SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2024, 21(4), 833.
https://doi.org/10.62617/mcb833

Article

Dynamic relationship between oral English pronunciation standard and mental health monitored by biosensor

Xuan Zhou[1], Wei Jia[2,*], Cuiping Shi[2]

[1] School of Foreign Languages, Baoding University of Technology, Baoding 071000, China
[2] Basic Course Teaching Department, Baoding University of Technology, Baoding 071000, China
*** Corresponding author:** Wei Jia, 13833025494@163.com

CITATION

Zhou X, Jia W, Shi C. Dynamic relationship between oral English pronunciation standard and mental health monitored by biosensor. Molecular & Cellular Biomechanics.2024; 21(4): 833. https://doi.org/10.62617/mcb833

ARTICLE INFO

Received: 19 November 2024
Accepted: 11 December 2024
Available online: 19 December 2024

Abstract: Oral English pronunciation is an important feature of language ability, especially among non-native speakers, as good pronunciation has a direct impact on communication efficacy and social integration. However, the difficulties connected with attaining a high standard of oral English pronunciation can lead to severe stress, anxiety, and other mental health disorders. The purpose of the research is to establish a dynamic correlation between oral English pronunciation standards and mental health, as monitored through biosensor data. The research aims to explore how variations in speech accuracy and fluency during English pronunciation tasks can reflect underlying psychological states, such as stress, anxiety, and overall emotional well-being. The study proposed a novel Improved Flower Pollination-tuned Resilient Deep Neural Network (IFP-RDNN) in this article, to predict the oral English pronunciation rating using biosensors. Electroencephalography (EEG)records patterns of cerebral waves using electrodes applied to the head to assess the electrical impulses in the cerebellum called EEG signals was acquired during the listening state and with the audio signals utilized in stimuli, as well as the spoken audio obtained from the subject. The data processing used a median filter to remove noise from the audio data. Fast Fourier transform (FFT) is used to extract the features from the preprocessed data. It is measured by biomedical data, can be predicted with the help of an optimization technique which draws inspiration called IFP helps to optimize the parameters effectively by mimicking natural pollination processes; RDNN is employed with the optimized parameters; it can predict oral English pronunciation ratings. Experimental results reveal that the spoken audio confirms the improvement in pronunciation throughout the trials. In a comparative analysis, the suggested method is assessed with various evaluation measures, such as F1-score (88.9%), recall (91.60%), precision (89.80%), and accuracy (90.3%). The result demonstrated the IFP-RDNN method to predict the oral English pronunciation rating using biosensors. The findings indicate a significant connection between the quality of oral English pronunciation and mental health, with deviations from standard pronunciation being associated with increased stress and emotional suffering.

Keywords: oral English pronunciation; mental health; biosensor; improved flower pollination tuned resilient deep neural network (IFP-RDNN)

1. Introduction

The development of English pronunciation standards and the promotion of mental health using innovative technology like biosensors have received a lot of attention in recent years. The importance of proficient oral communication has grown since English established itself as a universal language [1]. Being proficient in English, especially in pronunciation is important for linguistic correctness and helps to demonstrate one's capacity for interaction and integration in multicultural environments. The spoken English pronunciation was necessary for social, professional, and academic settings [2]. English phonetics is complicated and varies greatly among linguistic origins, non-native speakers have difficulty for specific criteria. Traditional pronunciation techniques include machine learning (ML) and repetitive exercises, students might suffer from worry, self-doubt, and stress. It might hinder language learning and possibly impact students' mental health [3]. To incorporate mental health concerns frameworks for language acquisition. Biosensors have become a transformative tool, providing a way to track how learners react to stress and anxiety while they work on their pronunciation [4]. The knowledge might help establish a stress-reduction and mental health-promoting learning environment. Teachers and students could modify the rhythm and subject matter of sessions to provide a more pleasant experience of physiological effects on pronunciation practice. For instance, instructors might use specific strategies to reduce a student's tension if a biosensor identifies increased anxiety during specific sounds or phrases [5]. Biosensors aid in the development of self-awareness in students, allowing them to identify stress and self-regulation techniques. The biosensor-driven method offers insights into physiological components for learning instead of traditional evaluations, which frequently ignore the learner's mental state. It was possible to implement a comprehensive approach that gives equal weight to mental health and speech development, resulting in a well-rounded educational experience [6].

1.1. Linguistic hierarchy in language learning

Learners might feel less motivated and have poor self-perception when they can't satisfy the pronunciation requirements. The relationship between mental health and pronunciation of an interdisciplinary approach recognizes the psychological difficulties of language acquisition [7]. The sentiments could be intensified by the pressure from society to speak English, which frequently made the worse ideas by linguistic hierarchy. Long-term psychological repercussions of such events might include consequences on mental toughness and self-worth. It was proactively addressed by integrating biosensors into language teaching. Additionally, biosensors enable language learning programs to be tailored to every learner's unique requirement and speed [8]. By customizing feedback based on biosensor data, students can progress at a pace that suits them, avoiding feelings of being overburdened. The strain of challenging sounds or dialects was lessened and a sense of accomplishment was fostered by customization. It is especially helpful in group situations when students might experience more pressure to perform on pace with their classmates [9]. Teachers might create a supportive atmosphere that prioritizes mental health as linguistic accuracy by using biosensors to detect student focus. Using biosensors in pronunciation instruction promotes more compassionate teaching strategies. Teachers might modify the methods to encourage rather pupils to learn more about emotional and psychological states. It fosters the learner's mental health by acknowledging their difficulties and promoting self-compassion with pronunciation [10]. Instructors might be trained to identify symptoms of stress and anxiety that help to modify the encouragement or feedback of individuals. The positive feedback encourages students to take chances and make errors as a vital component of language acquisition, without worrying about criticism or failure [11]. Pronunciation learning could be impacted by differences in learners' ability, motivation, and exposure to English. Oral pronunciation frequently emphasizes correctness and fluency, but it might exceed how language acquisition affects the mental health and growth of speaking confidence. Biosensors offer real-time physiological data to monitor stress and anxiety levels during pronunciation practice, providing a personalized language learning approach. This technology aids in improving pronunciation and promotes mental well-being by creating a supportive, tailored learning environment [12].

The study's aim is to develop a novel Improved Flower Pollination-tuned Resilient Deep Neural Network (IFP-RDNN) to predict the oral English pronunciation rating using biosensors. The research intends to explore how variations in speech accuracy and fluency during English pronunciation tasks can reflect underlying psychological states, such as stress, anxiety, and overall emotional well-being.

Contribution of the study

- The study's aim is to develop a novel Improved Flower Pollination-tuned Resilient Deep Neural Network (IFP-RDNN) to predict the oral English pronunciation rating using biosensors.
- This research establishes the dynamic relationship between oral English pronunciation standards and mental health, as monitored through biosensor data.
- The data processing used a median filter to remove noise from audio data. FFT was used to extract the features from the preprocessed data.
- The research intends to explore how variations in speech accuracy and fluency during English pronunciation tasks can reflect underlying psychological states, such as stress, anxiety, and overall emotional well-being.

The following six sections comprise the overall article. Related work is given in Phase II, the study's methodology was presented in Phase III, Phase IV represents the result, Phase V provide discussion, and the research is concluded in Phase VI.

2. Related works

Speaking and listening were greatly impacted by Ramzan and Javaid [13] for proper pronunciation. Language acquisition and performance were influenced by psychological variables. The students had significant difficulties with some monophthongs, particularly diphthongs. The psychological elements influence proper pronunciation. The experimental outcome demonstrated the significant barriers to acquiring proper pronunciation.

The Deep Belief Network (DBN) utilized for voice identification in oral English instruction by utilizing deep learning methods and speech recognition techniques by Wang [14]. Several typical issues with oral English instruction included the ineffectiveness of the teaching technique and the low spoken English proficiency of the students. Furthermore, it develops a multi-parameter assessment model to gauge college students' spoken English pronunciation proficiency. The experimental outcome demonstrated the speech rating trials of pronunciation.

Teachers' pronunciations were inconsistent for students and difficult to learn proper pronunciations, spoken English was extremely passive in typical classroom settings by Jing et al. [15]. Enhancing students' interest in oral English, encouraging them to talk, and helping them to understand English communication. The experimental outcome demonstrated the export language assessment.

Li and Huang [16] investigated the detection of pronunciation errors and quality evaluation, for evaluating learners' English pronunciation quality. English has gained more attention from the worldwide language, and learning the language orally was essential. The result findings demonstrated the effectiveness of speech quality assessment and mistake detection in pronunciation.

The restricted English learning environment and teaching settings make it difficult for English language learners to acquire spoken English Geng [17]. Language training and learning have constituted a new era due to advancements in artificial intelligence and instructional techniques. Language learning relies heavily on voice recognition and evaluation technology, with speech recognition as a fundamental building block. The experimental outcome demonstrated the impact of multimedia instruction in English.

The international Standard English teaching approach known as phonics by Zhou et al. [18] created a phonetic matching system by pronunciation. English pronunciation is labeled as the International Phonetic Alphabet (IPA), which raises the spoken language to worldwide standards. The experimental outcome demonstrated the use of new media technologies in spoken English instruction.

The fundamental aspect of language used in oral communication is known as pronunciation. Sharma [19] emphasized how useful it was for teaching segmental language and transcendental speech elements. The primary purpose of language is to facilitate speech-based communication. Sequences of regional and transcendental speech feature pronunciations that intend the speech. The experimental outcome demonstrated the pronunciation issues.

Wang and Zhao [20] investigated the automatic techniques of recognizing and assessing spoken English in English education by computer-assisted technologies. Learner's preference for computer-assisted spoken English instruction was growing. Learners might improve their pronunciation by using computer-assisted mixed technology to evaluate and improve their spoken pronunciation. The experimental outcome demonstrated the criticism of pronunciation.

Spoken English learners have five dimensions of variations, such as learning capacity, grammar structure, lexical utilization, pronunciation, and disparities in communicative circumstances. Liang and Ye [21] examined how learners' spoken English proficiency affected gender inequalities. The experimental outcome demonstrated the possible gender disparities in language acquisition.

The foundations of creating guidance tasks in Project-Based Learning (PBL) include the phonological and phonetic components of English pronunciation by Iskandar et al. [22]. The PBL might be used to guide the tasks and it enhances the educational achievement in English pronunciation and develops self-directed learning abilities. The experimental outcome demonstrated the English pronunciation learning guidance.

Student's pronunciation was typically improved by teachers using their subjective opinion while teaching oral English. The reconstruction of acoustics and language aspects of speech recordings could be used to determine students' oral pronunciation features Wu and Sangaiah [23]. However, the task ensures the integration of multimodal sentences. It utilized an improved spatial segmentation network for English voice recognition. The experimental outcome demonstrated the efficient processing of speech augmentation.

The history of global communication constitutes more demands on students' speaking and listening skills in English by Gui [24]. The issues can be resolved by raising students' impact on pronunciation while speaking. Students frequently struggle with weak articulation, inaccurate intonation, and lack of proficiency in both pronunciation and intonation. The experimental outcome demonstrated the intonation of English communication.

Through the use of Mobile Biosensor Networks called MBN-QE, Huang [25] suggested a realistic, ethical, and medically relevant evaluation technique for political education students in Chinese universities. Learning data analysis, automated testing, student rewards, and identity control were all applications for MBNs. The results implied that MBNs could be a useful resource for students studying political science.

The usage of wearable technology for managing stress, mental health concerns, and physiological parameter monitoring was described by Wu et al. [26]. Sensors could interpret changes in emotion and exhaustion levels, offer guided training functions, and make the first diagnosis. Applications for enhanced memory retention and stress relief could also be linked with such devices. Wearable technology, sensor kinds, data reception techniques, processing precision, and application dependability were some of the variables that might lead to data inaccuracies. The accuracy of future medical systems should be established and clinically assessed.

Problem statement

The relationship between English pronunciation standards and mental health knowledge enhances how pronunciation issues might impact people's psychological health. Students with poor pronunciation might suffer from anxiety, low self-esteem, and social rejection in a variety of social and educational settings, which aggravate mental health issues. In contrast, tension and performance anxiety constitute the pressure to adhere to certain pronunciation standards, particularly for non-native speakers. The impact of mental health on language learning and voice clarity was frequently disregarded. Since emotional strain could impair vocal and cognitive function. The pronunciation might increase self-esteem, lower anxiety, and improve psychological well-being.

3. Methodology

The Improved Flower Pollination-tuned Resilient Deep Neural Network (IFP-RDNN) was used to predict the oral English pronunciation rating using biosensors. Data pre-processing is used to preprocess the raw data. The median filtering is used to remove the noise from audio data. FFT was used to extract the features from the preprocessed data. IFP helps to optimize the parameters effectively by mimicking natural pollination processes and RDNN was employed to predict the oral English pronunciation ratings. **Figure 1** represents the overall paper flow.

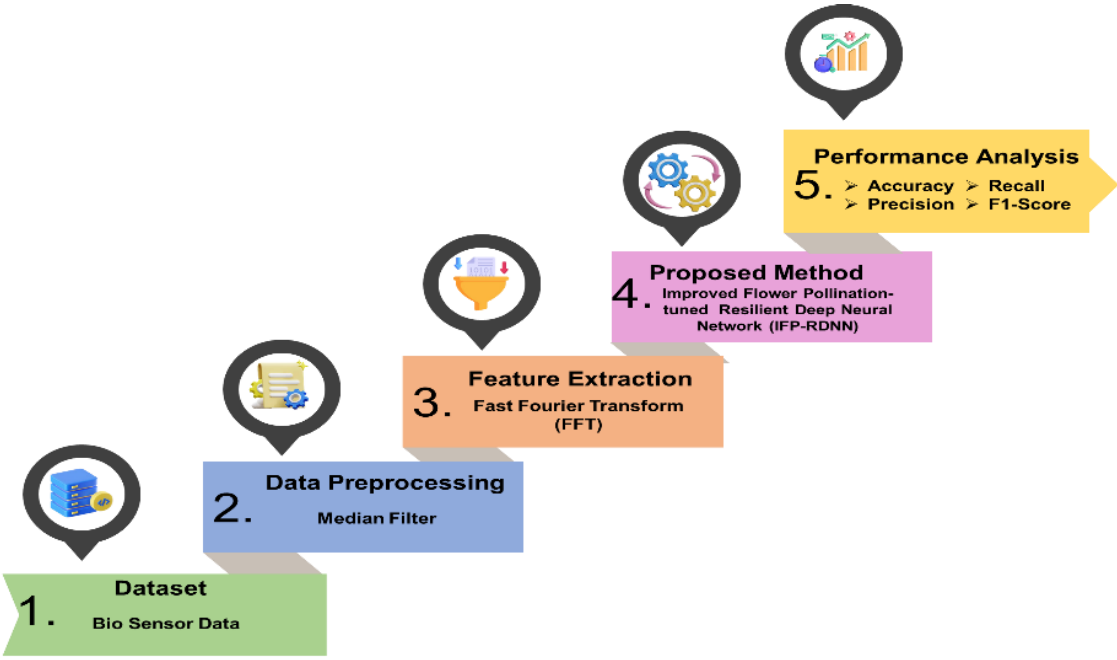

Figure 1. Overall flow for the research.

3.1. Dataset

The dataset was gathered from open-source platform (https://www.kaggle.com/datasets/ziya07/biosensor-oral-english-pronunciation-data/data). Using biosensor data from participants' pronunciation of English tasks and physiological indicators like electroencephalogram (EEG), heart rate, and GSR called Galvanic Skin Response, the dataset examines the dynamic relationship between oral English pronouncing quality and mental health. To examine variability in speech performance, the dataset integrates psychological metrics such as level of anxiety and Positive and Negative Affect Schedule (PANAS) scores with speech audio elements. It is a useful tool for language competence and emotional wellness research since it concentrates on stress levels and pronunciation scores, offering knowledge about how these aspects may interact with psychological conditions like stress and anxiety.

3.2. Data pre-processing

One of the most important steps in the evaluation of data and ML is data preprocessing, which turns raw data into a format that can be understood. Encoding, classifying variables, removing duplicates, scaling or normalizing data, and handling missing values are among the tasks that are involved. Preprocessing enhances the model's functionality. To eliminate noise from audio data, the median filter was employed in the data.

Median filter (MF)

The MF might be used to smooth variances in speech data, which makes it simpler to evaluate the quality and correctness of pronunciation in spoken English standards. As a member of the nonlinear filter family, the MF was a widely used image processing technique. Within a sliding window centered at each pixel point to create a smooth image, the procedure calculates the median value. It effectively eliminates high-energy sound, such as isolated bright dark areas, while maintaining edge data and minimizing extreme blurring. The brightness of a pixel at a certain point(y, x) in the image after median filtering was expressed in Equation (1) as follows:

$$H(y, x) = median\,(H(y', x')) \tag{1}$$

Coordinates (y', x')denote the domain's pixel locations, and "median" refers to the value of the output pixel that was produced when the filter's pixels were sorted by size. The use of median filtering allowed for the elimination of high-frequency visual components while drawing emphasis on low-frequency. Human senses are susceptible to image change, which increases the arbitrary visual effects of hostile situations and the invisibility of negative effects.

3.3. Feature extraction

Feature extraction is used in ML and data analysis to minimize the dimensionality of data. It cracks unprocessed data into a collection of useful characteristics that make analysis easier while preserving details. It improves the efficiency and interpretability of the model, concentrating on important characteristics. Feature extraction, including FFT used to extract the features from the preprocessed data.

FFT

FFT is used to evaluate the correctness and fluency of pronunciation in spoken English by analyzing the frequency components and speech sounds. Speech patterns frequently correspond with mental health by employing speech analysis to identify emotional states and cognitive problems. The process of breaking down a function or signal over time into a component of frequency function is known as FFT. The perception of the Fourier transform was essential for image processing, speech and communication, signal processing, and many other fields. Equation (2) describes the FFT mathematical form.

$$W(\zeta) = \int_{-\infty}^{\infty} w(s)\, f^{-2\pi j\zeta s} ds \tag{2}$$

Here, $W(\zeta) \in D$as Lebesgue integrals, andζ indicates the frequency. For practical use, FFT can be used to implement quickly. The discrete Fourier transform (DFT) was calculated by FFT, which yields the same outcome as evaluating the DFT directly. The primary distinction of DFT was significantly quicker and less precise. Equation (3) represents the DFT mathematical form.

$$W_l = \sum_{m=0}^{M-1} w_m f^{-\frac{j\pi m}{M}}, (l = 0, .., M-1) \quad (3)$$

Here, $w_0, .., and\ w_{M-1}$ represent thecomplicated integers. The FFT was used to translate and evaluate the spinning machine's data into the frequency plane from several experiments.

3.4. Improved flower pollination-tuned resilient deep neural network (IFP-RDNN)

A hybrid of Improved Flower Pollination with a Resilient Deep Neural Network offers a novel way to improve mental health evaluation and English pronunciation standards. Adapted from the pollination process in nature, the FPA optimizes parameters inside an RDNN structure, enabling enhanced performance when challenging and it ensures non-linear tasks, such as spoken English pronunciation. Due to RDNN's resilience while processing a variety of linguistic inputs, this hybrid approach achieves better accuracy in facilitating efficient pronunciation and phonetic clarity analysis. By effectively exploring and using the solution space, the FPA optimizes the weights and biases of the neural network for pronunciation evaluation and guarantees accurate feedback and adaptive enhancements. The IFP-RDNN approach incorporates mental health concerns to help students develop emotional resilience. Pronunciation instruction causes tension and anxiety in many language learners, which might affect their ability to learn. The exploration of participation in emotional reactions during pronunciation exercises integrates the evaluations of mental health. When paired with FPA's search capabilities, the layered architecture of RDNN detects trends in pronunciation accuracy as mental health. This hybrid paradigm combines linguistic and psychological elements to create a learning atmosphere where pronunciation is refined to preserve mental health.

3.4.1. Resilient deep neural network (RDNN)

RDNN approach seeks to improve both language acquisition and psychological wellness. RDNN modifies problems, such as inconsistent accents or noisy data, giving regular feedback for better pronunciation. Such systems might track student stress, motivation, and confidence by including mental health evaluation and fostering a positive learning environment. It enables the model to concentrate on both emotional resilience and verbal precision. The Gaussian mixture model (GMM) decreases the duplicate features, which shrinks the dimension of the DNN system and increases the recognition rate. Consequently, DNN's structure was modified by using GMM. The feedforward neural network (FNN) with several hidden layers was completely linked with resilient DNN. To generate the GMM probability density function, the Gaussian probability density function of each component was weighted for specific demands that were represented in Equation (4) as follows.

$$O(w|\lambda) = \sum_{j=1}^{N} x_j M_j(w) \quad (4)$$

Here, x represents the eigenvector's dimension as $M_j(w), j = 1,2,..,M$. N as subgroup probability. Every subgroup probability constitutes Gaussian probability distribution in $x_j, j = 1,2,...,N$ dimensions denoted as $M_j(w)$ were expressed in Equation (5).

$$M_j(w) = \frac{1}{(2\pi)^{C/2} |\sum_j|^{1/2}} \exp\left\{-\frac{1}{2}(w - \mu_j)^S \sum\nolimits_j^{-1} (w - \mu_j)\right\} \tag{5}$$

Here, $\mu_j$represents the mean vector, $\Sigma_j$represents the covariance matrix and S refers to the number of feature vectors. The eigenvectors are associated with the parameters that were expressed in Equation (6).

$$\lambda = \left\{x_j, \mu_j, \sum\nolimits_j n\right\}, j = 1,2,...,N \tag{6}$$

The training data's feature vector sequence as $W = \{w_s\}, s = 1,2,...,$ S and prospect probability were expressed in Equation (7).

$$O(W|\lambda) = \prod_{s=1}^{S} O(w_s|\lambda) \tag{7}$$

The Gaussian density function was expressed in Equation (8) as follows:

$$O(W|\lambda) = \sum_{s=1}^{S} 1g \left\{\sum_{j=1}^{N} x_j M_j(w_s, \mu_j, \sum\nolimits_j\right\} \tag{8}$$

Maximum Likelihood Estimation (MLE) parameter was used to estimate the technique that was utilized in the GMM system. The model is constantly modified into the prospect probability, $O(W|\lambda)$ achieves its concentration. Following the definition of the model parameters, the probability at maximum likelihood as $\lambda_j$was expressed in Equation (9).

$$\lambda_j = argmax\ O(W|\lambda) \tag{9}$$

DNN's input specifications were specified in equation (10) as follows.

$$q = \sum\nolimits_j x_j \times \frac{\mu_j - \mu_{j,ubm}}{(diag\{\Sigma_j\}^{1/2}} \tag{10}$$

Here, $\mu_{j,ubm}$represent the universal background model's (UBM) as the mean vector. The initial hidden layer's activation vector of the eigenvector was expressed in Equation (11).

$$\vec{g}^{(1)} = \sigma\left(X^{(1)S}\vec{w} + \vec{a}^{(1)}\right) \tag{11}$$

Here, $X^{(1)S}$represents the weight distribution of the first invisible layer in reverse, with dimensions $J \times M_1$ times. Where$\vec{a}^{(1)}$ a1 denotes the offset vector, and σ is a function that activates the buried layer. The vectors of activation $\vec{g}^{(j)}$were used to calculate the j second invisible layer expressed in Equation (12).

$$\vec{g}^{(j)}(\vec{w}) = \sigma(X^{(j)S}\vec{g}^{(j-1)}(\vec{w}) + \vec{a}^{(j-1)} \tag{12}$$

Here, $M_j$indicates the number of neurons in $the\ jth$ invisible layer, $X^{(j)S}$represents the weight distribution reverse of the jth layer, which has dimensions of $M_{j-1} \times M_j$. Where $\vec{a}^{(j)}$ represents the offset vectors of matching dimension. The ReLU function utilized as a hidden layerfunction of activation because of its strong performance in resilient DNN categorization and recognition assignments were expressed in Equation (13).

$$\sigma(b) = \begin{cases} b\ if\ b > 0 \\ 0 if\ b \leq 0 \end{cases} \tag{13}$$

The SoftMax function is used by the DNN's output layer to finish the output category. It ultimately achieves categorization and identification. The expression for the SoftMax function is expressed in Equation (14).

$$t(y_l) = \frac{f^{y_l}}{\sum_{i=1}^{I} f^{y_i}} \tag{14}$$

Here, y_l represents the vector's output layer's dimensions. The loss function provided by the SoftMax regression technique was expressed in equation (15).

$$I(x) = -\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{c} 1\{z^{(j)} = c\} log \frac{f^{x_i^S y^{(j)}}}{\sum_{k=1}^{c} f^{x_k^S y^{(i)}}} \tag{15}$$

Here, $log \frac{f^{x_i^S y^{(j)}}}{\sum_{k=1}^{c} f^{x_k^S y^{(i)}}}$represents the logarithmic values for the softmax function, with $z^{(j)}$ as the predicted part. $1\{\times\}$ represents a characteristic function, where $z^{(j)}$ indicates the predicted label or true label. Where$z^{(j)} = c$ function returns as 1 otherwise it returns as 0.

3.4.2. Improved flower pollination (IFP)

The Improved Flower Pollination algorithm simulates the interaction of pollens with flowers to solve complicated problems. It was inspired by the natural pollination process in flowers. In the context of spoken English pronunciation, it replicates the process of choosing the most pertinent linguistic elements and adjusting to various speech patterns, thereby optimizing speech models to better match the pronunciation standards. The IFP approach improves the flexibility and personalization of the process. The communication models of people with speech impairments or mental health disorders impact the speech and language for the evaluations of treatments.

Initial population generation:

Chaos is defined as a nonlinear dynamical system's unpredictable, erratic, and chaotic behavior in a starting state that is highly sensitive in a small space. A slight alteration at the beginning condition might result in a change of output. The link between a chaotic system's present state and its upcoming state was described by a separate logistic operation to construct the chaotic map. The initial population generation of IFP has resulted in the greatest improvement in convergence rate. The circle map outperforms chaotic maps to improve the modified multi-verse optimizer's rate of convergence. The circular map's chaotic sequences' have the highest degree of dispersion accounts for its efficiency. The non-recurrence probability of chaotic sequences increases with their dispersion degree. Therefore, the circle map utilized for population initialization could yield superior starting population generation resolutions that disperse the ideal solution. The usual random technique of initializing the population will be replaced with a circular map.

Frog leap search:

The local search approach used by the FP discourages information exchange among the superior solutions, which causes the IFP to converge slowly. The shuffled frog-leaping algorithm (SFLA) serves as the inspiration for the idea of reciprocal cooperation. The frogs' foraging habits served as the basis for the creation of the SFLA. The search space was used to start an initial population of frogs. It contains information about the quantity of food in each frog's location. Memeplexes were created from the initial population. The frog who snatches the most food during the procedure was assigned to the initial group. The process was continued until the frog ranks were assigned. Likewise, the frogs of the second and third ranks were assigned to the second and third groups, respectively. The initial group receives the frog rank. The frogs with the lowest and highest fitness for each memeplex were represented as $\vec{W}_{xj}^{(s)}$ and $\vec{W}_{aj}^{(s)}$. The frog's least fit will be updated according to the information that constitutes the best frog exchange represented in Equations (16) and (17).

$$\vec{W}_{xj}^{(s+1)} = \vec{W}_{xj}^{(s)} + ramd\left(\vec{W}_{aj}^{(s)} - \vec{W}_{wj}^{(s)}\right) \tag{16}$$

$$t_{max} \geq rand\left(\vec{W}_{aj}^{(s)} - \vec{W}_{wj}^{(s)}\right) \geq -t_{max} \tag{17}$$

Here, s denotes the current iteration, $rand \in (0,1)$ and t_{max} represents the maximum step size. The modified frog's fitness was assessed. The frogs were arbitrarily relocated to a different area of the swamp if their fitness didn't improve. The procedure was carried out repeatedly until the Memeplexes predetermined the maximum iteration number until it reached. Entire Memeplexes were combined. The shuffling procedure divides the population among groups and the local search of every Memeplexes entails the halting condition until it is satisfied.

Momentum coefficient:

Ensuring the optimization system's worldwide search exhibits dynamic behavior was essential for increasing the convergence rate. It was better to conduct more inquiry in the initial stages. The solutions were closer to the global optimum, which ensures the progressive transition to become more localized corruption in the final stages. IFP constitutes the search procedure. It takes a long time because of the pointless investigation at the end. The suggested IFP coordinates the Lévy flight to overcome the constraint were expressed in Equations (18) and (19).

$$\vec{W}_j^{s+1} = \vec{W}_j^{s} + xK\left(\vec{W}_j^{s} - h^*\right) \tag{18}$$

$$x = x_{min}\left(1 + \frac{x_{max}}{\sqrt{s}} \times tang\left(\frac{h_1}{h_*}\right)\right) \tag{19}$$

Here, x represents the inertia weight, with x_{max} being the maximum allowed inertia weight and x_{min} being the minimum. The ideal global fitness is indicated as h^* to obtain the highest fitness value attained by the original population the Lévy distribution stands for the current generation number and the hyperbolic tangent is indicated as $\tanh(\times)$ the inertia weight might be used to dynamically modify the worldwide search's step size in IFP. The iteration ensures optimal fitness value as better to ensure the outcome of inertia weight that eventually drops. The values of x_{max} *and* x_{min} are set from 1 and 10. Algorithm 1 shows the IFP-RDNN pseudocode.

Algorithm 1 Process of IFP-RDNN

```
data = load_data("oral_english_mental_health_data.csv")
preprocessed_data = preprocess_data(data)
train_data, test_data = split_data(preprocessed_data)
defcreate_resilient_dnn(input_size, hidden_layers, output_size):
    model = Sequential()
model.add(Dense(units = hidden_layers[0], input_dim = input_size, activation = 'relu'))
forlayer_sizein hidden_layers[1:]:
model.add(Dense(units = layer_size, activation = 'relu'))
model.add(Dense(units = output_size, activation = 'softmax'))
model.compile(optimizer = Adam(), loss = 'categorical_crossentropy', metrics = ['accuracy'])
returnmodel
defimproved_flower_pollination(population_size, max_iterations):
population = initialize_population(population_size)
foriterationinrange(max_iterations):
forindividualsinthepopulation:
ifisGlobalPollination():
individual = update_global_pollination(individual)
Else:
individual = update_local_pollination(individual)
fitness_scores = evaluate_fitness(population)
population = select_best_individuals(population, fitness_scores)
returnbest_solution_from_population(population)
input_size = len(train_data.features)
hidden_layers = [64, 32]
output_size = len(train_data.labels)
dnn_model = create_resilient_dnn(input_size, hidden_layers, output_size)
best_hyperparameters = improved_flower_pollination(population_size = 50, max_iterations = 100)
dnn_model.update_hyperparameters(best_hyperparameters)
dnn_model.fit(train_data.features, train_data.labels, epochs = 50, batch_size = 32)
test_predictions = dnn_model.predict(test_data.features)
evaluate_model(test_predictions, test_data.labels)
pronunciation_quality = analyze_pronunciation(test_predictions)
mental_health_correlation = correlate_pronunciation_with_mental_health(test_data.mental_health)
print("PronunciationQuality: ", pronunciation_quality)
print("MentalHealth Correlation: ", mental_health_correlation)
```

4. Experimental results

Using TensorFlow to complete the recommended task. Where Python software was installed for the procedure to be completed and the experiment was run on a 64-bit version of Windows 10. The Intel(R) Core (TM) i7-7770hq 2.8 GHz CPU and 8 GB of RAM are installed. The configuration information's of the experiment is presented in **Table 1**.

Table 1. Configuration details for experiment execution.

Component	Details
Operating system	64-bit version of Windows 10
Software installed	Python
Processor	Intel(R) Core (TM) i7-7770HQ, 2.8 GHz CPU
RAM	8 GB

4.1. Confusion matrix

A confusion matrix is a technique for assessing the performance of a classification model. It shows the numbers of real positive results (properly anticipated positives), genuine negative results (properly forecasted negatives), and false positive results (inaccurately predicted positives) for each result. This matrix offers information on the model's overall efficiency. It also helps to pinpoint the areas of inaccuracy and areas for model improvement. For unbalanced datasets, where basic accuracy might be deceptive which was helpful. It demonstrates how many events were correctly and incorrectly predicted across various categories. In this matrix, "True labels" and "Predicted labels" are categories. Most instances were properly predicted by the model; however, a few individuals were low predicted. It helps to determine how effectively the model operates and where it might be improved. **Figure 2** represents the confusion matrix.

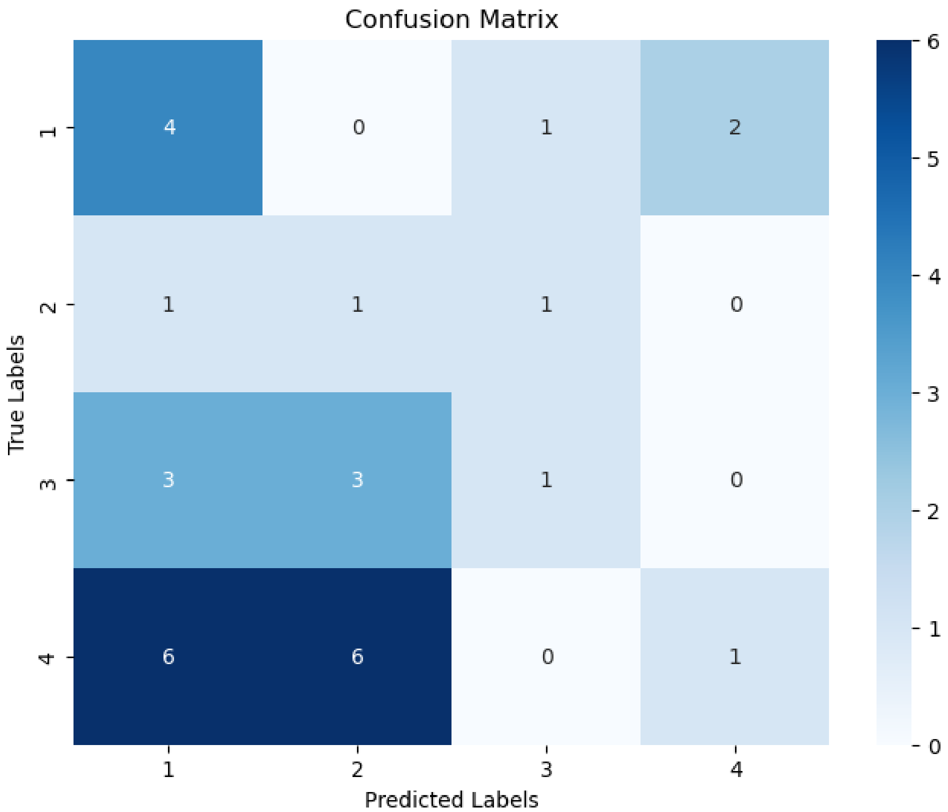

Figure 2. Confusion matrix for proposed method.

4.2. Accuracy and loss

The ratio of accurate forecasts to all predictions used to determine the model's accuracy can anticipate events properly. It expresses a broad idea of the model's performance in every class. Loss measures how far the model's predictions deviate from reality by comparing the expected values with the actual results. Since it reduces the difference between the real and projected values, a lesser loss suggests that the model was operating properly. A more precise assessment of the model's learning quality and predicted accuracy is provided by loss, whereas accuracy is a broader metric. **Figure 3** represents the accuracy and loss.

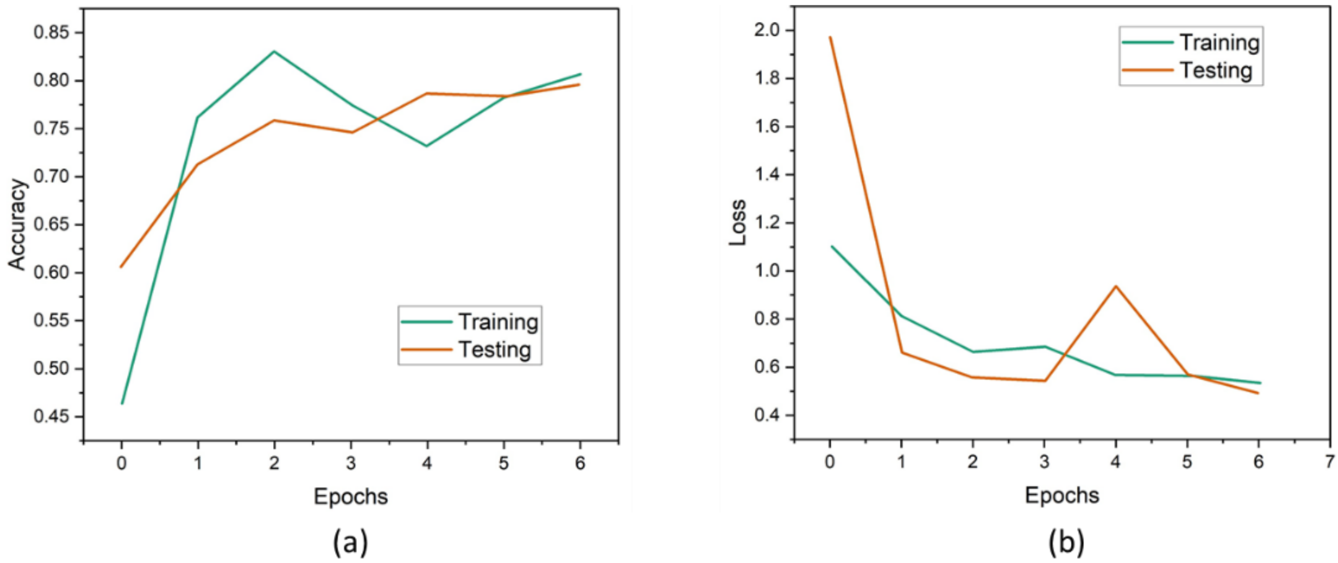

Figure 3. Graphical outcome of **(a)** accuracy and **(b)** loss.

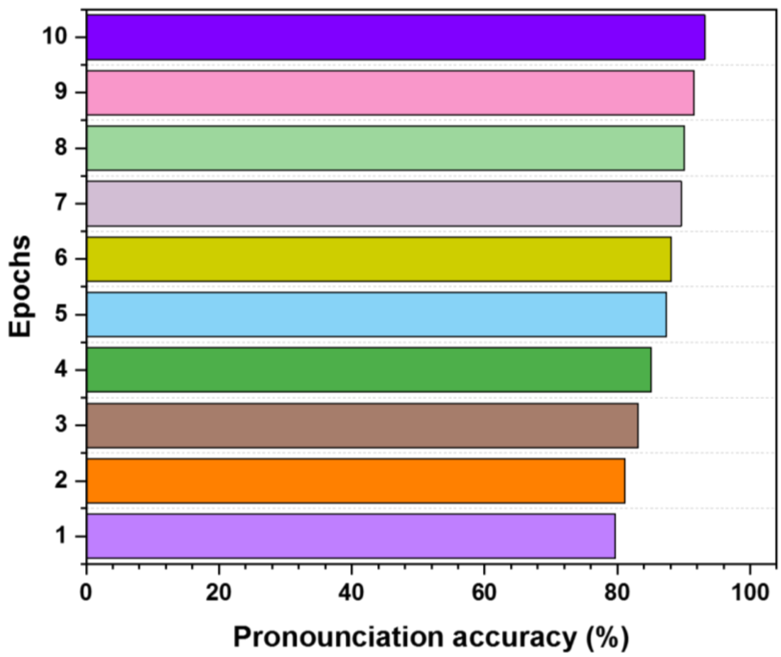

Figure 4. The pronunciation accuracy outcome values.

Pronunciation accuracy can articulate sounds, stress, intonation, and rhythm correctly in speech while adhering to the norms of the target or standard language. It is essential for communication since it guarantees speakers' comprehension and clarity. Accurate pronunciation makes a student more understandable and self-assured while speaking a language. Accuracy was frequently attained by voice technique and sound production refinement through practice and feedback. **Figure 4** illustrates the pronunciation accuracy.

The proposed strategy was assessed and its effectiveness was calculated using the following indicators: Recall (%), Accuracy (%), precision (%), and F1-score (%). An efficiency comparison across the proposed strategy and other traditional approaches was also presented. The traditional methods include a Resilient Deep Neural Network (RDNN).

Evaluating the model's accuracy by calculating the ratio of successfully expected to total occurrences, accuracy offers a robust assessment of the system's efficiency. Compared to traditional methods like RDNN have an accuracy of 86.5% and the proposed IFP-RDNN attains an accuracy level of 90.3%. A model's precision level indicates how accurately it anticipated results. The assessment is the proportion of precisely predicted positive results to the total expected benefits. Compared to traditional methods RDNN has a precision of 85.9%, and the proposed IFP-RDNN attains a precision level of 89.80%. **Figure 5** and **Table 2** illustrate an evaluation of accuracy and precision in comparison between suggested and traditional methods. The proposed method provided superior results in predicting the oral English pronunciation rating using biosensors.

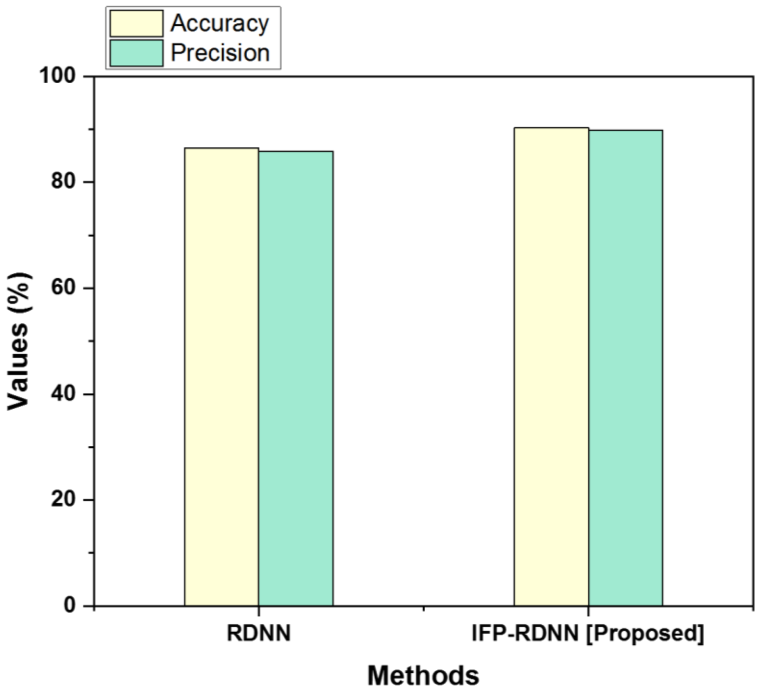

Figure 5. Results of model accuracy and precision performance analysis.

Table 2. Result parameters.

Methods	Accuracy (%)	F1- Score (%)	Recall (%)	Precision (%)
RDNN	86.5	85.9	85	85.4
IFP-RDNN[Proposed]	90.3	89.8	91.60	88.9

Recall is a statistic that assesses a model's capacity to locate all pertinent instances of a class. It assesses the model's accuracy in identifying every pertinent among the total number of real positives. Compared to traditional methods like RDNN with a recall of 85%, the proposed IFP-RDNN attained a recall level of 91.60%. F1-score is a statistic that was used to assess how well a classification model is performed. Compared to traditional methods, like RDNN with an F1-score of 85.4%, the proposed IFP-RDNN attained an F1-score level of 88.95%. **Figure 6** and **Table 2** present an evaluation of the F1-score in comparison between suggested and traditional methods. The proposed method provided better results for predicting oral English pronunciation rating through biosensors.

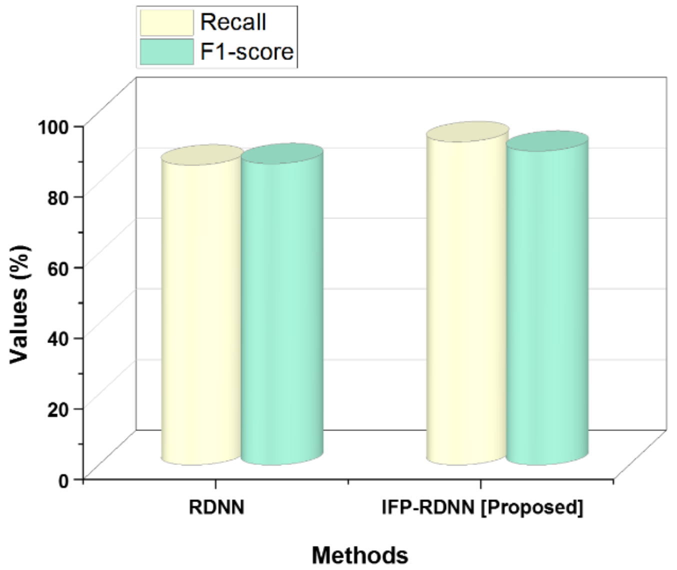

Figure 6. Result values for recall performance and F1-Score evaluation.

5. Discussions

The model's complexity of the FDNN method in oral English pronunciation standards and mental health might demand more computing power, which might make it less appropriate for real-time applications and settings with constrained computational resources. Furthermore, mental health was impacted by several intricate and multifaceted elements other than language proficiency; it might be difficult to measure the relationship between oral pronunciation standards and mental health. The standardization of spoken English pronunciation might be impacted by participants' cultural and linguistic diversity, making the application of global standards challenging. To overcome this, the IFP-RDNN model uses the advantages of the IFP for accurate parameter effectiveness, which helps it successfully navigate the challenges of mental health and oral English pronunciation standards. Because it is less computationally complex than the FDNN method, it is more suited for applications that operate in real-time, even in resource-constrained environments. With the ability to handle the complex nature of mental health, RDNN adjusts to a range of mental states, maintaining robustness. The approach offers more accurate and personalized knowledge of how pronunciation impacts mental health by integrating biosensor data. Furthermore, the problem of particular cultural and linguistic variation is addressed by IFP's natural optimization method, providing a more flexible option for language students from across the world. All things considered, the IFP-RDNN model offers a thorough, scalable, and effective way to forecast pronunciation quality by taking emotional well-being into account.

6. Conclusions

The Improved Flower Pollination-tuned Resilient Deep Neural Network (IFP-RDNN) was used to predict the oral English pronunciation rating using biosensors. The research is to explore how variations in speech accuracy and fluency during English pronunciation tasks can reflect underlying psychological states, such as stress, anxiety, and overall emotional well-being. The purpose of the research is to establish the dynamic correlation between oral English pronunciation standards and mental health, as monitored through biosensor data. EEG signals were acquired during the listening state with the audio signals utilized in stimuli, as the spoken audio obtained from the subject. The median filter removes noise about the recorded information was employed in the data processing. FFT is used to extract the features from the preprocessed data. The proposed method is implemented using Python software. Experimental results reveal that the spoken audio confirms the improvement in pronunciation throughout the trials. While comparing the proposed method to the traditional method RDNN, the suggested method achieved various evaluation measures, such as F1-score (88.9%), recall (91.60%), precision (89.80%), and accuracy (90.3%). The result demonstrated the IFP-RDNN method to predict the oral English pronunciation rating using biosensors. The findings indicate a significant connection between the quality of oral English pronunciation and mental health, with deviations from standard pronunciation being associated with increased stress and emotional suffering.

Limitations and future scope

Pronunciation learning could be impacted by differences in learners' ability, motivation, and exposure to English. Language anxiety, less confidence, and decreased motivation exceed pronounced words, which might have an impact on mental health. Oral pronunciation frequently emphasizes correctness and fluency, but it might exceed how language acquisition affects the mental health and growth of speaking confidence. Future research should examine the pronunciation guidelines in a global setting while acknowledging the variety of English dialects and how they affect learners' mental health in various cultural contexts.

Author contributions: Conceptualization, XZ, WJ and CS; methodology, XZ, WJ and CS; writing—original draft preparation, XZ, WJ and CS; writing—review and editing, XZ, WJ and CS. All authors have read and agreed to the published version of the manuscript.

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Lee, H., 2022. Developing an Approach to Support Instructors to Provide Emotional and Instructional Scaffolding for English Language Learners Through Biosensor-Based Feedback (Doctoral dissertation, University of Maryland, Baltimore County).
2. Cassar, C., McCabe, P. and Cumming, S., 2023. "I still have issues with pronunciation of words": A mixed methods investigation of the psychosocial and speech effects of childhood apraxia of speech in adults. International Journal of Speech-Language Pathology, 25(2), pp.193-205. 10.1080/17549507.2021.2018496
3. Vanova, M., Aldridge-Waddon, L., Jennings, B., Puzzo, I. and Kumari, V., 2021. Reading skills deficits in people with mental illness: A systematic review and meta-analysis. European Psychiatry, 64(1), p.e19. 10.1192/j.eurpsy.2020.98
4. Feijóo-García, P.G., Zalake, M., de Siqueira, A.G., Lok, B. and Hamza-Lup, F., 2021, September. Effects of virtual humans' gender and spoken accent on users' perceptions of expertise in mental wellness conversations. In Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents (pp. 68-75). 10.1145/3472306.3478367
5. Lundin, N.B., Jones, M.N., Myers, E.J., Breier, A. and Minor, K.S., 2022. Semantic and phonetic similarity of verbal fluency responses in early-stage psychosis. Psychiatry Research, 309, p.114404. 10.1016/j.psychres.2022.114404
6. Zeng, Y., 2021. Application of flipped classroom model driven by big data and neural network in oral English teaching. Wireless Communications and Mobile Computing, 2021(1), p.5828129. 10.1155/2021/5828129
7. Yang, Y., 2023. Machine learning for English teaching: a novel evaluation method. International Journal of Computer Applications in Technology, 71(3), pp.258-264.
8. Zhang, Y., 2022. Multimedia-assisted oral English teaching system based on B/S architecture. International Journal of Continuing Engineering Education and Life Long Learning, 32(6), pp.663-680. 10.1504/IJCEELL.2022.126869
9. Li, S., 2022, December. General Design of Automatic Correction System for English Pronunciation Errors Based on DTW Algorithm. In 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC) (pp. 1-5). IEEE. 10.1109/ICMNWC56175.2022.10031909
10. Fu, Y., Zhang, Z. and Yang, H., 2023. Design of Oral English Teaching Assistant System based on deep belief networks. Soft Computing, 27(22), pp.17403-17418. 10.1007/s00500-023-09211-8
11. Liu, Y. and Quan, Q., 2022. AI recognition method of pronunciation errors in oral English speech with the help of big data for personalized learning. Journal of Information & Knowledge Management, 21(Supp02), p.2240028. 10.1142/S0219649222400287
12. Alduais, A., Alarifi, H.S. and Alfadda, H., 2024. Using Biosensors to Detect and Map Language Areas in the Brain for Individuals with Traumatic Brain Injury. Diagnostics, 14(14), p.1535.
13. Ramzan, M. and Javaid, Z.K., 2023. PSYCHOLOGICAL FACTORS INFLUENCING PASHTO SPEAKING ESL STUDENTS'PRONUNCIATION OF ENGLISH VOWELS. Pakistan Journal of Society, Education and Language (PJSEL), 9(2), pp.52-63.
14. Wang, J., 2020. Speech recognition of oral English teaching based on deep belief network. International Journal of Emerging Technologies in Learning (Online), 15(10), p.100. 10.3991/ijet.v15i10.14041
15. Jing, C., Zhao, X., Ren, H., Chen, X. and Gaowa, N., 2022. An approach to oral English assessment based on intelligent computing model. Scientific Programming, 2022(1), p.4663574. 10.1155/2022/4663574
16. Li, Y. and Huang, G., 2022, April. An English Pronunciation Quality Evaluation Model Based on Multi-dimensional Features. In Journal of Physics: Conference Series (Vol. 2224, No. 1, p. 012061). IOP Publishing. 10.1088/1742-6596/2224/1/012061
17. Geng, L., 2021. Evaluation model of college English multimedia teaching effect based on deep convolutional neural networks. Mobile Information Systems, 2021(1), p.1874584. 10.1155/2021/1874584
18. Zhou, R., He, B., Lu, X., Gu, X., Dai, X., Peng, J. and Xi, Y., 2023. Research on the Integration of Natural Phonetic Pronunciation and International Phonetic Alphabet to Improve College Students' Oral English Expression Ability. Lecture Notes on Language and Literature, 6(15), pp.38-43. 10.23977/langl.2023.061507
19. Sharma, L.R., 2021. Significance of teaching the pronunciation of segmental and suprasegmental features of English. Interdisciplinary Research in Education, 6(2), pp.63-78. 10.3126/ire.v6i2.43539
20. Wang, Y. and Zhao, P., 2020. A probe into spoken English recognition in English education based on computer-aided comprehensive analysis. International Journal of Emerging Technologies in Learning (iJET), 15(3), pp.223-233.

21. Liang, E. and Ye, S., 2023. The influence of gender differences on Chinese English learners' oral ability. Journal of Education, Humanities and Social Sciences, 8, pp.813-818. 10.54097/ehss.v8i.4365
22. Iskandar, I., Dewanti, R., Sulistyaningrum, S.D. and Santosa, I., 2024. Scaffolding Assignments to Conciliate the Disinclination to Employ Project-Based Learning of English Pronunciation and Autodidacticism. International Journal of Language Education, 8(2), pp.199-227.
23. Wu, H. and Sangaiah, A.K., 2021. Oral English Speech Recognition Based on Enhanced Temporal Convolutional Network. Intelligent Automation & Soft Computing, 28(1). 10.32604/iasc.2021.016457
24. Gui, L., 2020. An analysis of the strategies for developing students' consciousness of pronunciation and intonation in college English teaching based on international communication. In E3S Web of Conferences (Vol. 214, p. 01041). EDP Sciences. 10.1051/e3sconf/202021401041
25. Huang, A., 2024. Speech Recognition Based on Mobile Biosensor Networks and Quality Evaluation of University Political Education. International Journal of High-Speed Electronics and Systems, p.2540128.
26. Wu, J.Y., Ching, C.T.S., Wang, H.M.D. and Liao, L.D., 2022. Emerging wearable biosensor technologies for stress monitoring and their real-world applications. Biosensors, 12(12), p.1097.