[1]Haiyan Li

Design and Development of an Oral English Evaluation System Based on Speech Recognition Technology and A Scoring Algorithm

Journal of Electrical Systems

Abstract: - Proficiency in oral English plays a significant role in different professional and academic settings. However, evaluating oral English proficiency precisely is challenging because of the subjective nature of oral communication. In this study, we developed an innovative Oral English Evaluation System (OEES) by combining the efficiencies of speech recognition technology and scoring algorithms. Initially, an intensive oral English database was collected, containing numerous persons' voice recordings. The collected voice recordings are pre-processed to remove the background noise, ensuring clarity in the voice signals. Then, an automatic speech recognition algorithm was developed using the recurrent neural network (RNN) to transcribe voice signals into text. This module trains the system to map the voice signals to their corresponding textual characteristics. Finally, an adaptive scoring algorithm was incorporated into the designed OEES to evaluate the oral English proficiency of the individuals. The adaptive scoring algorithm considers different factors like fluency, grammar, pronunciation, and vocabulary for evaluating the individual's oral English proficiency as "excellent," "good," and "poor." The presented framework was modeled in Python and validated across diverse natural English-speaking databases. The experimental results are assessed in terms of accuracy, precision, recall, and error rate. The implementation outcomes suggest that the proposed OEES framework accurately evaluates oral English proficiency.

Keywords: Automatic Speech Recognition Technology, Adaptive Scoring algorithm, Oral English Evaluation, Academic and Professional Pursuits

1. INTRODUCTION

The demand for reading, listening, writing, and speaking skills in the English language is increasing due to the rapid development and advancement of computing technologies and networks [1]. Assessing these skills, especially oral English, is significant in evaluating an individual's language proficiency. However, the manual assessment of oral English needs huge labor power, increasing the test cost. Hence, a computer-based oral assessment is designed to enhance the effectiveness of test administration. This computer-based system uses emerging technologies such as speech recognition algorithms (SRA) to automatically transcribe the voice signal into textual characteristics [2]. Generally, the SRA was developed using artificial intelligence techniques, which transcribe the voice signals into their corresponding texts. Further, they use natural language processing (NLP) techniques to examine oral English proficiency [3]. This usage of NLP offers a detailed evaluation of an individual's oral skills by providing feedback on grammar, fluency, pronunciation, and vocabulary usage. Generally, these algorithms are trained using a large, diverse speech database to precisely evaluate the correlations between the words [4]. However, these algorithms cannot classify the oral proficiency of each individual.

Hence, the studies utilized big data analytics and machine learning (ML) algorithms to perform classification of an individual's oral performance. Using big data analytics enables one to understand the intricate patterns in speech signals, assisting the system in providing feedback to improve performance [5]. On the other hand, the ML algorithm adapts to each speaker's oral proficiency levels, classifying their levels into different categories. This classification assists the trainers in optimizing the learner's oral proficiency levels [6]. However, training the ML models requires large databases, and they are computationally intensive. Moreover, these algorithms are limited to scalability and lack adaptability. Therefore, recent studies concentrated on designing scoring algorithms to evaluate the oral test in English [7]. The scoring algorithm classifies each person’s oral proficiency level by assessing their scores in pronunciation, vocabulary usage, fluency, etc. The traditional scoring algorithm uses a correlation statistical feature assessment technique for evaluating the oral English assessment score.

[1] Changchun College Of Electronic Technology, Changchun, Jilin, 130000, China

*Corresponding author e-mail: 15584404636@163.com

However, the pronunciation feature sets' frequency and phase characteristics are scattered, resulting in reduced accuracy and lower system stability [8]. On the other hand, the scoring algorithm based on sequence matching and big data analytics faces issues in mapping the attributes in textual characteristics, resulting in inaccurate classification. These drawbacks of the existing studies demand an innovative evaluation system for assessing oral English proficiency. Therefore, we proposed a collaborative technique in this study by integrating the efficiency of speech recognition technology and scoring algorithms. This study aims to develop an automatic speech recognition technique to transcribe the voice signals into corresponding texts, which are processed using the adaptive scoring algorithm to determine the oral English proficiency of each individual.

The enduring sections of the article were organized as follows: section 2 provides a detailed review of the existing works, section 2 presents the architecture of the proposed framework, section 4 analyzes the results of the developed work, and Section 5 depicts the article's conclusion.

2. RELATED WORKS

A few recent studies related to oral English evaluation system are reviewed below,

Ping Li et al [9] developed an automatic scoring approach to evaluate the oral English test. Conventional speech signal analysis focuses on the capture of informative attributes, which degrades the evaluation accuracy. This study developed an automatic scoring technique using sequence matching. This system adapts a spoken speech feature engineering algorithm and a dynamic, optimized spoken English model to evaluate oral English proficiency scores. This methodology was evaluated using different corpus voice recording databases and achieved 93.12% accuracy in assessing the oral English scores. However, this framework cannot process the spectral and spatial features in the audio signals.

Xin Wang et al. [10] designed and implemented a scoring algorithm for open-spoken English using the improved neural network (NN) approach. This algorithm evaluates the oral recording from both phonetic and text levels. In addition, this system evaluates the spoken speech and spoken content separately using various scoring algorithms. Advanced pattern identification and signal processing techniques are also used to develop an acoustic feature extraction framework. The implementation outcomes of the study manifest that it obtained 92% accuracy in assessing open-spoken English. However, this methodology lacks scalability.

Wang Jing et al. [11] presented an innovative system using speech recognition technology and speech synthesis to assist English learners in enhancing their oral pronunciation. The system enhances learners' pronunciation levels by automatically assessing the learners' pronunciation accuracy and offering targeted correction and training. The speech recognition module transforms the oral recordings into textual characteristics, and the speech synthesis approach generates corrected audio to help the learners improve their pronunciation. Although this methodology showed greater results, training the models is complex and time-consuming.

Xia Sun et al. [12] proposed an English speech-scoring approach using the NN technology. The primary concern of this study is to develop an English speech-scoring data system leveraging the efficiency of the NN algorithm. This study utilized the data (voice recordings collected from the students. The experiment displayed that the modified backpropagation algorithm obtained a higher oral English recognition rate than the conventional BP algorithm. However, this methodology requires fine-tuning to optimize its training process.

Lei Bao and Jing [13] developed an auxiliary teaching framework for spoken English using Speech Recognition Technology. This framework combined the efficiency of HMM speech cores and the DWJ algorithm for creating the teaching system. This study was validated using professional data, and the experiment results show that the proposed algorithm is reliable in improving standard English language proficiency. However, assessing the score of each individual is not possible.

3. PROPOSED COLLABORATIVE TECHNIQUE FOR ORAL ENGLISH EVALUATION

This study develops a collaborative framework for accurately evaluating oral English by processing voice inputs. The proposed methodology integrates the efficiencies of speech recognition technology and the scoring algorithm for assessing and classifying each person's oral English proficiency. The corpus English recording

database was initially collected and imported into the system. Then, adaptive filtering was applied to pre-process the collected voice recording by eliminating the background or unwanted noises from the voice signals. This increases the clarity of the signal, making it effective for subsequent analysis. Further, an automatic speech recognition module was created using a recurrent neural network (RNN) to transcribe the voice input into its corresponding textual characteristics. In this module, initially, feature extraction was performed using the Mel-Frequency Cepstral Coefficient (MFCC) to capture the most informative attributes from the pre-processed audio signals. Then, the RNN architecture was trained using the extracted feature sequences to understand and capture the correlations and long-range dependencies in the data. This enables the system to transcribe the audio signals into their corresponding texts. Finally, an adaptive scoring algorithm was created to evaluate the oral English proficiency of each individual. This algorithm determines the average score considering different aspects of oral English, such as fluency, grammar, pronunciation, vocabulary usage, and overall coherence and classifies the individual's performance into three different categories. Figure 1 provides the architecture of the developed framework.

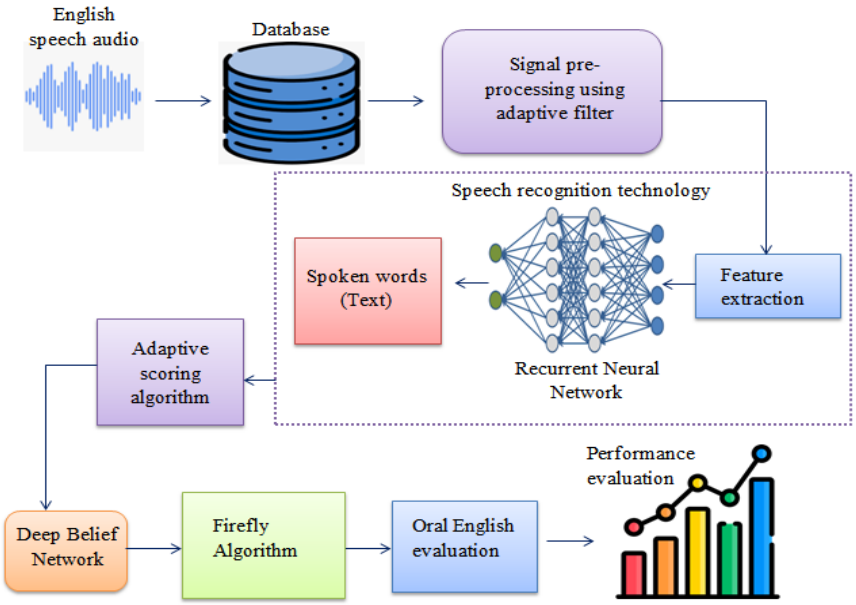


Figure 1: Architecture of Proposed Methodology

3.1 Data collection and pre-processing

The proposed work commences with the collection of corpus recording data. This study utilizes a well-established publicly available database under Pannous, a collaboration on enhancing speech recognition from the librosa library. This dataset contains audio files (voice recordings of different speakers). This audio database was pre-processed using an adaptive filtering technique to remove the background noise and improve speech clarity. Adaptive filtering is an approach widely used in signal processing to reduce background noises, echoes, and other interferences to improve the recorded voice's quality. It is mathematically expressed in Eqn. (1).

$$y(t) = \sum_{h=1}^{n} a(h).x[t-h] \quad (1)$$

Where $y(t)$ indicates the filter output at time t, $x(t)$ denotes the audio signal at time t, $a(h)$ represents the adaptive filter coefficients, which is continuously updated depending on the input and desired output, and n defines the filter order. The dataset is collected from 15 speakers, including males and females, and contains isolated spoken digits. Each speaker utters a digit 16 times, leading to 240 instances for each digit. The database contains 2400 different audio files, which are used to train the system for accurate oral English evaluation. This database was split in the ratio of 90:10 for training and validation of the proposed model.

3.2 Automatic Speech Recognition Technology

Automatic speech recognition defines the process of converting the audio signals into their corresponding textual characteristics. This module accepts the pre-processed audio signals as input, and then MFCC was used to capture the most informative attributes from the pre-processed signals. MFCC is a feature engineering technique widely used in speech processing to extract the spectral characteristics of audio signals by approximating the human auditory system's response to different frequencies. Initially, the pre-processed audio signal is divided into short overlapping frames, and then each frame is multiplied with a window function. In the presented study, we applied a Hamming window to each frame. This step minimizes the spectral leakage and then determines the spectrum magnitude of each frame by using Fast Fourier Transform (FFT). The estimated power spectrum is then passed through a filter bank of triangular filters. The triangular filters are spaced evenly on the mel scale, enabling them to capture spectral features of the audio signals, which is expressed in Eqn. (2).

$$M_r(f) = \begin{cases} 0 & if\ f < f_o \\ \dfrac{f - f_o}{f_c - f_o} & if\ f_o \le f < f_c \\ \dfrac{f_u - f}{f_u - f_c} & if\ f_c \le f < f_u \\ 1 & if\ f \ge f_u \end{cases} \quad (2)$$

Where $M_r(f)$ denotes the magnitude response of the *mth* triangular filter in the filter bank f_o f_c, and f_u indicates the lower, center, and upper frequencies of the triangular filters, respectively. Further, the logarithm of the filterbank energies are evaluated to approximate the human perception of sound intensity.Finally, the Discrete Cosine Transform (DCT) was applied to the log-filterbank energies to minimize the correlation between the signals. This step produces MFCC coefficients, indicating the audio signals' spectral attributes. These features are fed as input to the RNN for transcription of audio signals into texts.

3.2.1 Recurrent Neural Network

An RNN is a feed-forward neural network that uses a deep supervised learning strategy to perform sequence prediction functions. In the proposed work, we utilized RNN to transcribe the audio signals into their corresponding textual characteristics, which aids in evaluating the oral English proficiency of the individual. Here, we used the Long Short-Term Memory (LSTM) architecture of RNN for mapping the words corresponding to the voice inputs. Similar to conventional neural networks, the RNN contains three layers: input, hidden, and output. The unique feature of RNN structure is the presence of memory blocks in the hidden layer (LSTM), which enables it to store data over time. This unique component in the RNN enables it to capture long-range dependencies in the data over time. The LSTM unit contains cells, forget gate, input gate and output gate for capturing the patterns and correlations within the data for transcribing the audio signals into texts. The input layer of RNN accepts the extracted feature sequences from the MFCC block as inputs. The input layer forwards the feature vector into the LSTM block by converting the feature sequence into a suitable form, which the LSTM structure can quickly process. These input values can be preserved in the cell state only if the input gate of LSTM allows them. The input gate and cell state at time step *n* is expressed in Eqn. (3), and (4).

$$I_n = \sigma\left(w_{ig}.\left[h_{n-1}.x_n\right] + b_{ig}\right) \quad (3)$$

$$\hat{C}_n = \tanh\left(w_{cs}.\left[h_{n-1}.x_n\right] + b_{cs}\right) \quad (4)$$

Where σ indicates the sigmoid activation function, w_{ig}, and w_{cs} defines the weight matrices, and *b* denotes bias vector. The input gate decides and controls what input features need to be stored in the cell state. The forget gate regulates the weight of the state cell component, and the forget gate is evaluated using Eqn. (5).

$$F_n = \sigma\left(w_{fg}.\left[h_{n-1}.x_n\right] + b_{fig}\right) \quad (5)$$

Where F_n defines the forget gate, which regulates what previous or past information needs to be preserved and forgotten from the cell state. Considering the forget gate value, the new state of the memory cell is upgraded using Eqn. (6).

$$\hat{C}_n = I_n.\hat{C}_n + F_n.\hat{C}_{n-1} \quad (6)$$

Further, for the new state memory cell, the output gate value is expressed in Eqn. (7).

$$O_n = \sigma\left(w_{og}.\left[h_{n-1}.x_n\right] + b_{og}\right) \quad (7)$$

Where O_n indicates the output gate value. The outcome value of the cell is evaluated using the Eqn. (8).

$$h_n = O_n * \tanh\left(c_n\right) \quad (8)$$

This cell outcome encapsulates the information learned from the input set and serves as the network's memory. By selectively upgrading the LSTM cells considering the input, forget, and output gates, which enables the system to capture the correlation and long-term dependencies in the audio sequences. This unique feature of RNN architecture enables transcribing the audio signals into text (words) accurately. After several iterations and updation, the LSTM block passes the hidden state into the output layer, which produces the textual form of the audio signals. This transcribed textual sequence was forwarded into the adaptive scoring algorithm block for evaluating the oral English proficiency.

3.3 Adaptive scoring algorithm

The adaptive scoring algorithm is an optimized classification algorithm, which estimates the score for each individual to evaluate their oral English proficiency. This scoring algorithm accepts the textual representation of the audio signals as input, and evaluates the score considering different aspects like grammar, vocabulary usage, pronunciation, etc. The developed scoring algorithm utilizes the Deep Belief Network (DBN) and the Firefly optimization for precise evaluation of oral English. The DBN is an artificial neural network, which consists of a stack of Restricted Boltzmann Machines (RBMs). The RBM is a type of perceptron, and the DBN system contains two layers namely, explicit and hidden. The neurons between these two layers are interconnected via two directions (weights and bias). Typically, the DBN system involves two processes namely: pretraining, and fine-tuning. In pretraining phase, the DBN model is trained using the English corpus database via unsupervised learning algorithm. This process involves layer-by-layer training of the RBM network. This intensive training mechanism enables the system to determine the scores of oral English Proficiency levels. In this training, the system learns to extract the most informative features, which are significant for mapping oral English proficiency scores. These features include linguistic characteristics such as grammar, vocabulary usage, fluency, pronunciation, etc. After training, the model is tested with the unknown textual to evaluate its generalization ability. The output of the DBN is in the form of probability function (scores), which is expressed in Eqn. (9).

$$P_v\left(u \mid \beta\right) = \frac{\sum_{i=1}^{M} e^{-E_f\left(u,h|\beta\right)_{\max}}}{z\left(\beta\right)} \quad (9)$$

If the output probability P_v is greater than 0.8, then the oral English proficiency of the person is “excellent”, if the probability value lies between 0.6 and 0.8, then the oral English proficiency is “good”, and if the probability value lies below 0.6, then the oral English proficiency of the person is “poor.” As per these conditions, the system classifies the proficiency levels of each individual. The next step in DBN is fine-tuning in which the DBN parameters such as weights and bias are optimized to minimize the classification error or loss function. Although the DBN model produces accurate evaluation of oral English, its performances are greatly influenced

by these parameters. By fine-tuning the entire DBN module, we can improve the accuracy in oral English evaluation. Typically, this is done by using a backpropagation algorithm. In the proposed work, the fine-tuning of DBN parameters is performed using the Firefly optimization algorithm. The Firefly optimization is a meta-heuristic optimization algorithm designed based on the flashing characteristics of fireflies. Typically, the fireflies use their flashlight to attract their prey. The brightness of the fireflies are interconnected with the objective function. In the proposed work, the objective function is to minimize the classification error or loss by fine-tuning the DBN network. The brightness of the fireflies indicates the classification accuracy. If the accuracy is high, the brightness of the solution will be high. The fine-tuning phase commences with the initialization of DBN parameters such as weights and bias with population size, and maximum iteration count. After initialization, the fitness (brightness) of each parameter solution was determined based on the objective function. For each iteration, the FA technique refines and updates the parameter values using the random movement updation of fireflies, which is represented in Eqn. (10).

$$P_s(n+1) = P_s(t) + \alpha\varepsilon_n \quad (10)$$

After refining and updating the parameter values, the fitness was determined to the upgraded parameters. Further, the parameter solution with higher fitness was selected for pretraining the DBN network. This fine-tuning mechanism continues until reaching the maximum iteration count. This iterative refining process not only enhances DBN training, but also ensures adaptiveness in the system. The working of the proposed algorithm is presented in pseudocode format in algorithm 1.

Algorithm: 1
Input: audio signals
Output: “Excellent”, “Poor”, or “Good”
Start {
Initialize the input dataset;
Pre-process the audio signals using adaptive filtering algorithm;
Design automatic speech recognition system:
Extract feature using MFCC;
Design RNN;
LSTM processing;
Model training;
Output: Spoken words;
Adaptive scoring algorithm:
Design DBN;
Initialize maximum iteration, population size and FA parameters;
Define objective function;
For each iteration:

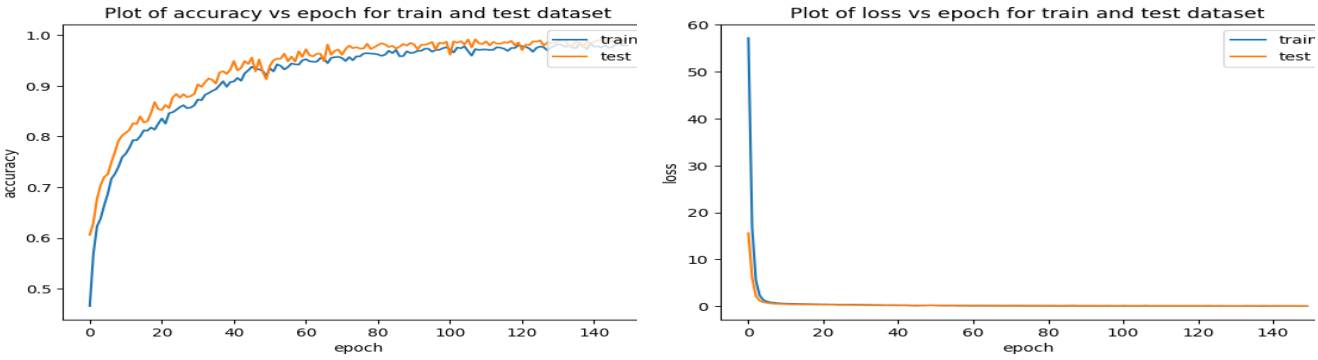
Determine fitness;
Update parameter solution;
Select parameter solution with high fitness;
Pretraining phase;
Determine probability value (scores);
if $P_v > 0.8$ "Excellent";
else if $0.8 > P_v > 0.6$ "Good";
else "Poor"
}End

4. RESULT ANALYSIS AND DISCUSSION

This study proposed a collaborative model for oral English evaluation by combining the efficiency of speech recognition technology and scoring algorithm. The presented framework was executed in Python language and the experimental computer environment is Windows10 operating system, CPU for Intel four core, memory 8 G, hard disk 1T. The results of the study were examined in terms of accuracy and evaluation error rate.

4.1 Model training performances

The training and testing performances of the proposed framework are assessed in this section in terms of accuracy and loss. Initially, the input database was split into training and validation subsets.



(a) (b)

Figure 2: Model training and validation performances: (a) accuracy, (b) loss

Here, 10% of the dataset was utilized for validation and the balance 90% of the dataset was employed for model training. The training and validation outcomes are assessed over increasing the epochs from 0 to 140. The batch size is 64/64, and the learning rate is fixed at 0.001. The training performance evaluates how effectively the proposed methodology learns the patterns and correlations in the training data for accurately evaluating the oral English proficiencies. The developed methodology achieved high training accuracy of 0.97 and minimum training loss of 0.04. This demonstrates that the designed methodology fits well into the training set. Figure 2 (a, b) depicts the training and testing performances of the proposed algorithm. The validation performances

evaluate the system's generalization ability. It determines how precisely the proposed model applies the identified patterns on the validation set. The proposed framework acquired high testing accuracy of 0.96, and lower loss of 0.07. This highlights that the developed algorithm obtained quickly learns the patterns in the data and evaluates the oral English proficiency accurately.

4.2 Performance analysis

In this section, the performance of the proposed strategy was evaluated over different voice databases (T1, T2, T3, and T4). T1 indicates the training model dataset containing 1208 natural spoken English voice recording.

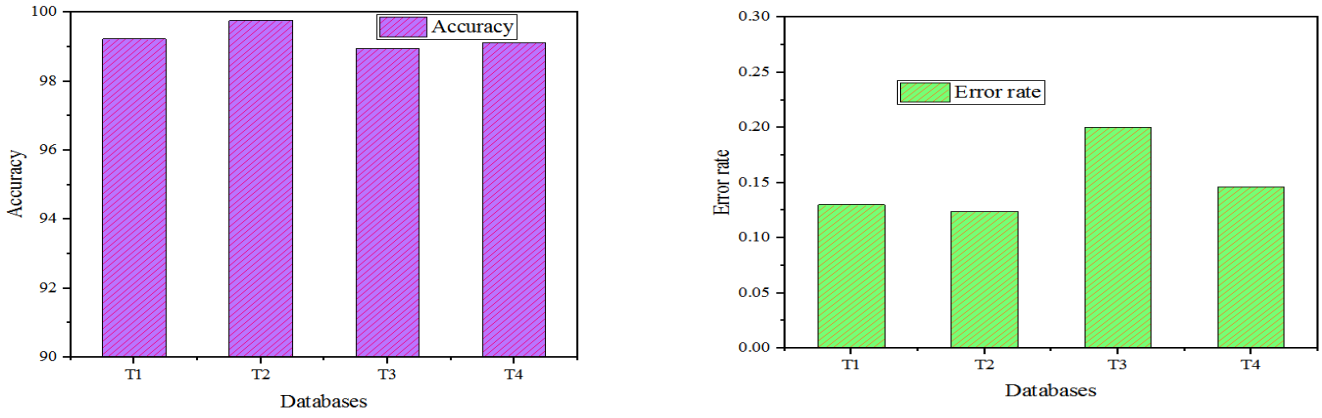


Figure 3: Performance of system over different databases

T2 defines the natural English speaking dataset comprising 122 Phoneme clusters. T3 indicates the professional English learner voice recording database, and T4 represents the database containing the amateur learner of oral English speech recordings. The performance parameters such as accuracy and error rate are evaluated by validating the proposed model for the above-mentioned databases. The developed framework achieved accuracy of 99.23%, 99.76%, 98.95%, and 99.12%, respectively for T1, T2, T3, and T4 databases. On the other hand, the designed approach obtained an error rate of 0.13%, 0.124%, 0.20%, and 0.146%, respectively, for T1, T2, T3, and T4 databases. Figure 3 presents the system performances over different datasets. This evaluation of the proposed methods' performances over diverse databases depicts that the proposed methodology achieved high accuracy of scoring and minimum evaluation error, highlighting its applicability over real-time environments.

4.3 Comparative evaluation

In this section, we compare the performances of the proposed strategy with the existing algorithms. The existing techniques used for performance comparison include Backpropagation neural network (BPNN), AdaBoost algorithm, Deep Neural Network (DNN), and Fuzzy-based speech recognition technology (FbSRT). Here, we evaluated and compared the outcome parameters such as accuracy and error rate with the above-stated conventional algorithms.

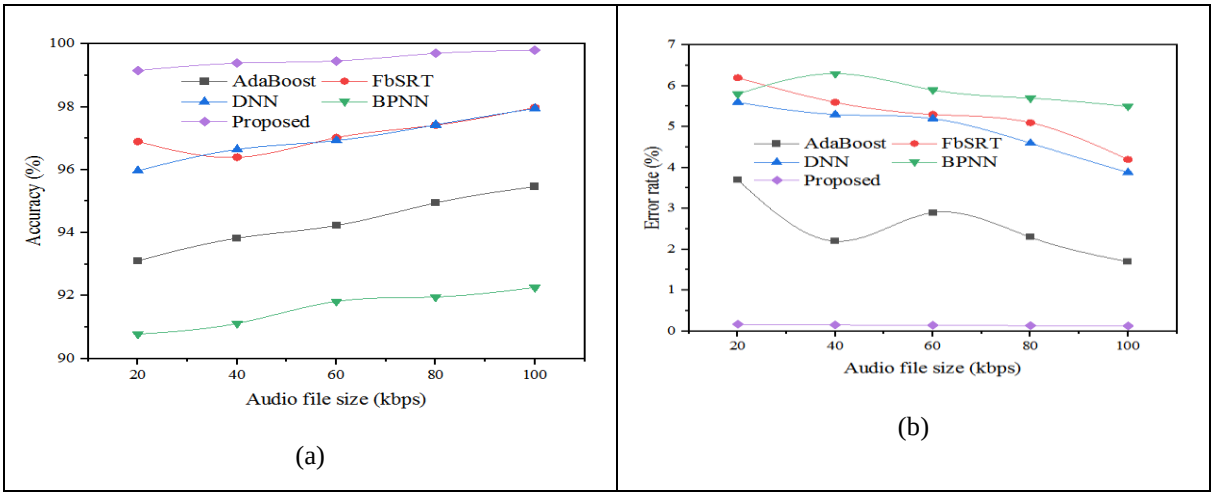


Figure 4: Comparative evaluation: (a) accuracy, (b) loss

The scoring accuracy defines how precisely the proposed methodology evaluates the oral English proficiency of the individuals. To validate that the designed algorithm achieved better accuracy, it is validated with conventional algorithms such as Adaboost, FbSRT, DNN, and BPNN. The accuracies of the existing models are evaluated by varying the audio file size from 20 to 100kbps. On average, these conventional techniques obtained an accuracy of 94.23%, 97.02%, 96.93%, and 91.82%, respectively, while the developed framework acquired a higher accuracy of 99.46%. This highlights that the proposed strategy accurately evaluates individuals' oral English proficiency levels. Consequently, we evaluated the evaluation error rate to determine the incorrect scoring and classification made by the system. The evaluation error determines the deviation between the actual and predicted scores. Figure 4 (a, b) compares accuracy and loss. The existing techniques, such as Adaboost, FbSRT, DNN, and BPNN, obtained an average error rate of 2.90%, 5.60%, 5.20%, and 5.90%, respectively, over increasing audio file sizes. However, the proposed framework obtained a minimum error rate of 0.147%, demonstrating that it accurately evaluates the oral English proficiency scores. Moreover, reducing the error rate over increasing audio file sizes depicts the model's scalability and adaptability, making it optimal for real-world applications.

6. CONCLUSION

This study developed an innovative oral English evaluation system using speech recognition technology and a scoring algorithm. The proposed framework utilized the corpus voice recording database, and it is pre-processed using the adaptive filter. An automatic speech recognition module was created using MFCC and RNN to transcribe the audio signals into textual characteristics. Further, an adaptive scoring algorithm was developed using the combination of Deep Belief Network and Firefly optimization algorithm to determine the scores of oral English. Moreover, this collaborative algorithm categorizes each individual’s oral English proficiency based on the scores. The developed work was executed in Python language and validated using four different voice databases. The experimental results suggest that the developed framework achieved a high accuracy of 99.46% and a minimum error rate of 0.147. Furthermore, we compared the existing works such as Adaboost, FbSRT, DNN, and BPNN to validate the proposed model's effectiveness. The comparative analysis demonstrated that the developed methodology performed better than the conventional models, making it more effective and optimal for real-time oral English evaluation.

REFERENCES

1. Wong, Shereen, and Melor Md Yunus. "A review of teachers’ perceptions of the use of social networking sites for the teaching and learning of English." Journal of Language Teaching and Research 14.2 (2023): 249-359.

2. Xie, Tianshi, Dallin Bailey, and Cheryl Seals. "SRA System Design: Using Deep Learning to Analyse Experimental data for Speech Researchers." 2022 International Conference on Theoretical and Applied Computer Science and Engineering (ICTASCE). IEEE, 2022.

3. Rodriguez-Ruiz, Jorge, Alvaro Alvarez-Delgado, and Patricia Caratozzolo. "Use of Natural Language Processing (NLP) Tools to Assess Digital Literacy Skills." 2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop. IEEE, 2021.

4. Wang, Yinping. "Detecting pronunciation errors in spoken English tests based on multi-feature fusion algorithm." Complexity 2021 (2021): 1-11.

5. Himeur, Yassine, et al. "AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives." Artificial Intelligence Review 56.6 (2023): 4929-5021.

6. Brena, Ramon F., et al. "Automated evaluation of foreign language speaking performance with machine learning." International Journal on Interactive Design and Manufacturing (IJIDeM) 15.2-3 (2021): 317-331.

7. Zhao, Chunrong. "An Innovative Strategy Towards Oral English Assessment Using Machine Learning, Data Mining, and Blockchain Techniques." (2023).

8. Wang, Qing, et al. "A Spatial-Temporal Graph Model for Pronunciation Feature Prediction of Chinese Poetry." IEEE Transactions on Neural Networks and Learning Systems (2022).

9. Li, Ping, Hua Zhang, and Sang-Bing Tsai. "Design of automatic scoring system for oral English test based on sequence matching and big data analysis." Discrete Dynamics in Nature and Society 2021 (2021): 1-10.
10. Wang, Xin. "Research on Open Oral English Scoring System Based on Neural Network." Computational Intelligence and Neuroscience 2022 (2022).
11. Jing, Wang. "Speech recognition sensors and artificial intelligence automatic evaluation application in English oral correction system." Measurement: Sensors (2024): 101070.
12. Sun, Xia. "Design and implementation of English speech scoring data system based on neural network algorithm." The International Conference on Cyber Security Intelligence and Analytics. Cham: Springer International Publishing, 2022.
13. Bao, Lei, and Jing Lv. "An Auxiliary Teaching System for Spoken English Based on Speech Recognition Technology." Scientific Programming 2022 (2022).