

## Research Article

# Development Countermeasures of College English Education Based on Deep Learning and Artificial Intelligence

Fei Wu,<sup>1</sup> Yu Chen,<sup>1</sup> and Dan Han <sup>2</sup>

<sup>1</sup>Department of General Education, LuXun Academy of Fine Arts, Shenyang 110003, China

<sup>2</sup>Department of World Language, Dalian University of Foreign Languages, Dalian 116044, China

Correspondence should be addressed to Dan Han; handan@dlufl.edu.cn

Received 25 April 2022; Revised 7 June 2022; Accepted 13 June 2022; Published 22 June 2022

Academic Editor: M. Praveen Kumar Reddy

Copyright © 2022 Fei Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

College English education aims at cultivating students' English application ability and promoting the development of students' English communication level. However, at present, many college students still stay in English learning to cope with the examination. The cultivation of oral English and other comprehensive ability is little. Therefore, the development direction of college English education should be based on the actual needs of students in social work and life, and formulate a plan more suitable for contemporary development needs. In this context, to better provide countermeasures for the development of college English education, this paper provides a new English education model based on artificial intelligence (AI), and evaluates students' comprehensive English ability through deep learning (DL). This paper not only constructs a college English teaching model and evaluation system through artificial intelligence methods but also evaluates the quality of English pronunciation through DBN. It also designs comparative experiments, and the consistency between manual evaluation and DBN-based evaluation is obtained. The results show that the English pronunciation evaluation system based on DBN has a high consistency with manual evaluation. Among them, in the speech rate evaluation, the adjacent consistency rate reaches 99.58%, showing that the evaluation model constructed in this paper is effective and verifying the feasibility of applying the artificial intelligence method to college English education.

## 1. Introduction

With today's increasingly obvious globalization, the importance of English as an international language is self-evident. As a professional learning platform, universities have never stopped English education. But in terms of the current English education environment, the center of English education is on the examination, such as the CET examination and other English examinations, which have no obvious effect on the cultivation of students' comprehensive abilities such as oral English. Furthermore, due to the great difference between English language features and Chinese language features, it is more difficult for students who are accustomed to Chinese language features to learn English. Therefore, how to overcome the problems arising in the learning of different language systems has become the focus of language scholars' research, which also involves the

consideration of the teaching development direction of college English education. The development of society has led to the development of various technologies, and these new technologies have also injected new impetus into the reform of college English teaching mode. Coupled with the advent of neural networks in recent years, there are new ways to review English. In the traditional evaluation, teachers are mainly based on classroom tests and examinations to grade students' conditions. This result evaluation method is difficult to analyze knowledge mastery according to students' personal conditions, which is not conducive to students' English learning.

In such context, related research for college English teaching of college English education development countermeasures can provide a certain theoretical basis for the AI and DBN methods commonly used in college English education development countermeasures study. On the one

hand, the study of AI and DBN role field has been extended. On the other hand, it also makes the article with more modern features, which has certain theoretical and practical significance. It also uses new technology to explore a new model for the development of college English education, which can provide theoretical support for the choice of strategies for the development of college English education.

This paper will construct the college English teaching process through AI technology and evaluate as well as analyze students' spoken English through deep learning methods. The results confirm that the oral English evaluation method based on DBN has a positive effect. The innovation of this paper is to construct a model of college English education through artificial intelligence, analyzing the pronunciation learning in college English through DBN, as well as applying the advanced technology to the mode reform and the construction of a speech evaluation system. In the pronunciation evaluation experiment, a test plan is also designed to analyze the feasibility of the constructed model.

## 2. Related Work

College English education, as an English language education in higher education, has attracted many research scholars to study it. Wen et al. made statistics of vocabulary in college English textbooks of different countries and built a research system of standard English vocabularies on this basis [1]. Li made a theoretical modeling study on the characteristics and rules of college English learning activities through cognitive strategies in educational principles, and established a statistical analysis model based on theoretical knowledge to study the effect of oral English teaching. He also analyzed the results of which evaluation methods were used to improve oral English learning [2]. Gong believed that there were too many writing forms in college English teaching. In the actual teaching process, teaching methods should be reformed, from paying more attention to writing form to the development of practical English ability [3]. Starting from the development of modern technology, Nie put forward the argument that the development of college English education can be based on modern technology [4]. Liu, on the other hand, studied the learning mode of MOOCs and believed that the MOOCs system had injected new impetus into the development of English education in China [5]. Hao and Su made a theoretical analysis of the teaching method of the "flipped classroom". He conducted an empirical study on the application of this teaching method to college English classes through simulation experiments [6]. Zhang used the Bayesian network method to study English grammar learning [7]. Lin considered that mind map learning strategies could help students organize their thinking and improve their English writing ability [8]. From the existing research, there are few data analyses on the effects of college English education under the combined mode. Therefore, the research on development strategies from the perspective of big data has become the direction considered by researchers.

At present, algorithm analysis based on deep learning (DL) has been glowing in many fields, among which the data analysis effect of deep neural network (DNN) has also been confirmed by researchers in many fields. Ravanelli et al. modeled speech classification recognition based on a gated cyclic unit neural network and studied the automatic recognition function of speech based on this [9].

Yazdani et al. used DNN to evaluate and parallelize automatic recognition of speech [10]. Pakoci et al. studied the recognition of speech by building an acoustic model using a deep neural network trained on sequence decision [11]. Bo et al. performed remote speech processing via deep neural networks [12]. These related studies all reflect the feasibility of neural networks in speech recognition. AI, as a popular subject for the joint development of many disciplines such as computer science, is also a research hotspot in many disciplines. Glauner et al. solved the problem of detecting nontechnical losses including electricity theft, faulty meters, or billing errors in electrical engineering through artificial intelligence [13]. Havinga et al. studied the use of artificial intelligence for the collaborative detection of events by sensor nodes in wireless sensor networks [14]. Caviglione et al. applied artificial intelligence to the development of network information security engineering to detect modern malware [15]. Artificial intelligence algorithms are currently widely used in science and engineering but few in liberal arts. In actual research, using a single method is easy to deviate from the final result, and it is often necessary to synthesize the algorithm based on the research data.

## 3. College English Teaching Mode and Speech Recognition Model Method Based on Deep Learning and Artificial Intelligence

### 3.1. Artificial Intelligence and Deep Learning

*3.1.1. Artificial Intelligence.* AI, which is part of computer science, is a new technological science that responds in a similar way to human intelligence [16]. Artificial intelligence encompasses knowledge in multiple disciplines. As shown in Figure 1.

Figure 1 shows some of the fields included in AI, including mathematics, philosophy, computer science, psychology, and cybernetics, as well as cognitive science, information theory, and other disciplines, which involve many disciplines. It can be seen from this that AI technology is developed from multi-disciplinary theoretical knowledge and time manipulation [17, 18]. Under the joint action of multi-disciplinary fields, the application fields of AI are also wide. As shown in Figure 2.

Figure 2 shows the common fields where AI technology is applied. The fields shown in the figure include neural networks, machine learning, complex systems, intelligent search, natural language processing, and pattern recognition. In addition, AI can also be applied to combined scheduling, inference planning, etc. AI technology has been recognized by scholars in many fields due to its advantages of multi-disciplinary integration. Using AI technology in big

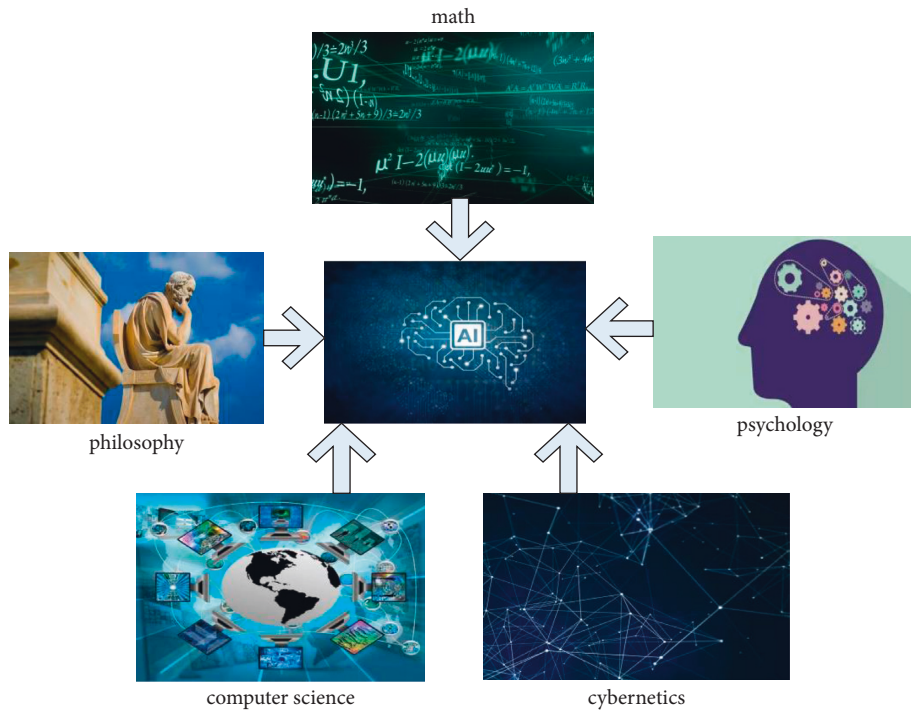


FIGURE 1: Areas included in AI.

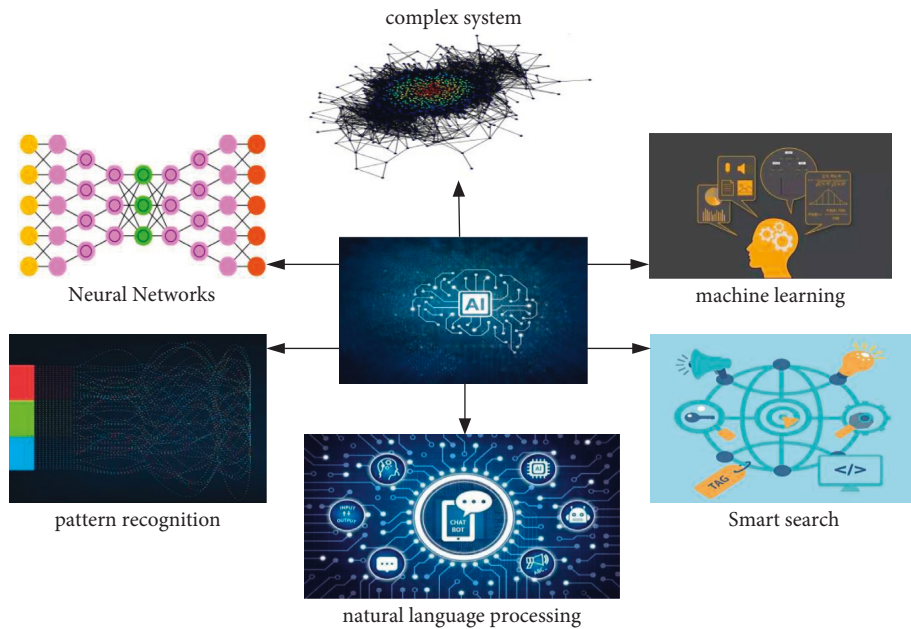


FIGURE 2: AI application areas.

data analysis can save a lot of time and obtain similar data results. In the future, AI technology will also conduct more explorations and attempts in multiple subdisciplines.

**3.1.2. Deep Learning.** Deep learning is a technical approach to the algorithmic analysis of data through a neural network model. The training principle of deep learning is to constantly update the parameters in the training and gradually transform the analyzed data features from low to high through layer-to-

layer operation. It will classify and output features finally. The simple process of deep learning is shown in Figure 3.

Figure 3 is a simple flowchart of deep learning. As can be seen from the simple flowchart in Figure 3, DL needs repeated training in the process, and the loss function is also required to participate in data training analysis to ensure the minimum error of data analysis results. During this process, weight and other training parameters are constantly updated. Then it finally inputs the final result.

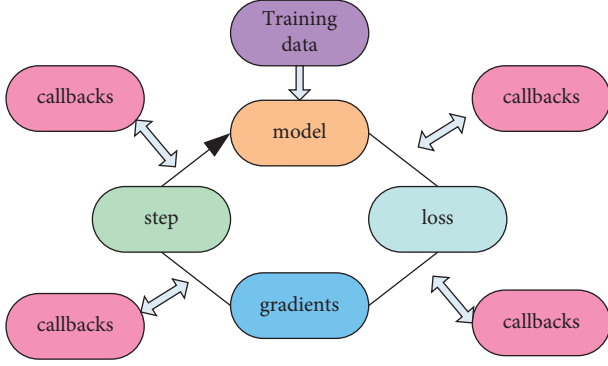


FIGURE 3: Deep learning flowchart.

DNN is developed from the traditional neural network. The traditional network model is a single-layer nonlinear mapping. DNN adds more hidden layer nodes based on the traditional neural network. The traditional neural network and DNN models are shown in Figure 4.

As shown in Figure 4, the network structure of DNN is similar to that of the traditional neural network, except that the hidden layers of the two models are different. In the deep neural network structure, the increased hidden layer nodes can help the dataset continuously update the weights during the training process so that the final training result will be closer to the expected value.

**3.1.3. BP Neural Network.** BP neural network is a supervised algorithm model, which can not only transmit information forward but also transmit errors back through layer training. Its network structure model and training process are shown in Figure 5.

As shown in Figure 5, the training process of the network is mainly based on the comparison between the target output of the training set and the actual output to obtain the error value. It continuously adjusts the training parameters to reduce the error value, and then the functional relationship between the input and output can be obtained through training.

In the forward propagation process of information, it is assumed that the input value of the  $a^{\text{th}}$  node of the hidden layer is  $A_a$ . The input signal of the linear transformation is  $S_c$ , and  $\omega_{ac}$  represents the connection weight between the  $a^{\text{th}}$  node and the  $c^{\text{th}}$  node of the hidden layer. Both  $x$  in formulas (1) and (2) represent the number of input nodes. The relationship among them is shown as:

$$A_a = \sum_{c=0}^x \omega_{ac} \cdot S_c. \quad (1)$$

The output  $B_a$  of this layer is calculated as:

$$B_a = \phi_1 \left( \sum_{c=0}^x \omega_{ac} \cdot S_c \right). \quad (2)$$

Supposing that the input of the  $r^{\text{th}}$  node of the output layer is  $R_r$ , the output of this layer is  $B_r$ . The  $y$  in formulas (3) and (4) represents the number of hidden layer nodes.

According to the previous calculation principle, the expressions of these two values are as formulas (3) and (4):

$$R_r = \sum_{a=0}^y \omega_{ra} \cdot B_a, \quad (3)$$

$$B_r = \phi_2 \left( \sum_{a=0}^y \omega_{ra} \cdot B_a \right). \quad (4)$$

The  $\phi$  in the article represents the activation function. The subscript letter 1 represents the hidden layer activation function, and the subscript letter 2 represents the output layer activation function.

During the back-propagation process of the error, the function expression for calculating the error should be defined first, and then the error value of each training layer should be calculated using the error function. The gradient descent method should also be continuously adjusted to minimize the error. Assuming that the total number of input samples in the training set is  $L$ , the number of samples is set to 1, 2, ...,  $L$ . When the sample numbered  $L$  enters the training process, supposing that the actual output value is  $B_r^L$  and the target output value is  $M_r^L$ , the calculated error of the  $L^{\text{th}}$  sample is as follows:

$$e_L = \frac{\sum_{r=1}^L (M_r^L - B_r^L)^2}{2}. \quad (5)$$

According to the error of a single sample, the total error of the entire training set can be obtained as follows:

$$e = \sum_{L=1}^L e_L. \quad (6)$$

The parameters in the algorithm are adjusted by the gradient descent method. Assuming that the offset of the output layer is  $\delta_r$ , the adjustment amount required for the weights and offsets of the output layer is as formulas (7) and (8):

$$\Delta \omega_{ra} = -\varepsilon \frac{\partial e}{\partial \omega_{ra}}, \quad (7)$$

$$\Delta \delta_r = -\varepsilon \frac{\partial e}{\partial \delta_r}. \quad (8)$$

In formulas (7) and (8),  $\varepsilon$  is the learning rate.

After the partial derivative is calculated, the final adjustment amount of the output layer weight and bias is as formulas (9) and (10):

$$\Delta \omega_{ra} = \varepsilon \sum_{L=1}^L \sum_{r=1}^L (M_r^L - B_r^L) \phi_2'(L_r) \cdot B_a, \quad (9)$$

$$\Delta \delta_r = \varepsilon \sum_{L=1}^L \sum_{r=1}^L (M_r^L - B_r^L) \phi_2'(L_r). \quad (10)$$

According to the same calculation principle, the final adjustment amount of the weights and biases of the hidden layer can be obtained as formulas (11) and (12):

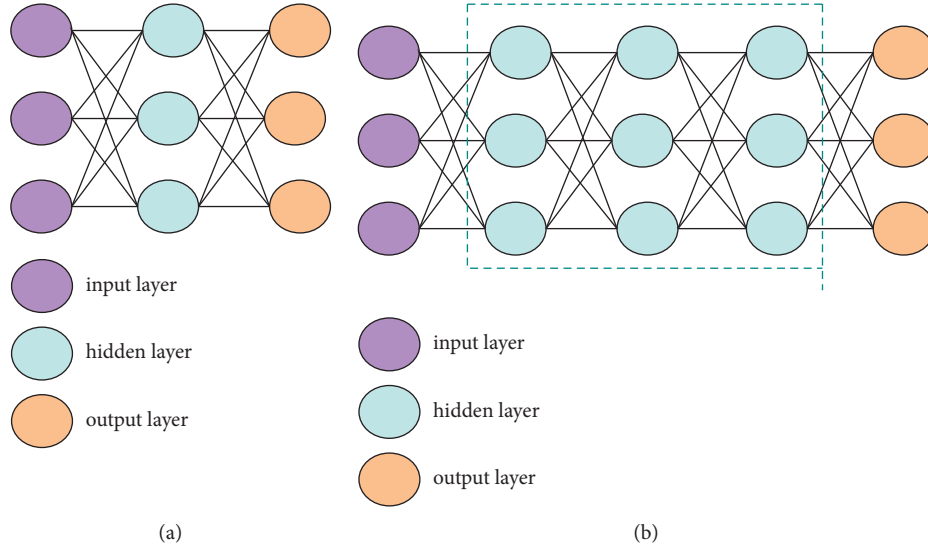


FIGURE 4: Traditional neural network and deep neural network model structure. (a) Traditional neural network structure (b) DNN architecture.

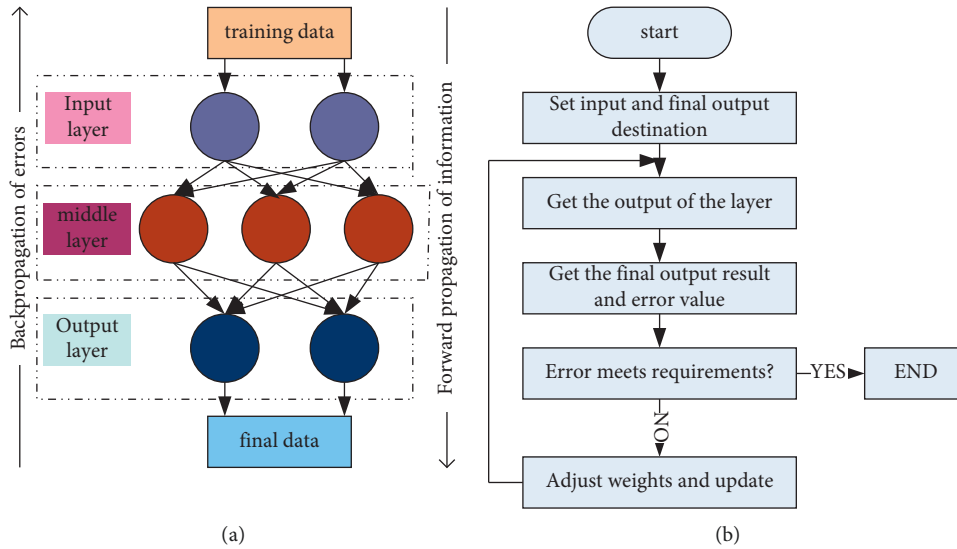


FIGURE 5: BP structure model and training flowchart. (a) BP network structure model (b) BP algorithm process.

$$\Delta\omega_{ac} = \varepsilon \sum_{L=1}^L \sum_{r=1}^L (M_r^L - B_r^L) \phi_2'(L_r) \cdot \omega_{ra} \cdot \phi_1'(A_a) \cdot S_c, \quad (11)$$

$$\Delta\delta_a = \varepsilon \sum_{L=1}^L \sum_{r=1}^L (M_r^L - B_r^L) \phi_2'(L_r) \cdot \omega_{ra} \cdot \phi_2'(A_a). \quad (12)$$

3.1.4. *Restricted Boltzmann Machine (RBM)*. RBM is developed from the Boltzmann machine (BM). The difference between the two is that the nodes of the same layer of RBM are independent of each other, while the nodes of the same layer of BM are related [19]. The two network structures are shown in Figure 6.

Figure 6 is a schematic diagram of the network structure of BM and RBM. From the structure diagram, RBM is improved based on BM, which reduces the connection between the same layers and can save a lot of computing time [20].

Assuming that  $F$  is the total number of visible layer nodes and  $G$  is the total number of hidden layer nodes. The energy function of RBM can be expressed as formula:

$$C(F, G) = - \left( \sum_{t=1}^F \sum_{q=1}^G F_t G_q \omega_{tq} + \sum_{t=1}^F F_t \delta_t + \sum_{q=1}^G G_q \delta_q \right). \quad (13)$$

Assuming that the node state of the visible layer is known, the activation probability of the hidden layer node is obtained as formula:

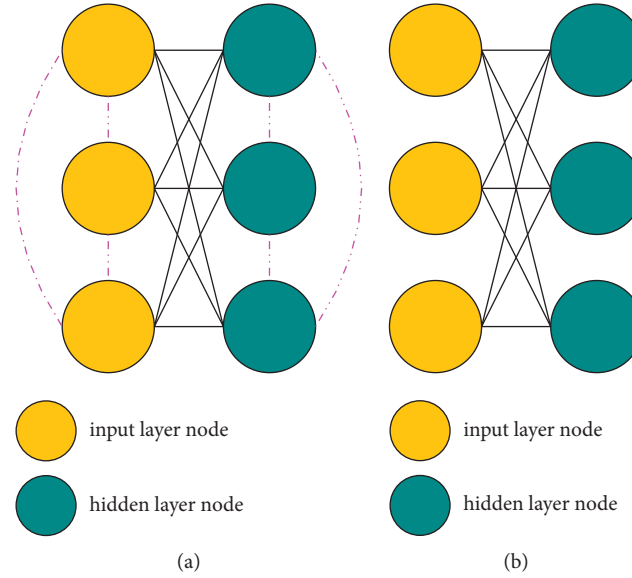


FIGURE 6: BM and RBM network structure. (a) BM (b) RBM.

$$p = \phi_{\text{sigmoid}} \left( \delta_q + \sum_{t=1} F_t \omega_{tq} \right). \quad (14)$$

Meanwhile, the activation probability of the visible layer node can be obtained as:

$$p = \phi_{\text{sigmoid}} \left( \delta_t + \sum_q G_q \omega_{tq} \right). \quad (15)$$

**3.1.5. Deep Belief Network (DBN).** DBN is composed of multiple layers of RBM and one layer of the BP neural network. The output of the lower layer of RBM is used as the input of the next layer of RBM until the output of the last layer of RBM becomes the input of the BP neural network [21, 22]. The DBN structure is shown in Figure 7.

Figure 7 shows the structure and flowchart of DBN. According to the structure in the figure, DBN is trained layer by layer through a multi-layer RBM, and finally, a layer of BP neural network is used for supervised learning. A large number of studies have shown that DBN can obtain a better global parameter through the layer-by-layer training of RBM, and the error backpropagation of the BP neural network can effectively improve the training accuracy of the network.

**3.2. Construction of College English Teaching Mode under Artificial Intelligence.** At present, most universities generally use multimedia technology for English teaching, which to a certain extent promotes the cultivation of students' listening skills and also facilitates teachers' explanations of English texts and classroom exercises. However, there are also some drawbacks, that is, the main body of the classroom is still the teacher, it is difficult for students to take the initiative in the classroom. In addition, in college English education, the

teaching scale of a class is often large. Regardless of whether students are in the classroom or outside the classroom, they rarely communicate with teachers, and teachers have no in-depth understanding of the actual situation of students' learning. The introduction of AI into college English teaching can help teachers better judge students' English learning and conduct targeted teaching activities according to the results of evaluation and analysis. According to the language characteristics of English, some AI technologies that can be used in English teaching are shown in Table 1.

Table 1 shows the current technologies related to English teaching. At present, some functions of these technologies need to be selectively used, because most of them cannot be fully combined with the characteristics of English teaching activities.

According to the characteristics of AI technology combined with college English teaching, the construction of an AI-based college English teaching mode and evaluation system is shown in Figure 8.

As shown in Figure 8, adding AI can give full play to students' classroom initiatives. AI can help students organize learning information so that students can understand their shortcomings in learning, and teachers can also better find The learning of my students. The establishment of an evaluation system through AI reduces the workload of teachers, saves the time of manual screening, and reflects the mistakes of students more clearly. In the process, it can also deal with the interests of students, motivating for students to stimulate their interest in English learning.

### 3.3. Speech Recognition and Evaluation

**3.3.1. Speech Recognition.** In speech recognition, the speech information is first collected by the speech collector, and then the speech signal is preprocessed. In the preprocessing process, there are preemphasis, framing, and windowing. After

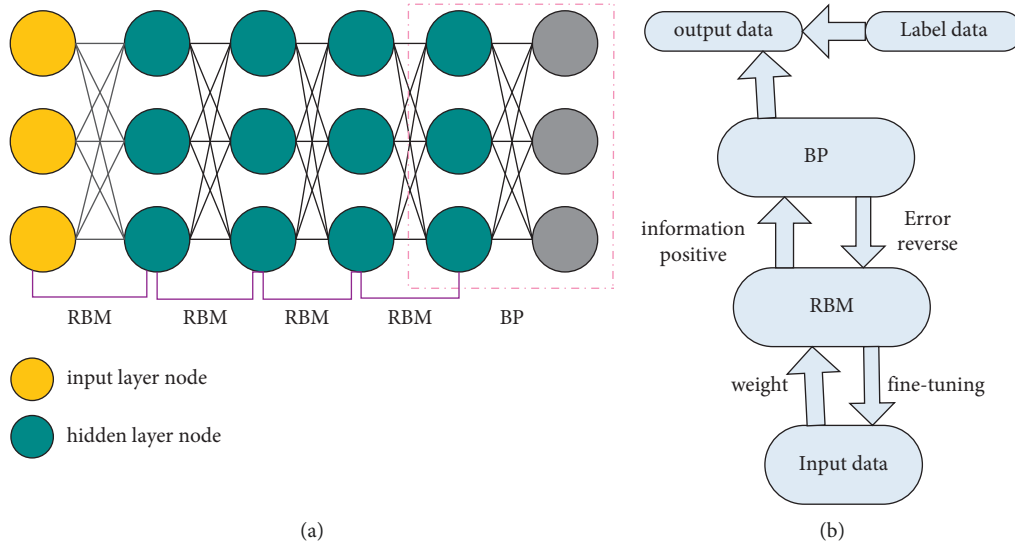


FIGURE 7: DBN structure and flowchart. (a) DBN structure (b) DBN process.

TABLE 1: AI technology related to English teaching.

	Related to “listening and speaking” abilities	Related to “reading and writing” ability	Other related AI technologies
Research areas	Speech segmentation, pronunciation modeling, and other technologies	Word attributes labeling, segmentation, and other technologies	
Specific method	Voice recognition, speech recognition, speaker recognition	Word segmentation, word embeddings, word sense disambiguation, sentiment analysis	AI robot based on question-answering technology
Insufficiency	Does not match the English teaching function	Teacher-student ratio mismatch	

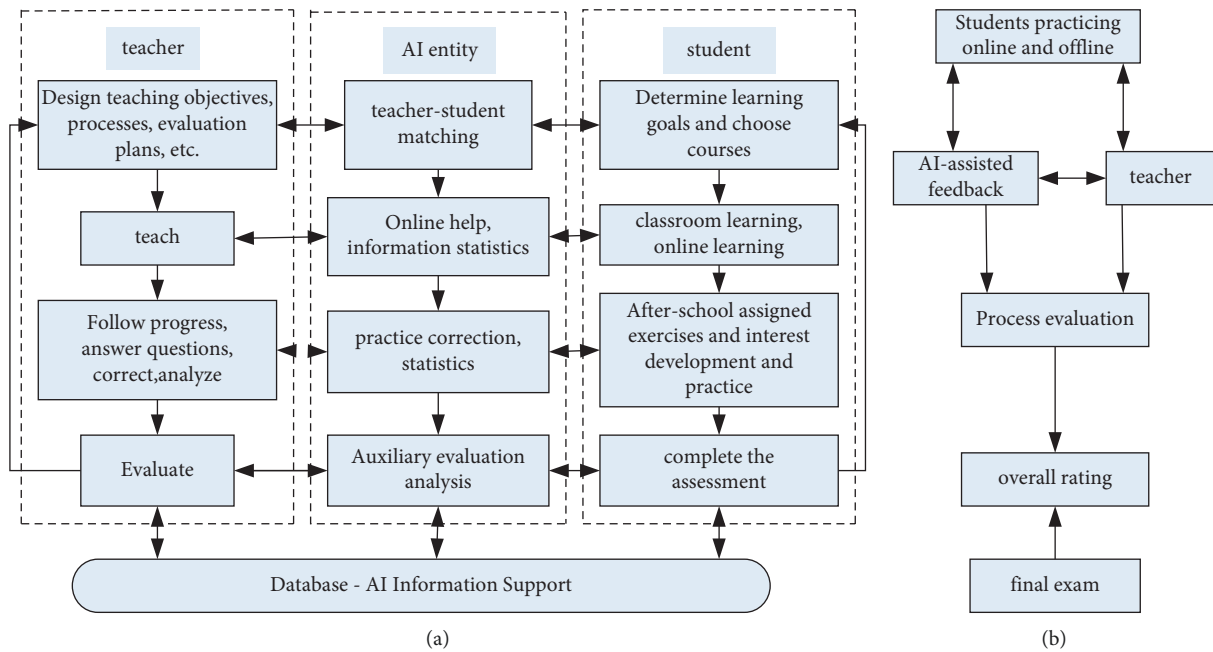


FIGURE 8: AI-based teaching mode and evaluation system. (a) AI-based college English teaching mode (b) evaluation system.

the preprocessing, the speech feature parameters are extracted when mainly frames are extracted, trained, identified and pattern-matched. Finally, the output is processed.

In the preprocessing stage, the signal is filtered as follows:

$$Hz(x) = 1 - \frac{\mu}{x}. \quad (16)$$

In formula (16),  $\mu$  represents the preemphasis coefficient, which is between 0.95 and 0.98.

Since the speech signal is not stable, it is generally assumed that the signal in a short period (10 ms~30 ms) is stable in the research. Therefore, the speech signal needs to be framed in the preprocessing stage. Usually, the number of frames per second is as follows:

$$Z_s = T^{-1} (0.01 < T < 0.03). \quad (17)$$

Since there will be a certain offset between each frame, window processing will be performed after framing, so as to reduce signal discontinuity caused by framing. The window function in the windowing process is as follows:

$$f(k) = \begin{cases} 0, k < 0 \vee k \geq N - 1, \\ 0.54 - 0.46 \cos\left(\frac{2\pi k}{N-1}\right), 0 \leq k \leq N - 1. \end{cases} \quad (18)$$

The windowing process is as follows:

$$\lambda_k = \sum_{r=-\infty}^{\infty} \Pi[I(k)]f(k-r). \quad (19)$$

In formula (19),  $I(k)$  represents the input speech signal sequence.  $\lambda_k$  represents the first-time sequence after processing, and  $\Pi$  represents a certain transformation.

Speech features are extracted using Mel frequency cepstral coefficients (MFCC), and Mel frequency is used to extract speech features based on human ear theory. The approximate relationship between the MFCC and the fundamental frequency is as follows:

$$MFCC(g) = 2595 \log\left(1 + \frac{g}{700}\right). \quad (20)$$

First, in the preprocessing stage, the short-term signal  $s(k)$  is obtained, and then the fast Fourier transform is used to convert  $s(k)$  into the frequency domain signal  $S(r)$ . The short-term energy spectrum  $E(L)$  is calculated, and the signal is continued, which is converted to get the frequency  $M(L)$  in the Mel frequency domain and the Mel domain energy spectrum  $E(M)$ . Then,  $E(M)$  is processed through the filter  $H_j(r)$  to get the logarithmic representation of the filter, which is given by:

$$E(H) = \text{In}\left(\sum_{r=0}^{N-1} E(M)H_j(r)\right). \quad (21)$$

The logarithmic representation of the filter is then discrete cosine transformed to obtain the MFCC coefficients as follows:

$$MFCC(q) = \sum_{j=0}^{N-1} E(H)\cos\left(\frac{\pi q(j-0.5)}{H}\right). \quad (22)$$

**3.3.2. Pronunciation Quality Assessment.** The pronunciation quality evaluation of spoken English is mainly based on a comprehensive evaluation of the intonation, rhythm, etc. of oral English. Since the sentence form and rhythm of Chinese are quite different from those of English, these are the points that need to be paid attention to in spoken English. In English sentences, the pause rhythm of sentences and paragraphs is also very important to the understanding of English content. To a certain extent, the real meaning expressed by the sentences can be better understood, and the reader's perception of the important information in the sentence content can be judged by the rhythm of reading English. Under normal circumstances, the evaluation of English pronunciation quality is not only considered from the content, that is, the integrity of the content, but it is more important to consider the pronunciation, intonation, speed and rhythm, the grasp of the light and stress, and the fluency of the process. This article will evaluate the quality of English pronunciation from the aspects of intonation, speed, and rhythm. The details are shown in Table 2.

As shown in Table 2, the evaluation of the evaluation indicators of speech rate, intonation, and rhythm is mainly based on the comparison and evaluation of standard speech.

**3.3.3. DBN-Based Speech Recognition.** Combined with the speech recognition steps mentioned above, the speech signal is first preprocessed and its feature is extracted. Feature fusion is performed, and then the data is input into the DBN. The original feature input here is the SDC parameter after operation based on MFCC. Layer-by-layer pre-training is performed by DBN, and weights and parameters are initialized and updated. Parameters are fine-tuned in the last layer of DBN BP neural network.

## 4. College English Teaching Experiment

**4.1. Dataset.** This paper used DBN to evaluate English pronunciation according to the evaluation system under the artificial intelligence teaching mode and compared it with manual evaluation. The subjects were 30 junior non-English majors in a normal university, including 15 males and 15 females. Subjects were recorded by recording software. To verify the recognition effect of the model, a dialogue was selected from the third lesson of the first volume of the college English listening textbook for the test experiment. The specific settings of the two groups of experiments are shown in Table 3.

It can be seen from the parameter settings in Table 3 that the sampling rate and coding number of the two groups of experiments are consistent, and the MFCC characteristic parameters of the experiments are also unified to the 13th order.



TABLE 2: Evaluation index table.

Index	Method
Intonation	MFCC
Tone of voice	Autocorrelation function method, DTW
Speed of speech	Based on speech duration
Rhythm	DTW

TABLE 3: Dataset parameter settings.

-	Test experiment	Pronunciation evaluation experiment
Sampling rate	16 KHz	16 KHz
Coding	16	16
Recording software	—	CoolEdit

In the pronunciation evaluation experiment, the subjects recorded a total of 8 sentences, all of which are commonly used sentence patterns and common sentences in spoken English.

*4.2. Pronunciation Evaluation Experiment Based on DBN.* First, the test data into the model is input and model verification experiments can be conducted. Then, the influence of the number of hidden layer nodes and the number of hidden layers on the recognition effect were studied. Different input durations for analysis were finally selected.

In the pronunciation evaluation experiment, the data preprocessing process will process the speech feature parameters of variable length and convert them into features with the same length. The specific processing process is as follows: First perform average segmentation processing on the characteristic parameters of the speech signal. This paper is divided into 4 segments. Then continue to perform average segmentation processing on each segment. The value of the sub-segmentation in this paper is 3, and each segment is divided into 4 segments. Then calculate the mean of the small segments in the segment to obtain the mean vector of each small segment, and finally combine all the small segment mean vectors into a matrix. The speech signal designed in this paper was 16 frames. The 8 sentences include that

- (1) Don't bury your head in the sand;
- (2) You have got to dig in your heels;
- (3) It's a pleasure working with you;
- (4) I had a wonderful time here;
- (5) You've got a point there;
- (6) You can't please everyone;
- (7) How do you like your new job?
- (8) What goes around comes around?

### 4.3. Experiment Results

*4.3.1. Test Results.* The recognition effects of different hidden layer numbers and hidden layer nodes are shown in Figure 9.

According to the data in Figure 9, the more hidden nodes in the model are, the lower the error rate of the training model

will be. The lower the error rate of the model is, the higher the training accuracy of this method will be, and the more accurate the recognition and evaluation of pronunciation will be. When choosing different time series, it also has an impact on the recognition effect. From the data in the figure, when the selected time series is longer, the training effect of this method is better, which may be related to the characteristics of the voice. The short-term voice does not express the meaning of the sentence obviously, so it is not conducive to the recognition of the system. However, the increase in the number of hidden layer nodes does not significantly improve the recognition effect of such sequences. For data with short time series, although the recognition effect is slightly worse than that of long time-series data, when the number of hidden layer nodes increases, the recognition effect of this type of data is improved even higher. It can be seen that, whether it is short-time series or long-time series data, the more the number of hidden layer nodes is, the better the recognition effect of the model will be.

When the number of hidden layers is larger, it does not mean that the training accuracy of the model will be better. According to the trend line in Figure 9, whether it is long-term sequence data or short-term sequence data, when the number of hidden layers reaches a certain value, a state with the best effect will appear. For short time-series data, when the number of hidden layers is 6, the error rate is the lowest at this time, while for two sequences with slightly longer time, when the number of hidden layers is 4, the recognition effect is the best.

According to the analysis of the test results, the validity and feasibility of the model constructed in this paper are illustrated, and the accuracy of the training results is also related to the selected voice data. Therefore, this paper will select the optimal combination mode to set the parameter values according to the test results.

To better highlight the effect of the model, the test set will be used for comparative experiments to analyze the recognition effect under different recognition models. This paper chooses several different methods to train the data. The specific experimental results are shown in Figure 10.

It can be seen from Figure 10 that the recognition model constructed in this paper is better than other recognition models, which verifies the effectiveness of this model in speech recognition.

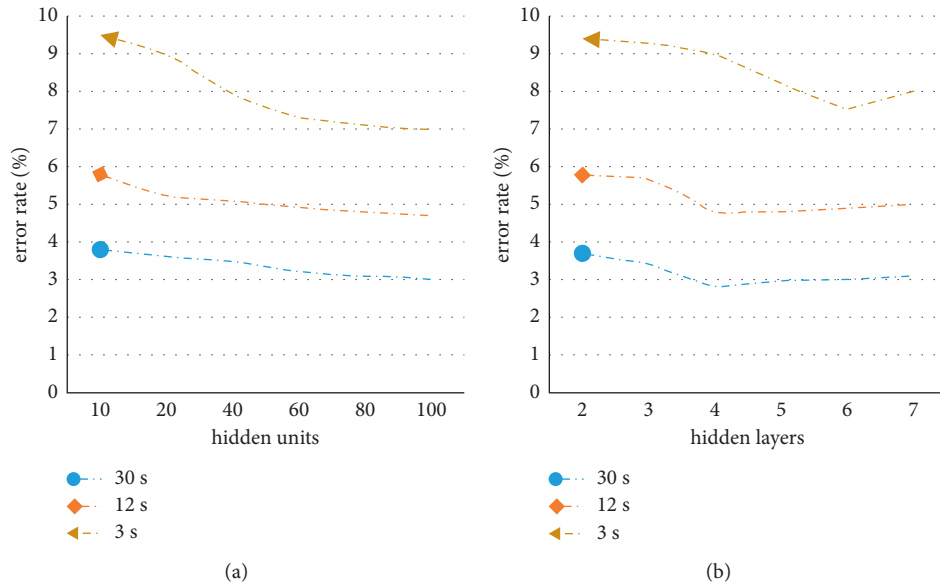


FIGURE 9: Influence trend of hidden layer parameters on recognition effect.

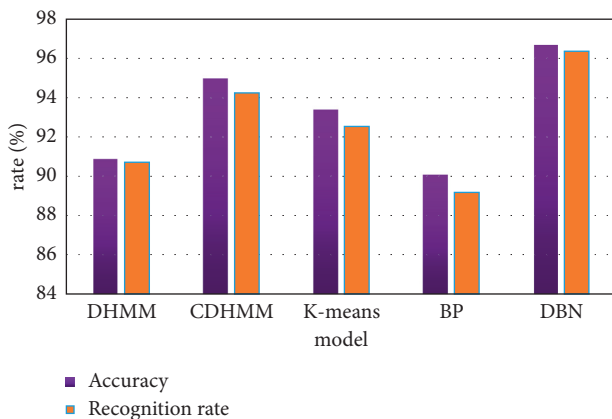


FIGURE 10: Accuracy and recognition rate under different models.

#### 4.3.2. Pronunciation Evaluation Experiment Results.

According to 30 subjects, each with 8 sentences, a total of 240 sentences in the pronunciation of English speaking, the pitch, speed, intonation, and rhythm were evaluated and scored. First, two college English teachers in the school scored it manually, and then it was scored by a machine to calculate the consistency of the two scores. The relevant evaluation indicators were the inhibition rate and the adjacent consistency rate. Consistency rate was the ratio of the total number of consistent evaluations of the two methods in the total sample, and the adjacent consistency rate was the ratio of the total number of samples evaluated by the two methods to be consistent and similar to the two methods in the total sample. To better perform mathematical calculations, the score was set to 1–4. The higher the score was, the better the pronunciation quality would be. The specific grades are shown in Table 4.

As shown in Table 4, the manual evaluation would be based on the students' pronunciation performance. In terms

of intonation, the score is mainly based on whether the sentence is complete and whether there are obvious pronunciation errors. Speaking speed is mainly based on whether the sentence speed is too fast or too slow to score. In terms of intonation, it is scored according to the pronunciation of stress and whether there is a distinction between light and stress. The score of rhythm is mainly based on whether the pronunciation of the sentence is coherent and whether the sentence segmentation accurately expresses the meaning of the sentence.

According to the results of the two scoring systems, the results of the consistency of the scoring results of the two methods are shown in Table 5.

As shown in Table 5, among the four evaluation indicators of intonation, speech rate, intonation, and rhythm, 208/198/199/206 are consistent in evaluation, and 30, 41, 38, and 30 are adjacent. According to the experimental results, the agreement rate and adjacent agreement rate results of the two methods are shown in Figure 11.

As shown in Figure 11, the consistency between manual evaluation and machine evaluation is relatively high. In the pitch evaluation index, the consistency rate is 86.66%, and the adjacent consistency rate reaches 99.16%. In the speech rate evaluation index, the consistency rate is 82.5%, and the adjacent consistency rate reaches 99.58%. Among the intonation evaluation indicators, the consistency rate is 82.91%, and the adjacent consistency rate is 98.75%. Among the rhythm evaluation indicators, the consistency rate is 85.83%, and the adjacent consistency rate is 98.33%. The consistent results of these evaluations indicate that the evaluation system in this paper is feasible.

Through comparative analysis, the effectiveness of the DBN-based English pronunciation evaluation system is verified, indicating that artificial intelligence has a certain enlightening effect on the development of college English education. In university education, more multimedia and

TABLE 4: Evaluation score table.

Score	Intonation	Speed of speech	Tone of voice	Rhythm	Overall
4	No obvious pronunciation mistakes	Moderate	Accurate and natural	Accurate accent pronunciation	Pronunciation is generally excellent
3	No serious pronunciation mistakes	Slightly faster or slower	More accurate and natural	Accent pronunciation is more accurate	Pronunciation in general good grasp
2	There are pronunciation errors that interfere with comprehension	Fast or slow	Basically accurate but not natural enough	Normal accent	Pronunciation in general grasp the general
1	Serious pronunciation errors that interfere with comprehension	Too fast or too slow	Inaccurate and unnatural	Mispronunciation of accent	Pronunciation in general poor grasp

TABLE 5: Evaluation index results.

	Consistent	Adjacent	Others
Intonation	208	30	2
Speed of speech	198	41	1
Tone of voice	199	38	3
Rhythm	206	30	4

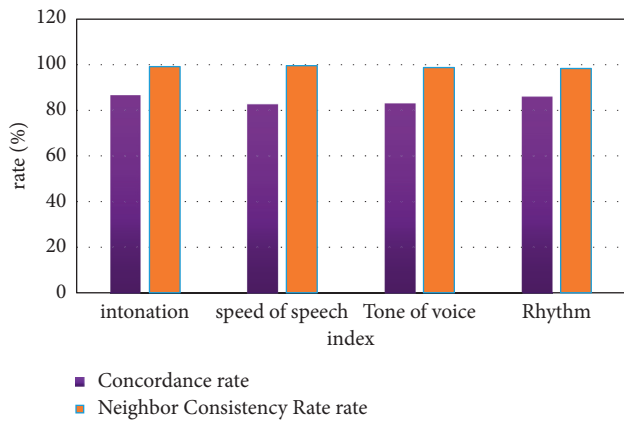


FIGURE 11: Consistency and adjacent consistency of two evaluation methods.

Internet technologies can be used to evaluate and analyze students’ learning. The classroom must be student-centered to ensure the dominant position of students in the teaching process. In the follow-up research, other goals in comprehensive English ability can be analyzed and combined with artificial intelligence.

### 5. Conclusions

This paper constructs a college English teaching model through AI and establishes an English pronunciation recognition evaluation system through DBN through the evaluation system in the constructed teaching model. It also verifies the validity of the constructed evaluation model through test experiments. Finally, non-English major subjects are selected to conduct a pronunciation quality evaluation test, and the agreement rate and adjacent agreement rate are selected to compare and analyze the evaluation results of the manual evaluation and the machine evaluation model constructed in this paper. According to the final experimental results, the pronunciation evaluation method

model constructed in this paper has a high consistency with the results of the artificial evaluation. According to some problems existing in the current college English education, such as the shortage of teachers and the single evaluation, the lack of students’ dominant position in the classroom, and the lack of spoken English teaching activities in the pronunciation evaluation of this paper, similar problems in the current environment, it is possible to consider using technologies such as AI in the context of the Internet to break through. The experiment of English pronunciation evaluation in this paper proves the effectiveness of this method.

### Data Availability

The data used to support the findings of the study can be obtained from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### References

- [1] Z. Wen, B. Wang, D. Coniam, and B. Xie, “Calibrating the CEFR against the China standards of English for college English vocabulary education in China,” *Language Testing in Asia*, vol. 7, no. 1, pp. 1–18, 2017.
- [2] X. Li, “Characteristics and rules of college English education based on cognitive process simulation,” *Cognitive Systems Research*, vol. 57, pp. 11–19, 2019.
- [3] J. . Gong, “An innovation method of college English education in the new media environment[J],” *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 13, pp. 449–455, 2017.
- [4] Z. Nie, “A study on the application of continuing education mode in college English teachers in Internet+ era[J],” *Boletin Tecnico/technical Bulletin*, vol. 55, no. 11, pp. 504–511, 2017.
- [5] K. Liu, “An empirical study on college English O2O teaching system based on MOOC education[J],” *Revista de la Facultad de Ingenieria*, vol. 32, no. 14, pp. 784–789, 2017.
- [6] Y. Hao and J. Su, “A study on college English flipped classroom education mode under the background of educational information,” *Revista de la Facultad de Ingenieria*, vol. 32, no. 15, pp. 174–177, 2017.
- [7] Y. . Zhang, “Improvement and optimization of function modules of college English distance education based on intercultural theory,” *Revista de la Facultad de Ingenieria*, vol. 32, no. 14, pp. 41–46, 2017.
- [8] C. J. Lin, “An online peer assessment approach to supporting mind-mapping flipped learning activities for college English

- writing courses,” *Journal of Computers in Education*, vol. 6, no. 3, pp. 385–415, 2019.
- [9] M. Ravanelli, P. Brakel, M. Omologo, and Y. Bengio, “Light gated recurrent units for speech recognition,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 2, pp. 92–102, 2018.
- [10] R. Yazdani, A. Segura, J. M. Arnau, and A. Gonzalez, “Low-power automatic speech recognition through a mobile GPU and a viterbi accelerator,” *IEEE Micro*, vol. 37, no. 1, pp. 22–29, 2017.
- [11] E. Pakoci, B. Popović, and D. J. Pekar, “Improvements in Serbian speech recognition using sequence-trained deep neural networks,” *SPIIRAS Proceedings*, vol. 3, no. 58, p. 53, 2018.
- [12] W. Bo, K. Li, F. Ge et al., “An end-to-end deep learning approach to simultaneous speech dereverberation and acoustic modeling for robust speech recognition[J],” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 8, pp. 1289–1300, 2017.
- [13] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, “The challenge of non-technical loss detection using artificial intelligence: a survey,” *International Journal of Computational Intelligence Systems*, vol. 10, no. 1, p. 760, 2017.
- [14] P. Havinga, N. Meratnia, and M. Bahrepour, “Artificial intelligence based event detection in wireless sensor networks [J],” *University of Twente*, vol. 85, no. 6, pp. 1553–1562, 2017.
- [15] L. Cavaglione, M. Gaggero, J. F. Lalande, W. Mazurczyk, and M. Urbanski, “Seeing the unseen: revealing mobile malware hidden communications via energy consumption and artificial intelligence[J],” *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 4, pp. 799–810, 2017.
- [16] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, “Neuroscience-inspired artificial intelligence,” *Neuron*, vol. 95, no. 2, pp. 245–258, 2017.
- [17] E. Burton, J. Goldsmith, S. Koenig, B. Kuipers, N. Mattei, and T. Walsh, “Ethical considerations in artificial intelligence courses,” *AI Magazine*, vol. 38, no. 2, pp. 22–34, 2017.
- [18] Y. Guan and F. F. Ren, “Application of artificial intelligence and wireless networks to music teaching,” *Wireless Communications and Mobile Computing*, vol. 2021, no. 12, 16 pages, Article ID 2811398, 2021.
- [19] S. K. Pandey, R. R. Janghel, A. V. Dev, and P. K. Mishra, “Automated arrhythmia detection from electrocardiogram signal using stacked restricted Boltzmann machine model,” *SN Applied Sciences*, vol. 3, no. 6, pp. 1–10, 2021.
- [20] K. Tan, F. Wu, Q. Du, P. Du, and Y. Chen, “A parallel Gaussian–Bernoulli restricted Boltzmann machine for mining area classification with hyperspectral imagery,” *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 2, pp. 627–636, 2019.
- [21] O. Oyedotun and A. Khashman, “Iris nevus diagnosis: convolutional neural network and deep belief network,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 25, no. 2, pp. 1106–1115, 2017.
- [22] L. Shiqi, T. Shengwei, Y. Long, Y. Jiong, and S. Hua, “Android malicious code classification using deep belief network,” *Ksii Transactions on Internet & Information Systems*, vol. 12, no. 1, pp. 454–475, 2018.