designs an English learning support system based on an intelligent housing environment, covering students' force training, speaking practice, reading ability development, vocabulary consolidation, systematic learning of grammar, comprehensive English assessment, completion of after-school homework, and study plan development. The system not only supports students' daily independent practice, but also realizes scenario teaching through intelligent housing environment, so that students can deepen their English application comprehension in highly realistic virtual scenarios. The proposed research breaks through the physical boundaries of traditional learning scenarios to create a comfortable, scientific and immersive learning space for students. It not only realizes the in-depth exploration of teaching resources and improves teaching efficiency, but also provides a practical paradigm for housing environment design.

2. Intelligent Housing Environment in English Language Learning in Practice

Intelligent housing environments are effective across the pre-, in-, and post-course aspects of English language learning through the integration of multimedia technologies, learning systems, and personalized instructional strategies.

2.1. Pre-learning

Setting multimedia devices in intelligent housing environment can improve students' pre-study before English class [7]. Teachers combine the psychological growth of students to produce short videos, which are distributed to students through the learning system. The short videos cover the main content of the course, promote students' understanding of the course through the form of breaking through small games, and raise some questions that can be found out the answers in the text, guiding the students to read the English text with the questions. The English learning system can supervise students' pre-study, and teachers can check students' pre-study through the background of the system, so as to improve the effect of after-class pre-study of English as well as the effect of classroom teaching, and to help students cultivate the good habit of pre-study before class.

2.2. Classroom instruction

In classroom teaching, for the semantics of English and the use of the environment, the use of multimedia technology for the scenario display, can strengthen the students' emotions, so that students can be immersed in the social background of the text, the relationship between the characters and the characters' emotional expression [8]. For example, “Developing ideas” text mainly expresses the tragic life of the Jews in the turbulent times of World War II, so in the teaching of the text, teachers can use multimedia projection equipment to play the documentary of the Jewish life in World War II to students, to promote a deeper understanding of the text, and improve the quality of teaching while improving students' learning. Improve the quality of teaching and learning at the same time to improve student learning enthusiasm.

2.3. Consolidation at the end of the lesson

The mastery of learning content is mainly based on classroom learning, followed by consolidation after class. With the development of multimedia technology, learning systems can be used to develop online learning programs and after-school testing for students, teachers in the background of the system to view the progress of student testing, as well as the mastery of knowledge, through the learning system to exercise the students' ability to listen, speak, read and write. You can also set up some rankings of the time and quality of completing tasks within the class, so that the completion of after-school tasks into a confrontational, fun game, motivate students to participate more actively, and turn the boring after-school homework into a knowledge competition between students to catch up with each other and help each other.

3. English Learning Support System Design

Intelligent housing realizes the intelligence of the English teaching process through a variety of multimedia devices such as networks, computers, cameras, sensor lighting systems, temperature and humidity sensing systems, projectors, speakers and so on [9]. At the same time, these all devices can also be used to collect the voice files of the teacher's lectures and students' oral practice in the English classroom, which can be used as the raw data of the English learning support system, and then provide a data carrier for the optimization of the design [10].

3.1. General structure

Virtual technology and deep learning are combined in intelligent housing to construct an English learning support system [11]. The system covers the data layer, algorithm layer, intermediate logic layer, and performance layer, and completes the oral training process through relevant equipment, and the overall structure is shown in Figure 1. The data layer accomplishes data transmission and storage through the collaboration of data server, synchronization server and file server. The algorithm layer uses deep learning algorithms to complete the semantic understanding and recognition of utterances, the intermediate logic layer completes speech recognition and intelligent dialog through the AI module, and the script module performs business logic processing on the speech data in order to obtain the logic data. The image module is used to recognize and simulate scenes, characters, actions, and expressions, and the three modules collaborate to obtain the corresponding business rules. The performance layer is equivalent to the client, where the user performs operations such as system login, system settings, and logging out.

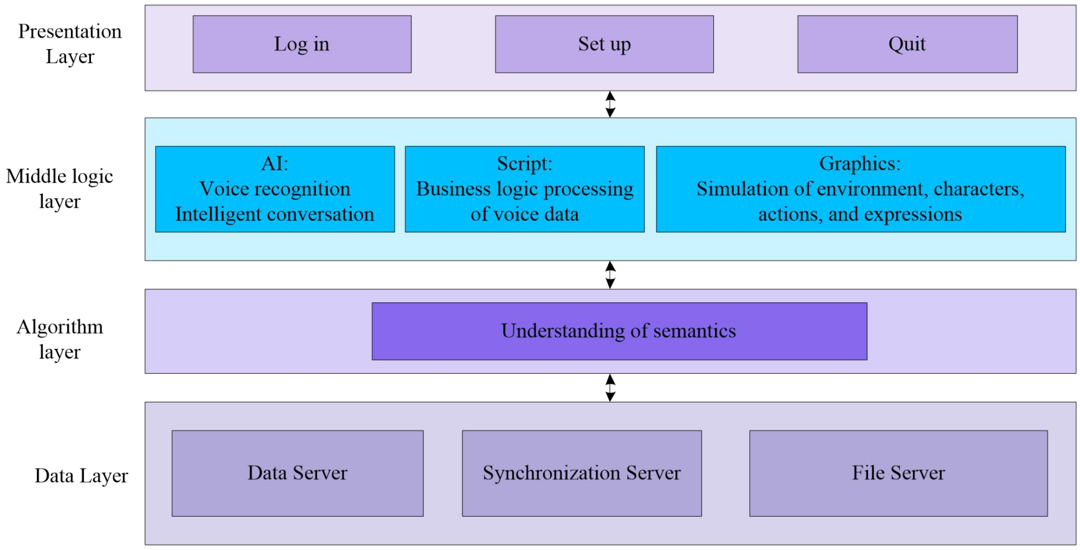

Figure 1. System overall structure

3.2. Semantic understanding based on role information

In the design of English learning support system, the main deep learning algorithms to understand the semantics of English, learning algorithms are usually used in the Transformer-based pre-training model (BERT) to represent the language. The BERT model is not only simple in structure, but also through the self-attention mechanism and the parallel computing ability to enhance the processing of the language. Recurrent neural network (RNN) due to the lack of memory units makes the network gradient existence of the disappearance of the situation, in order to solve the problem of the RNN to increase the memory units, the evolution of the long short memory neural network (LSTM), the LSTM network can be the global data collection and storage. Since the LSTM model has ideal modeling ability for textual temporal information, in order to better extract and understand the English semantics, this paper integrates the BERT model with the LSTM model to construct BiLSTM neural network with the support of intelligent housing environment, to realize the deep understanding of English semantics [12].

3.2.1. Semantic Understanding Model Construction

In order to improve the understanding of English semantics by the English learning support system, a semantic understanding model is constructed on the basis of BiLSTM neural network, and the structure of the semantic understanding model is shown in Fig. 2. The semantic understanding model is divided into semantic information extraction stage, history influence vector extraction stage and semantic understanding stage combined with history influence vector. The semantic information extraction stage mainly focuses on extracting the feature information of English utterances, the historical influence vector extraction stage extracts the historical language of the characters and the influence weights of the historical language, and the semantic understanding stage combined with historical influence vectors predicts and analyzes the real intention of the historical language of the objects, so as to realize the semantic comprehension of the English utterances [13].

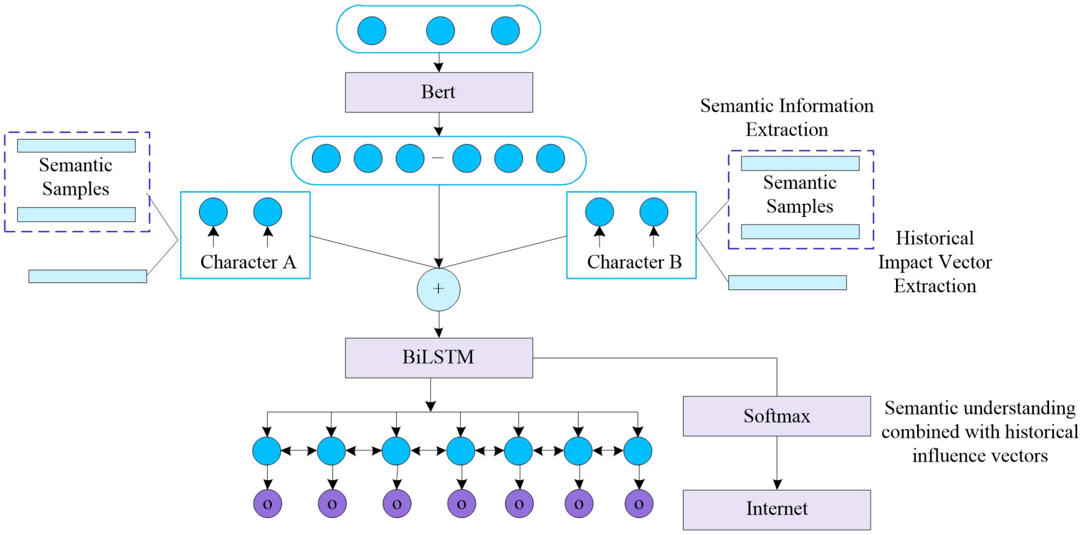

Figure 2. Semantic understanding model structure

3.2.2. Semantic Information Extraction

The premise of English semantic understanding is semantic feature extraction, in this paper, BERT pre-training model is used to extract semantic features. First of all, the BERT pre-training model is trained on semantic text in a fine-tuned way to get the loss value of the BERT pre-training model, and then in using the computation as well as the reverse propagation backpropagation process of the network, and finally the semantic information features are extracted.

If $L_0 = \{(S_i, y_i^i, y_i^s)\}$ is the input to the BERT pre-training model and satisfies $i \in [1, n]$, then where S_i and n are the English utterance samples as well as the number of samples, respectively, and y_i^i and y_i^s are the destination and slot-implanted labels of the utterance samples i, respectively. After the utterance sample of the segmentation processing input model utterance sample can be represented as $S_i = (c_1, c_2, c_3, \cdots, c_T)$, at this time, this utterance sample will be input into the BERT pre-training model, and the embedding vectors of each vocabulary word will be obtained, the At this point the semantic features of the utterance are $S = [h_1, h_2, \cdots, h_t]$.

3.2.3. Historical impact vector extraction

Assuming that the semantics of the utterance currently being input is S_{cur}, the number of historical utterance samples to be encoded is set to n, and after the input of the utterance samples is complete, a historical utterance semantic matrix is constructed $S_{bef} = [S_1, S_2, \cdots, S_n]$. If the input utterance semantic S_{cur} is fused and processed with the historical utterance semantic matrix, a new historical utterance semantic matrix can be obtained $S_{bef}^{'} = [S_1^{'}, S_2^{'}, \cdots, S_n^{'}]$ Accordingly, the influence weight of the historical utterance on the current input utterance is obtained, and the semantic vector after the fusion processing of the two utterances is calculated as:

$$S_t^{'} = S_{cur} + S_t \tag{1}$$

Where t denotes a certain historical utterance, and the semantic vectors before and after the fusion of this historical utterance are S_t and $S_t^{'}$, respectively, and the influence weights of the t historical utterance are calculated using Softmax. Then, the calculation results are input into the new semantic matrix of historical utterances $S_{bef}^{'}$ to obtain the influence weights of multiple utterances, Eq:

$$a_t = Softmax\left(tanh\left(W_{att}^T \times S_t + b_{att}\right)\right) \tag{2}$$

Where $tanh(\cdot)$ is the activation function, W_{att}^T is the hidden layer weight matrix, and b_{att} is the bias term. If the new semantic matrix of historical utterances is input into the BERT pre-training model, the weight distribution of historical utterances is obtained, thus visualizing the logical relationship between historical utterances and with the current utterance.

In the process of English teaching students' utterance dialogues present multiple rounds, the semantic comprehension model needs to identify the character roles, so it is necessary to combine the role information with the historical utterance weight information, and under the calculation of BiLSTM neural network, to get out the influence vector of the historical utterances, the calculation formula is:

$$a_t = Softmax\left(tanh\left(W_{att}^T \times S_t + b_{att}\right)\right) \tag{3}$$

where $BiLSTM_a$ and $BiLSTM_b$ denote the BiLSTM neural networks for the speaker's a and b roles, and $S_{t,role\ a}$ and $S_{t,role\ b}$ denote the historical utterance samples of speaker a and b, and $\alpha_{t,role\ a}$ and $\beta_{t,role\ b}$ denote the historical impact vectors of speaker a and b roles.

3.2.4. Semantic Understanding Combining Historical Influence Vectors

In order to improve the accuracy of English semantic comprehension, the current utterance is purpose-identified and slot-planted filled by BiLSTM neural network in the context of intelligent environment. The influence vector of historical utterance role information V_{bef} is used as the initial value of BiLSTM neural network, and $S_{cur} = [S_1, S_2, \cdots, S_n]$ is used as the input of the model to get the all hidden layers' State.

To achieve semantic understanding combining historical influence vectors through intent classification, slot value classification, and calculation of loss functions for slot filling and intent recognition, the following three processes need to be analyzed:

(1) In the semantic intent classification process, the vector obtained by fusing the output vectors of the hidden layers of the BiLSTM neural network is computed with the formula:

$$V_{cur} = BiLSTM\left(S_{cur}, V_{bef} \times W_{bef}\right) \tag{4}$$

where S_{cur} denotes the semantics of the input utterance and W_{bef} denotes the weight matrix of the historical utterance. The Intentional Classification Prediction is computed as:

$$Intent = Softmax\left(V_{cur} \times W_{Intent}\right) \tag{5}$$

where W_{Intent} is the historical utterance weight matrix.

(2) Conditional Random Field (CRF) is mainly used in the process of slot-value classification, which is based on logical relations to judge the input utterances positively or incorrectly, so as to obtain the best recognition results of the utterances. Therefore, the CRF model score is calculated as:

$$Score\left(l|s\right) = \sum_{j=1}^{m}\sum_{i=1}^{n}\lambda_j f_j\left(s, i, l_i, l_{i-1}\right) \tag{6}$$

where s and l are the input utterance and lexicality, respectively, l_i and l_{i-1} are the labeled lexicality of i and its next word, respectively, and $f_j\left(s, i, l_i, l_{i-1}\right)$ denotes the the feature function of the input English utterance, λ_j is the weight coefficient.

(3) The utterances are fused to classify the input utterances by slot planting through the CRF model, which is calculated by the formula:

$$Slot\left(V_i\right) = CRF\left(Softmax\left(h_i\right)\right) \tag{7}$$

Where h_i is the hidden layer output vector of i words in the LSTM model. If an English utterance contains n words or m feature vectors, the sum of the feature vectors of multiple vocabularies is the rating of the input utterance. By exponentiating the ratings, the labeling probability value of the current utterance is obtained, which is proportional to the confidence of the labeling result, and is calculated as:

$$p\left(l|s\right)=\frac{exp\left[score\left(l|s\right)\right]}{\sum_{l'}exp\left[score\left(l|s\right)\right]}=\frac{exp\left[\sum_{j=1}^{m}\sum_{i=1}^{n}\lambda_j f_j\left(s,i,l_i,l_{i-1}\right)\right]}{\sum_{l'}exp\left[\sum_{j=1}^{m}\sum_{i=1}^{n}\lambda_j f_j\left(s,i,l_i,l_{i-1}\right)\right]} \tag{8}$$

Based on the above, the whole semantic comprehension model operates collaboratively in the intelligent housing environment relying on computers, speech recognition and other devices to collect and process speech input, learning trajectories, etc. by themselves. Students can not only better understand the semantic intent of the current English expression, but also significantly improve the adaptation level of intelligent English teaching.

4. System application analysis

4.1. System performance

In order to verify the response performance of the English learning support system designed in this paper to client requests in an intelligent housing environment, random forest algorithms, machine learning algorithms, unsupervised learning algorithms, and deep learning algorithms proposed in this paper are selected to construct a variety of English learning support systems for comparative implementation. Each system issues questions and makes post-test modifications to students' tests through the multimedia learning terminals in the intelligent housing, and at the same time inquires about students' pre-course prep work and post-course consolidation. Therefore, the experiment focuses on evaluating the response time of different algorithms to user requests in an intelligent scenario, and the performance of different English learning support systems is shown in Table 1.

It can be seen that the English learning support systems under the Random Forest algorithm and the unsupervised learning algorithm take a long time to respond to the statistical requests for questions, corrections, and pre- and post-course consolidation on the user's end. The main reason is that the high complexity of the Random Forest model leads to a large number of prediction samples and a long computation time, which further lengthens the response time of client requests. The unsupervised learning algorithm data preprocessing process is cumbersome, resulting in increased computation and slowing down the request processing speed. In contrast, the English learning support system based on deep learning algorithms proposed in this paper, combined with intelligent housing, takes less than 30s to respond to statistical requests for question generation, correction, and pre- and post-course pre- and post-consolidation, which is significantly better than other algorithms. This indicates that the English learning support system built by combining intelligent housing and deep learning can provide students with a more convenient and intelligent learning experience.

Table 1. Performance of different English learning support systems

Response time/s	**English learning support system**			
	Random Forest Algorithm	**Machine Learning Algorithms**	**Unsupervised learning algorithms**	**Deep Learning Algorithms**
Test questions	104.57	74.38	100.25	28.36
Correction of test questions	98.83	71.65	92.79	29.98
Statistics before class	100.03	65.39	90.37	24.74
After-class consolidation statistics	99.56	75.17	96.44	26.83

4.2. Semantic categorization

In this paper, 380 semantic samples are randomly selected from the English utterance database, and these semantic samples are inputted into the English learning support system, and the Random Forest algorithm, machine learning algorithm, unsupervised learning algorithm, and deep learning algorithm in this paper are used to classify the intention and slot-plant classification in the input English utterances, respectively, and the semantic classification accuracy rate of each model is shown in Fig. 3. In the students' English speaking practice, only the English learning support system based on the deep learning algorithm has the highest accuracy of intent classification and slot-plant classification in semantics, with an accuracy of 95.1% for intent classification and 96.3% for slot-plant classification. It indicates that the deep learning algorithm combined with intelligent housing shows better results in English semantic recognition and can support personalized teaching and feedback more effectively.

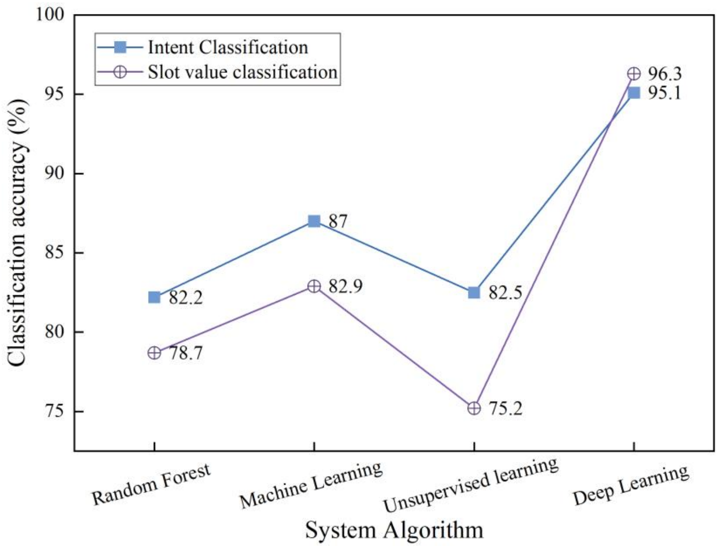

Figure 3. Semantic classification accuracy of each model

5. Conclusion

Intelligent housing mainly refers to the transformation of multimedia technology to improve the intelligence and comfort of the house, students can learn English in the intelligent housing environment, through the assistance of multimedia equipment, to improve the interest and enthusiasm for English learning, and under the guidance of multimedia video, in-depth understanding of the social background of the English text, the relationship between the characters, and the characters of the expression of emotions. At the same time, combined with deep learning algorithms, the English learning support system can summarize and analyze students' pre-study, post-study consolidation, and knowledge mastery to help teachers better understand students' weaknesses in English learning, realize accurate Q&A in class, and improve teaching effectiveness. The analysis results show that the English learning support system based on deep learning algorithms in this paper can respond to all client requests within 30s, and the system performance has obvious advantages. In the classification of English semantics, the English learning support system in this paper has the highest accuracy in classification, and the intent classification and slot plant classification of semantics are above 95%, which has high accuracy in semantic understanding. Combined with real-time speech acquisition and personalized learning in intelligent housing, it helps the system understand students' language input more accurately.

Funding

2025 Project of Beijing Institute of Petrochemical Technology: Project of Innovation and Entrepreneurship Training Program for College Students; Graduate Education and Teaching Reform Research and Practice--Curriculum Construction and Teaching Reform (Project No. : YJ25-204).

Funded by 2025 University-level Education and Teaching Reform Project: Innovation and Practice of Blended English Teaching Model for Non-English Major Postgraduates Based on the Production-Oriented Approach.
Project No.: YJ25-204

References

1. Wang, Y., Ye, H., Xiong, J., Nie, Y., Jiang, L., & Zhang, A. (2024). Digitization impact on future housing building industry mode. *Journal of Building Engineering,* 110202.
2. Al‐Raeei, M. (2025). The smart future for sustainable development: Artificial intelligence solutions for sustainable urbanization. *Sustainable development*, 33(1), 508-517.
3. Yang, B., Lv, Z., & Wang, F. (2022). Digital twins for intelligent green buildings. *Buildings,* 12(6), 856.
4. Agustin, W., Wahyudin, A. Y., & Isnaini, S. (2021). Language learning strategies and academic achievement of English Department students. *J. Arts Educ,* 1(1), 19-29.
5. Wang, Y. (2022). Research on English situational teaching assisted by multimedia and network. *Advances in Multimedia,* 2022(1), 8142553.

6. Lee, D., Kim, H. H., & Sung, S. H. (2023). Development research on an AI English learning support system to facilitate learner-generated-context-based learning. *Educational technology research and development,* 71(2), 629-666.
7. Li, Y. (2024). Construction of Smart Classroom for English Teaching under Modern Information Technology. *International Journal of New Developments in Education,* 6(12).
8. Dubovi, I. (2022). Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology. *Computers & Education,* 183, 104495.
9. Zhou, Y. (2022). Construction and application of college English multiple intelligence teaching model based on internet of things. *Mathematical Problems in Engineering,* 2022(1), 5014131.
10. Liu, L., & Tsai, S. B. (2021). Intelligent recognition and teaching of English fuzzy texts based on fuzzy computing and big data. *Wireless Communications and Mobile Computing,* 2021(1), 1170622.
11. Liu, X., Zhang, X., & Baziar, A. (2023). Hybrid machine learning and modified teaching learning-based english optimization algorithm for smart city communication. *Sustainability,* 15(15), 11535.
12. Xiang, Y., Chen, Y., Fan, W., & Ye, H. (2024). Enhancing computer-aided translation system with BiLSTM and convolutional neural network using a knowledge graph approach. *The Journal of Supercomputing, 8*0(5), 5847-5869.
13. Gao, E., Guo, J., Pang, X., Bo, D., & Chen, Z. (2024). Exploring pathways to comprehension performance in multilanguage smart voice systems: insights from Lasso regression, SEM, PLS-SEM, CNN, and BiLSTM. *Humanities and Social Sciences Communications,* 11(1), 1-20.