[1]Cuimin Sun*

Enhancing English Oral Teaching through Pyramidal Convolution Shuffle Attention Neural Network and Sea-Horse Optimizer using Virtual Reality Technology

Journal of Electrical Systems

Abstract: - The ultimate objective of teaching English to students is to help them become self-sufficient language learners and users, proficient in efficient language learning techniques, and capable of transmitting information in English. As a result, good English language instruction requires language communication training for both students and teachers, as well as between students. Compared to classroom instruction, English learning could easily facilitate English learning and provides a comfortable environment by reducing the drawback of the conventional classroom, which could lead to lower ratings for mental strain, absence of communication, and fear of making mistakes. To avoid these challenges, the pyramidal convolution shuffle attention Neural Network with sea-horse optimizer is proposed for classifying pronunciation, speaking proficiency, fluency, and intonation, of the English oral teaching. Initially, the data's are gathered via the dataset of oral English teaching in virtual reality dataset. Afterward, the data's are fed to pre-processing. In pre-processing segment; it removes the noise and enhances the input images utilizing federated neural collaborative filtering. The pre-processing output is fed to Feature extraction segment. Here, four statistical features such as kurtosis, mean, skewness, and standard deviation are extracted based on Adaptive and concise empirical wavelet transforms. After that, the extracted features are given to the pyramidal convolution shuffle attention neural network optimized with sea-horse optimizer algorithm for effectively classify the pronunciation, speaking proficiency, fluency, and intonation. The proposed EOT-VRT-PCSANN-SHO approach is implemented in MATLAB. The performance of the proposed EOT-VRT-PCSANN-SHO approach attains 99%, 98%, 97.5%, and 97%, as high accuracy, 98%, 98.5%, 95%, and 99% in F1 score, and 98.7%, 98%, 99%, and 97.5%, in precision, are high, when compared with existing methods.

Keywords: English Oral Teaching, Virtual Reality Technology, Pyramidal Convolution Shuffle Attention Neural Network, Sea-Horse Optimizer, Effective Communication.

I. INTRODUCTION

Currently, technology for virtual reality creates an extremely realistic environment and is based on computers. Simultaneously, individuals can utilize innovation to extend themselves into this virtual climate [1, 2]. Artificial intelligence, automatic control, and other similar technologies are frequently utilized. Likewise, when individuals are in the virtual climate, they can control the virtual climate and accomplish specific purposes [3, 4]. In this setting, people make up most of the population. Lately, the utilization of data innovation [5] in the field of training is increasingly broad. In oral English teaching, because of the rising fame of teaching English in China customary language teaching strategies [6] can presently not address individuals' issues, which is an ever increasing number of clear in schools and colleges.

Lately, the use of data innovation [7, 8] in the field of schooling is increasingly broad. In oral English instructing, because of the rising fame of teaching English in China, customary language teaching techniques [9] can at this point not address individuals' issues, which is an ever increasing number of clear in schools and colleges [10]. Virtual reality (VR) is primarily a significant virtualization and reality synthesis technology that is constantly evolving alongside computer virtualization technology [11]. During the time spent the persistent advancement of VR innovation, exhaustive innovation applications incorporate PC designs, framework programmatic experience and math [12]. Detecting and estimation, computerized picture handling, are to establish an augmented experience climate with cutting edge reproduction that can be seen by the client [13, 14]. The primary goal of incorporating virtual reality (VR) technology into classroom instruction is to create a conducive learning environment and enhance student learning [15]. In English teaching practice, the utilization of VR innovation can understand the natural association between the mental world, the discernment world and the development of the information framework [16]. Then, at that point, by making learning circumstances, students can take part in experiential picking up as per their own desires and needs, and exercises can boost students' independence and really further develop learning impacts [17].

In light of the previous perceptions, in here found that profound brain network innovation driven school oral English teaching changes have turned into a pattern [18, 19]. Instructors will actually want to devote more energy to genuine teaching function thus, and the nature of guidance will be gotten to the next level. Accordingly, a clever oral dialogue system in view of neural network innovation is basic, and the nature of talk is reliant upon exact oral assessment [19].

[1*]Lecturer, College of Basic Education, Nantong Institute of Technology, Nantong, Jiangsu, 226002, China, 13962921992@163.com

First and foremost, it has eliminated the six components of vocabulary richness, grammar, topic relevance, content richness, and fluency in speech. Besides, an assessment model that interfaces explicit PCSANN layers in a feed-forward manner was created. Utilizing the component portrayals of target words in various PCSANN layers, it can get more extravagant setting data and significantly diminish how much model boundaries. At long last, it was directed a reproduction try. The exploratory outcomes displays that the recommended structure is precise in assessing communicated in English and can successfully help the change of communicated in English teaching in schools and colleges. The contributions of the work is as follows

- The PCSANet model, linked with VR technology, improves speech recognition and analysis for English oral teaching. It utilizes pyramidal convolutional neural networks and shuffle attention mechanisms to extract relevant features from spoken English.
- By accurately recognizing and analyzing students' speech patterns, pronunciation, and fluency, the PCSANet enables personalized feedback and assessment, facilitating targeted improvements in oral communication skills.
- The shuffle attention mechanism employed in PCSANet improves the focus on important speech features during training and evaluation. This mechanism enables the model to selectively attend to relevant linguistic cues, such as stress, intonation, and rhythm, which are crucial for effective oral communication in English.
- For improving attention mechanisms, the PCSANet model improves the quality and accuracy of feedback provided to learners, leading to improved speaking skills.
- The SHO is used to optimize the training process of the PCSANet model. It efficiently searches for optimal model parameters, leading to enhanced model performance and convergence speed.
- The addition of SHO with PCSANet helps to enhance the training efficiency and effectiveness, ensuring optimal utilization of computational resources.
- The integration of VR technology provides an immersive learning experience for English oral teaching. Learners can connect in realistic and interactive virtual environments, simulating real-life scenarios for practicing spoken English.
- The grouping of PCSANet and VR technology allows learners to receive instant feedback on their pronunciation, intonation, and overall speaking skills. This immersive approach improves motivation, engagement, and retention, resulting in more effective learning outcomes.
- The PCSANet model, connected with VR technology, allows personalized and adaptive learning knowledge. By analyzing individual learners' speech patterns and progress, the system can provide tailored feedback, exercises, and learning materials.
- The adaptive nature of the system ensures that learners receive instruction and practice that aligns with their specific needs and learning styles, promoting efficient and effective English oral teaching.
- The proposed method contain enhanced speech recognition and analysis, improved attention mechanisms, and optimized training processes, immersive learning experiences, personalized and adaptive learning, and advancements in educational technology.

The rest portions of this manuscript are organized as follows: part 2 examines a survey of the literature; part 3 outlines the proposed approach, While the results and a discussion are presented in part 4, the conclusion is presented in part 5.

II. RECENT RESEARCH WORK: A BRIEF REVIEW

Many studies were proposed in the literature related to deep learning based English oral teaching using virtual reality technology; a few recent works are reviewed here,

Sun [20] have utilized that mix of fuzzy frameworks and deep learning model has given and displays how vulnerability might be actually decreased by using preparing fluffy standards. This study embraced the F-D2QLN to investigate coordinating portable based instructive games into advanced education English learning. The TSK , FIS was utilized to beat the troubles of vulnerability, vagueness, and imprecision intrinsic in normal language handling tasks to further develop the “decision-making process the Q-learning algorithm”. By consistently learning and further developing the growing experience in light of the student's exhibition and progress,‘ double-deep Q-learning in game-based English learning plans to make a versatile’, customized, and powerful opportunity for growth for every one of a kind student.

Liu [21] have fostered that significance of advancing the change of verbal English showing in China's English educating environment. It trusts that to advance the change of oral English instructing, an verbal training climate should be accessible. In any case, the ongoing normal issue in verbal English training in

schools and colleges is that the verbally expressed discussion objects are not adequately standard, or there is no individual who can converse with. In this way, a savvy verbally expressed exchange framework in light of large information and neural network innovation is especially significant, and the nature of discourse relies upon exact spoken discourse assessment. It has initially extricated six elements of elocution quality, familiarity, content extravagance, theme significance, punctuation, and jargon lavishness. Besides, here suggested an assessment model that interfaces explicit TDNN layers in a feed-forward way, utilizing the element portrayal of target words in various TDNN layers, which can get more extravagant setting data and enormously lessen how much model boundaries.

Zhou [22] have presented that VR innovation, is an extensive data innovation, which gives novel plans to English teaching. Using VR technology to teach English is the subject of this article. Utilizing the teaching simulation technique, here examined the plausibility of utilizing VR innovation to tackle the issues of unfortunate oral English, unfortunate English articulation capacity, and absence of English reasoning skill.

Xie et al. [23] have suggested that a blended strategies learn about utilizing virtual reality (VR) apparatuses for fostering understudies' oral capability in learning Chinese as a subsequent language. Twelve students played tour guides for six places over the course of a semester: two without VR equipment and four with. For Presentations 1, 3, 4, and 6, oral presentations were recorded at four distinct times throughout the semester, class observations were taken, reflection papers were written by participants, and person interviews were conducted at the conclusion of the semester.

Muhammad [24] have developed that decide students' perceptions in rural secondary schools on the execution of VR as an option in contrast to the conventional study hall setting, all the more explicitly in the setting of an oral ability example. Moreover, the impression of the understudies incorporates the overall discernment, inspiration levels, interest as well as the expertise improvement of the understudies when exposed to a VR study hall and their impacts on the referenced viewpoints. A overall of 39 students from Sekolah Menengah Kebangsaan Desa Kencana participated in this study. This investigation used a multiple method approach.

Luo [25] have developed that a plenty of choices to concentrate on a language, particularly for non-native speakers of English, utilizing one of the vivid stages, VR. Endeavoring individuals should work on their familiarity and capability to remain cutthroat in the business area. In this case, students would be well served by thinking about business-specific English. The examination looks at how VR permits students to encounter business-related circumstances in an active way. Additionally, it discusses scenarios in which students could realize their full professional potential. A blended philosophy study exhibits how computer generated reality innovations might be utilized to improve English communication capacities.

Zhai [26] have utilized that English live teaching for undergrads for instance, right off the bat dissects the issues existing in the showing live transmission process and proposes relating arrangements from two parts of educators and understudies, to work on the nature of "educating" and "learning". The results are supposed to give reference feelings to the public web-based live educating. Second, the information technology behind the excellent webcasts and short videos is closely related. This paper takes the video live transmission system as the research scenario and behaviors top to bottom exploration on the spectrum sensing in view of cognitive radio innovation.

Wang [27] have utilized that English open-spoken scoring system using convolutional neural networks (CNNs) as NN technology. The framework is capable of providing a comprehensive evaluation of the oral recording at both the phonetic and text levels. ‘The expressed substance is obtained via text record of the recording by an external discourse acknowledgment motor. The framework will separately score the expressed discourse and the expressed substance through several scoring models and add the scoring results as the final score’. An acoustic sensor is embraced to gather elocution signs of communicated in English. The quality of spoken pronunciation can be distinguished using cutting-edge technologies like automatic pattern recognition and signal processing. Comparative semantic units are set apart between acoustic component arrangements, which utilize the equal calculation handling method of multi-registering centers of current GPU and permit numerous units to execute the correlation calculation simultaneously freely.

III. PROPOSED METHODOLOGY

In this section, English oral teaching with virtual reality technology system using pyramidal convolutional shuffle attention neural network optimized with sea-horse optimizer algorithm (EOT-VRT-PCSANN-SHO) is discussed. The block diagram of the proposed EOT-VRT-PCSANN-SHO English oral teaching identification

system is represented in Fig.1. Data collection, pre-processing, feature extraction, and English oral categorization are some of its four phases. Consequently, a thorough explanation of each step is provided below.

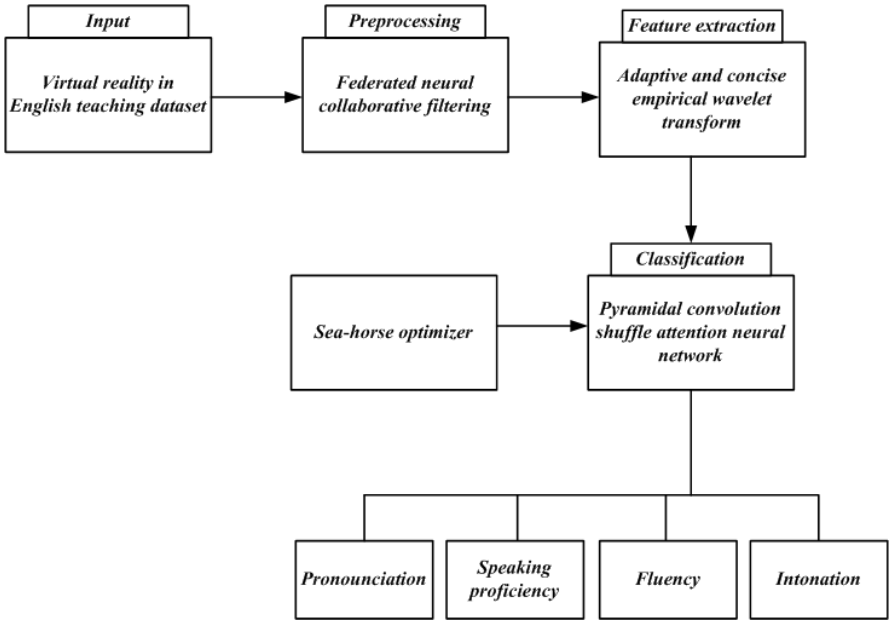

Fig.1: Block diagram of proposed EOT-VRT-PCSANN-SHO

A. Data Acquisition

This analysis uses data from an oral exam from a university situational English course. the extraction of 650 test-takers' responses to an identical open-ended oral question from the examination data between 2012 and 2018 and the manual scoring data from the teacher .The complete dataset is split into two subsets for testing and training: the 150 data pieces in the "test set." And the 500 data pieces in the "training set"

B. Pre-processing using Federated Neural Collaborative Filtering

In order to enhance the cleanliness of the data, to employed the federated neural collaborative filtering (FNCF) [28]. In this step, FNCF performs the data pre-processing, which is utilized for reducing the noise in the dataset. The primary issue pertains to the process of updating the item profile. The agreed-upon samples are subsequently employed to produce a random matrix IR_{ij} , which is determined by the item profile's size, here $j \in C,\ j \neq i$.

Finally, the updated item profile is expressed by:

$$MI_{t+1}^{i} = I_{t+1}^{i} + \sum_{i \in C; i < j} IR_{ij} - \sum_{i \in C; i > j} IR_{ji} \tag{1}$$

When i and j in an ordered pair $(i, j), i < j$ agree on a seed, a random matrix called IR_{ij} is formed, and the marked estimated weights are represented by MI_{t+1}. The coordinated server computes the following after gathering each MI_{t+1}^{C}:

$$I_{t+1}^{SUM} = \sum_{i \in C} MI_{t+1}^{i} \tag{2}$$

The I_{t+1}^{SUM} parameter, which is generated, comprises the total of the weight updates associated with the profile of the item that requires aggregation. The aggregated weights can be produced by the server in the following manner in the most basic form of an aggregation step:

$$I_{t+1} = \frac{MI_{t+1}^{SUM}}{|C|} \tag{3}$$

However, $|C|$ indicates the count of chosen participant at step-time *t*.But still, this form of aggregation fails to account for individual item adjustments, which leads to slower convergence.

Following local training and seed exchange, to produce $2(|C|-1)$ arbitrary matrices along with $|C|-1$ arbitrary vectors, from which, $|C|$ specifies that count of participants in the current cycle. The number of parameters in each model under consideration grows in accordance with the count of items $|I|$ and the size D that is specified in the profile. Every one generates a distinct result for the GMF model, which comprises a solitary linear layer equipped with one single processing unit.

$$(|C|-1).(D.|I|+D+1+|I|) \tag{4}$$

The factors are determined based on the specified seeds. In this context, $D+1$ indicates the count of inputs and biases in the neural framework $D.|I|$ implies the count of values in the item profile, and $|I|$ represents the count of factors comprising the random interaction vector. When the neural design is exceeded the model transforms into conventional MF, and the count of the produced parameters reduces to $(|C|-1)\cdot(D+1+|I|)$. In contrast, the MLP model incorporates a framework that consists of a minimum of a single hidden layer. The quantity of factors that must be produced consequently escalates in accordance with the count of processing units in every hidden layer and the specified hidden layers. With greater precision, an individual produces a designated quantity of parameters.

$$(|C|-1).\left(D.|I|+2D+h1+\left(\sum_{i=1}^{n-1}h_i\,.h_{i+1}\right)+h_n+\left(\sum_{i=1}^{n}h_i\right)+1+|I|\right) \tag{5}$$

After the agreement on the seeds, the total number of parameters to be generated is determined. These parameters include $2D.h1$, which represents the input size, $\sum_{i=1}^{n-1}h_i\,.h_{i+1}$, which represents the count of weights, and $\sum_{i=1}^{n}h_i+1$, which represents the network bias. The value of h_i represents the count of processing units on the i[th] hidden layer. Particularly necessary parameters pertain to the item profile $2D.I$ in the NeuMF model, which is formed by concatenating GMF and MLP.

$$(|C|-1).\left(2D.|I|+2D+h1+\left(\sum_{i=1}^{n-1}h_i\,.h_{i+1}\right)+D+h_n+\left(\sum_{i=1}^{n}h_i\right)+1+|I|\right) \tag{6}$$

The count of items $|I|$, The arbitrary of dimension D,and the count of processing units h_i are established before any calculation begins in the stage of federated recommender training. The number of customers taking part in each aggregation round then dictates the increase of parameters. Under MF-SecAvg, clients carry out basic matrix operations following a local update using parameters that are produced at random according to a preset order. They then distribute the final result by performing element-wise subtraction or addition of the computed weights. The feature extraction phase receives the pre-processed data at the end.

C. Feature Extraction using Adaptive and Concise Empirical Wavelet Transforms

Adaptive and brief empirical wavelet transforms are used in this step to assist clarify the important features under pre-processing [29]. From pre-processing output, it has significant characteristics of grayscale statistic features like skewness, kurtosis standard deviation, and mean are extracted with the help of ACEWT. The actual, unpretentious feature extractor is called ACEWT. To provide accurate virtually English oral teaching detection, the main objective of the ACEWT feature extraction technique is to take out discriminative characteristics. First, the spectral density of power of the data can be computed using equation (7).

$$PSD_{image}=\lim_{\alpha\to\infty}E\left[output\ \ pre-processed\ \ image^{2}\right] \tag{7}$$

Then set the minima point of data power spectral density as boundaries $(B_1\ to\ B_z)$. Subsequently, the boundaries are divided into several components using an adaptive and brief empirical wavelet transform. By

this each component in the data are reconstructed based on the function of empirical scale and function of empirical wavelet [28]. Equation (8) can be used to calculate the empirical scale function.

$$ESF = \begin{cases} 1; & |B| \le (1-\gamma)B_z \\ \cos\left[\frac{\pi}{2} * TF\left(\frac{1}{2\gamma B_z}\left(|B| - (1-\gamma)B_z\right)\right)\right]; & (1-\gamma)B_z \le |B| \le (1-\gamma)B_z \\ 0; & others \end{cases} \quad (8)$$

Equation (9) can be used to calculate the empirical wavelet function.

$$EWF = \begin{cases} 1; & (1-\gamma)B_z \le |B| \le (1-\gamma)B_{z+1} \\ \cos\left[\frac{\frac{\pi}{2} * TF\left(|B| - (1-\gamma)B_{Z+1}\right)}{2\gamma B_{Z+1}}\right]; & (1-\gamma)B_{z+1} \le |B| \le (1-\gamma)B_{z+1} \\ \sin\left[\frac{\frac{\pi}{2} * TF\left(|B| - (1-\gamma)B_{Z+1}\right)}{2\gamma B_{Z+1}}\right]; & (1-\gamma)B_z \le |B| \le (1-\gamma)B_{z+1} \\ 0; & others \end{cases} \quad (9)$$

Where, TF represents the transition function; γ depicts the coefficient. By this, the features are extracted. Below are the features that were extracted using the ACEWT approach.

1) Grayscale statistic features

The Grayscale statistic features like skewness ,mean, kurtosis and standard deviation are explained below,

Mean

To find the mean of the grayscale statistic properties, utilize equation (10).

$$Mean = \frac{1}{p*q}\sum_{x=0}^{p-1}\sum_{y=0}^{q-1} ACEWT(x,y) \quad (10)$$

Here, $\mathrm{ACEWT}(x,y)$ indicates Adaptive and concise empirical wavelet transforms (ACEWT) matrix, p and q is the data height and breadth of Pixels in the ACEWT matrix.

Standard Deviation (SD)

Equation (11) can be used to calculate the Grayscale statistic features' Standard Deviation.

$$SD = \sqrt{\frac{1}{p*q}\sum_{x=0}^{p-1}\sum_{y=0}^{q-1}\left(ACEWT(x,y) - Mean\right)^2} \quad (11)$$

Skewness: Eqn (12) can be used to calculate the skewness of the features in the grayscale statistics.

$$Skewness = \sqrt{\frac{1}{p*q}\sum_{x=0}^{p-1}\sum_{y=0}^{q-1}\left(ACEWT(x,y) - Mean\right)^3} \quad (12)$$

Kurtosis: Equation (13) can be used to find the Kurtosis of the features of the grayscale statistic.

$$Kurtosis = \sqrt{\frac{1}{p*q}\sum_{x=0}^{p-1}\sum_{y=0}^{q-1}\left(ACEWT(x,y) - Mean\right)^4} \quad (13)$$

Then these extracted features are given into oral English teaching classifier.

D. Oral English Teaching Classification using Pyramidal Convolution Shuffle Attention Neural Network

In this section, English teaching Classification using PCSANN [30] is discussed. A pyramidal convolution captures more complex characteristics without adding to the processing burden by utilizing a series of convolution kernels at various sizes with varying spatial resolution and depth. Furthermore, the convolution layer's capacity to classify detailed features is derived from the pyramidal convolution's ability to identify spatial feature correlations at various levels. Additionally, it successfully resolves the issue of down sampling-related local information loss in the interim. The features are splitted into groups, that is expressed in equation (14)

$$Y = \left[Y_1, Y_2 \ldots . Y_N\right] \quad (14)$$

Where, Y represents the feature map, N represents the number of groups. Then the function efficiency is enhanced by the use of the linear function. Finally, global information is introduced through embedded with the original assets, after its activation sigmoid function to obtain a class representation, it is given in equation (15)

$$Y_{a1} = \gamma\left(h_w(q)\right).Y_{a1} = \gamma\left(p_1 q_1 + k_1\right).Y_{a1} \tag{15}$$

Where, γ indicates the sigmoid activation function, h_w indicates the linear function, q_1 indicates the average pooled function, and p_1, k_1 are obtained from network training. Spatial perception can be preserved attention improvement. Finally, worldwide information is embedded by multiplying the value of the original asset is activated with a sigmoid function and the spatial attention function is given in equation (16)

$$Y_{a2} = \gamma\left(p_2 . HR\left(Y_{a2}\right) + k_2\right) - Y_{a2} \tag{16}$$

Where, HR denotes the group norm normalization function, q_2 is the normalized feature. To restore the original dimension, merge the grouped blocks once more. Equation (17) is obtained once the two branches have been split and combined following the completion of both attentions' learning and recalibrating of the features.

$$Y_{a2} = \left[Y_{a1}, Y_{a2}\right] \times P \times T \tag{17}$$

Where, P, T are denoted as the sub features. Then every sub functions are grouped. Finally, channel grouping procedure is carried out. PCSANN is made up of a fully connected layer, an average pooling layer, and a total of four residual blocks. The completely connected layer's output vector moves forward through a sigmoid layer, this is expressed in equation (18),

$$\tilde{l}\left(w|J\right) = 1/\left(1 + \exp\left(-lh\left(w|J\right)\right)\right) \tag{18}$$

Where, J is the input data, $\tilde{l}\left(w|J\right)$ is the probability score, the global branching is given in equation (19)

$$P(C) = \frac{-1}{W}\sum_{w=1}^{W} p_w \log\left(\tilde{l}\left(w\ |J\right)\right) + \left(1 - p_c\right)\log\left(1 - \tilde{l}\left(w|J\right)\right) \tag{19}$$

Where, p_w represents the true label, w and W represents the total number of teaching categories. Finally, PCSANN accurately classifies the English teaching. The optimization procedures needed to choose the most suitable variables to confirm precise detection are typically not provided by PCSANN. Therefore, for the optimization process to work, the PCSANN weight parameter γ, is crucial.

1) Sea Horse Optimizer

The temperament, mobility, and predation patterns of sea horses serve as inspiration for SHO, an innovative nature-inspired meta-heuristic algorithm [31]. The three crucial mechanisms of mobility, predation, and breeding are all part of the SHO algorithm. To stabilize the exploitation and exploration of SHO, the global and local search rules are adapted to the corresponding social actions of migration and predation. And once the first two behaviors are achieved, the breeding behavior is completed. Fig 2 displays the flowchart of the algorithm. The following are the SHO steps:

Step 1: Initialization

Initialize the input parameter as weight parameter of the PCSANN.

Step 2: Random Generation

Each resolution is randomly produced to involving the upper bound and lower bound of a identify difficulty, indicated by $UB and LB$, correspondingly. The formulation of the i^{th} individual Y_i in search space $\left[UB, LB\right]$ is

$$Y_i = \left\lfloor y_i^1, \ldots\ldots, y_i^{Dim} \right\rfloor \tag{20}$$

$$y_i^j = rand \times \left(UB^j - LB^j\right) + LB^j \tag{21}$$

Here , i denoted as the positive integer in the range $\left[1, Dim\right]$, $rand$ is denoted as the random value in $\left[0,1\right]$, The jth dimension in the ith individual is designated as y_i^j. $UB^j and LB^j$ are denoted as the upper bound and lower bound of the j[th] variable of the optimized problem.

Step 3: Fitness Function

Taking the least optimization crisis as an instance, the entity with the smallest fitness is regarded as the elite entity indicated as Y_{elite}, which is obtained as below

$$Y_{eite} = opt(\gamma) \tag{22}$$

Step 4: Movement Behavior

For the primary performance, the unusual progress patterns of sea horses just about pursue the normal distribution $randn(0,1)$. Use $r_1 = 0$ as the cut-off point, allocating half to global search and the other half to local mining, to maintain a balance between the performance of exploration and exploitation. So actions can be separated into the subsequent 2 phases.

***Step 4.1*:** Exploitation Phase

The sea horse's spiral movements in relation to the water's vortex. When the usual arbitrary value r_1 is on the right side of the cut-off, it typically discovers the local SHO exploitation. Sea horses that follow the spiral motion are terrible when it comes to the elite Y_{elite}. Specifically, the Lévy flight aims to duplicate the size of sea horses' group steps, which benefits the sea horse with the highest chance of migrating earlier and escaping SHO's harsh local exploitation. Meanwhile, the rotation angle is regularly modified by the mode of spiral movement of the sea horse in order to extend the neighborhoods of the current local solutions. The sea horse's novel position in this instance can be mathematically described as follows:

$$Y_{new}^{1}(t+1) = Y_i(t) + Levy(\lambda)\left(\left(Y_{elite}(t) - Y_i(t)\right) \times a \times b \times c + Y_{elite}(t)\right) \tag{23}$$

where the Coordinate components in three dimensions (a,b,c) under the spiral movement are represented by $a = \rho \times \cos(\theta)$, $b = \rho \times \sin(\theta)$, and $c = \rho \times \theta$, correspondingly. These elements are helpful in keeping search agents' positions updated. The logarithmic spiral constants define the stems' length $m and n$ is represented by the symbol $\rho = m \times e^{\theta n}$. The distribution function for Levy Flight,is indicated as $Levy(c)$, is written as follows.

$$Levy(c) = s \times \frac{o \times \sigma}{|k|^{\frac{1}{\lambda}}} \tag{24}$$

Where, λ is denoted as the random value between [0,2], s is denoted as the fixed constant, $o and k$ are denoted as the random numbers between [0,1], σ is formulated as below

$$\sigma = \left(\frac{\Gamma(1+\lambda) \times \sin\left(\frac{\pi\lambda}{2}\right)}{\Gamma\left(\frac{1+\lambda}{2}\right) \times \lambda \times 2^{\left(\frac{\lambda-1}{2}\right)}} \right) \tag{25}$$

Step 4.2: Exploration Phase

The sea horse's Brownian motions in relation to the waves. The SHO is investigated beneath the drifting movement when r1 is situated on the left side of the cut-off point. Here, staying away from the SHO's local extremum is crucial to the search procedure. Brownian motion can effectively simulate an extra moving length of the sea horse to guarantee that it is well explored in the search space. Its arithmetical appearance for this case is

$$Y_{new}^{1}(t+1) = Y_i(t) + rand * l * \beta_t * \left(Y_i(t) - \beta_t * Y_{elite}\right) \tag{26}$$

Where β_t indicates the Brownian motion random walk coefficient, which is essentially a random number that generally follows the normal distribution, and l is the constant coefficient. Then it given by the below Equation

$$\beta_t = \frac{1}{\sqrt{2\pi}\exp\left(-\frac{y^2}{2}\right)} \tag{27}$$

Step 4.3: Update the Phases

In total, these two situations can be set up in the following ways to reach the sea horse's novel location at iteration t.

$$Y_{new}^{1}(t+1) = \begin{cases} Y_i(t) + Levy(\lambda)\left(\left(Y_{elite}(t) - Y_i(t)\right) \times a \times b \times c + Y_{elite}(t)\right), & r_1 > 0 \\ Y_i(t) + rand * l * \beta_t * \left(Y_i(t) - \beta_t * Y_{elite}\right), & r_1 \le 0 \end{cases} \tag{28}$$

Where, $r_1 = randn()$ is denoted as a normal random number.

Step 5: Predation Behavior

The sea horse's chances of success or failure in its pursuit of zooplankton and small crustaceans are both high and low. Given that there is a 90% possibility that the sea horse will successfully capture food, the accidental number r_2 of SHO is meant to differentiate between these 2 outcomes and is set to a critical value of 0.1. Since the elite, to a sure, designates the estimated location of the victim, the predation achievement highlights the exploitation aptitude of SHO. If $r_2 > 0.1$, it denotes that the sea horse was successful in its predation.; in other words, the sea horse creeps up on the elite prey, outpaces it, and eventually catches it. Otherwise, both respond at the opposite speed from before the predation fails, indicating that the sea horse is trending to search the search space. This predatory tendency can be articulated mathematically as follows:

$$Y_{new}^{2}(t+1) = \begin{cases} \alpha * \left(Y_{elite} - rand * Y_{new}^{1}(t)\right) + (1-\alpha) * Y_{elite}, \ if \ r_2 > 0.1 \\ (1-\alpha) * \left(Y_{new}^{1}(t) - rand * Y_{elite}\right) + \alpha * Y_{new}^{1}(t), if \ r_2 \le 0.1 \end{cases} \tag{29}$$

Here, r_2 represents a random value between [0,1], $Y_{new}^{1}(t)$ represents after moving at iteration t, the sea horse's new position t, and α is a linear function of iterations that lowers the sea horse's moving step size as it hunts victim. Here is the formula for this.

$$\alpha = \left(1 - \frac{t}{T}\right)^{\frac{\mu}{T}} \tag{30}$$

here, T is denoted as the maximum number of iterations.

Step 6: Breeding Behavior

The population's fitness values are used to divide it into groups of men and women.. Notably, since male sea horses are the ones who procreate, the SHO approach divides the individuals with the highest fitness values in half, classifying them as dads and the other half as moms. In addition to preventing the over-localization of noval solutions, this division will make it easier for the next generation to inherit good qualities from their mothers and fathers. The job that sea horses are assigned mathematically can be succinctly expressed as

$$\begin{aligned} f &= Y_{sort}^{1}(1 : pop/2) \\ m &= Y_{sort}^{2}(pop/2 + 1 : pop) \end{aligned} \tag{31}$$

Where, The entire Y_{new}^{2} in ascending sequence of fitness values is represented by the symbol Y_{sort}^{2}. f *and* m are denoted as the father and mother which is indicates the male and female populations, correspondingly.

To produce novel children, random mating occurs among females and men. It is thought that every pair of sea horses gives birth to just one child in order to facilitate the easy execution of the suggested SHO algorithm. The fifth offspring's expression is as follows.

$$Y_i^{offspring} = r_3 Y_i^{f} + (1 - r_3) Y_i^{m} \tag{32}$$

Where, The positive integer in the range of $1 - pop/2$ is known as i, r_3 is denoted as the random number between [0,1], The individuals selected from the female and male populations who were chosen at random are identified as Y_i^{f} and Y_i^{m}, correspondingly.

Step 7: Termination Process

If the optimal solution is gets, the process will be stop. Or else it goes to the fitness step.

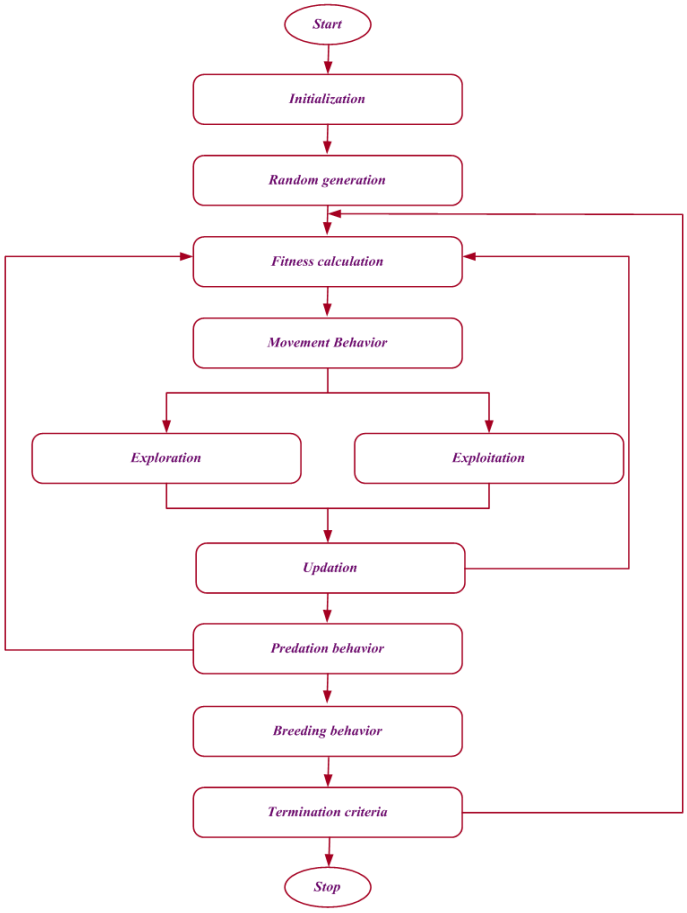

Fig 2: Flowchart of the SHO algorithm

IV. RESULT AND DISCUSSION

This part discusses the experimental outcome of the proposed approach. Computers running Windows 7 with an Intel Core i5 2.50 GHz CPU, 8GB RAM, and simulation software are used. Next, MATLAB is used to simulate the proposed method under a number of performance criteria, including F1-score, accuracy, precision, and retention rate. The obtained outcome of the proposed EOT-VRT-PCSANN-SHO method are analyzed with the existing methods, such EOT-VRT-CNN [27], EOT-VRT-F-D2QLN [20], and EOT-VRT-TDNN [21] methods, correspondingly.

A. Performance Measures

This is a significant phase in choosing the optimal classifier. The performance is evaluated by looking at performance measures including retention rate, F1-score, precision, and accuracy. The confusion matrix is considered in order to scale the performance indicators. False Positive/ Negative ,True Positive/ Negative, values are required to scale the confusion matrix.

- True Positive (TP): Samples in which the true class label is exactly the same as the predicted class label when the count is positive.
- True Negative (TN): Samples in which the true class label is exactly the same as the predicted class label when the count is negative .
- False Positive (FP): The number of samples in which the real class label is imprecise and the predicted class label implies a positive value.
- False Negative (FN): The number of samples in which the real class label is imprecise and the predicted class label implies a negative value.

1) Accuracy

It is measured by following equation (33),

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{33}$$

2) Precision

Precision is a metric which quantifies the count of correct positive prediction made. This is computed via following equation (34)

$$\Pr ecision = \frac{TP}{TP + FP} \tag{34}$$

3) F1 Score

This is determined by equation (35),

$$F1Score = \frac{TP}{TP + \frac{1}{2}\left[FP + FN\right]} \tag{35}$$

4) Retention rate

This is scaled by equation (36),

$$\text{Retention Rate} = \frac{TP}{AP} \times 100 \tag{36}$$

B. Performance Analysis

Figure 3 to 8 depicts simulation results of proposed EOT-VRT-PCSANN-SHO method. Then, the proposed EOT-VRT-PCSANN-SHO method is likened to existing EOT-VRT-CNN, EOT-VRT-F-D2QLN, and EOT-VRT-TDNN methods.

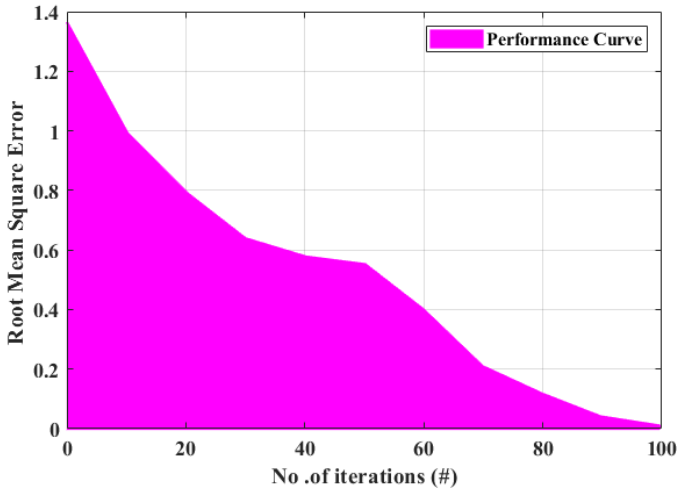

Fig 3: Performance of RMSE Analyses

Fig 3 shows the performance of RMSE analyses. The 100th iterations, the RMSE attains 0.01 error.

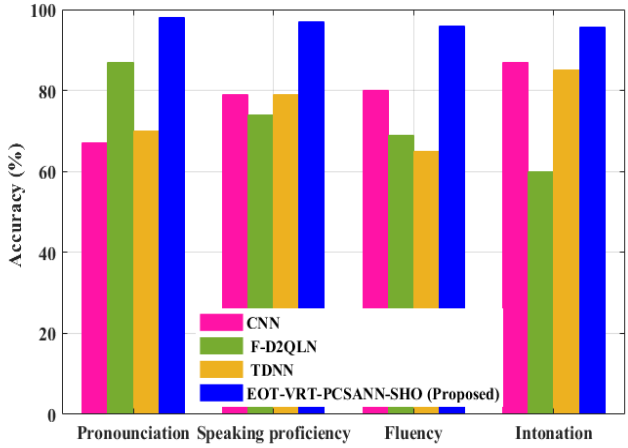

Fig 4: Performance of accuracy analyses

Fig 4 displays the performance of the accuracy analyses. The pronunciation attains 65%, 85%, 68%, and 99%, the speaking proficiency attains 79%, 75%, 79.3%, and 98%, the fluency attains 80%, 67%, 63%, 97.5%, and the intonation attains 84%, 60%, 84%, and 97%, the proposed is higher when compared to the EOT-VRT-CNN, EOT-VRT-F-D2QLN, and EOT-VRT-TDNN, respectively.

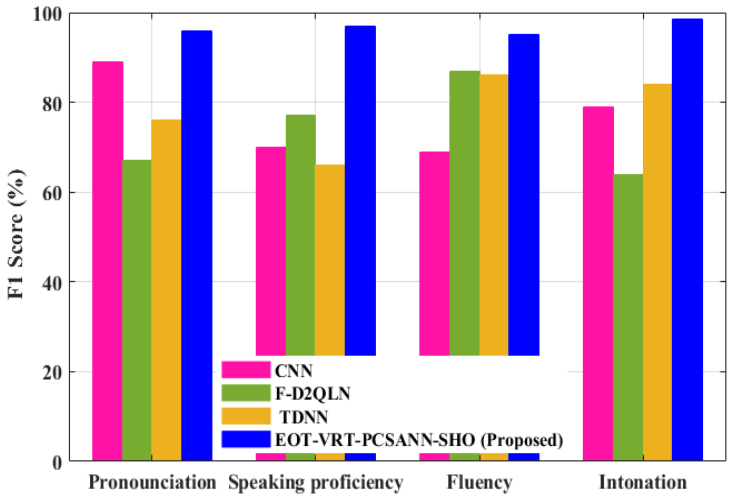

Fig 5: Performance of F1 score

Fig 5 displays the performance of F1 score. The pronunciation attains 65%, 85%, 68%, and 98%, the speaking proficiency attains 70%, 77%, 64%, and 98.5%, the fluency attains 67% 86%, 85%, and 95%, and the intonation attains 79%, 63%, 83%, and 99%, the proposed is higher when compared to the EOT-VRT-CNN, EOT-VRT-F-D2QLN, and EOT-VRT-TDNN, respectively.

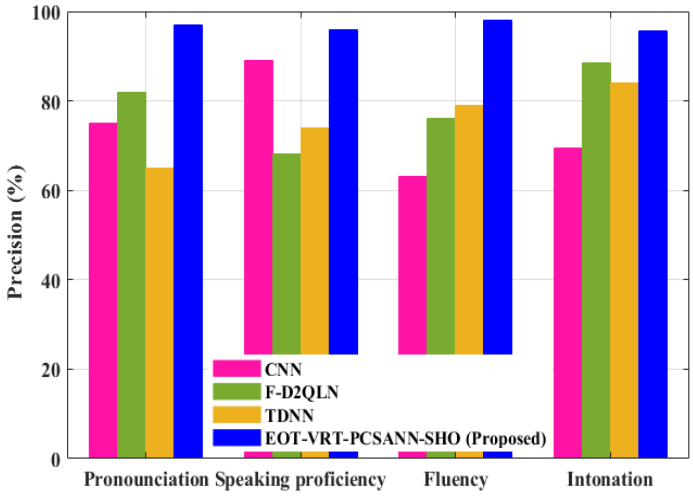

Fig 6: Performance of the precision analyses

Fig 6 shows the performance of the precision analyses. The pronunciation attains 76%, 81%, 63%, and 98.7%, the speaking proficiency attains 90%, 65%, 75%, and 98% the fluency attains 62%, 77.5%, 79%, and 99%, and the intonation attains 69%, 85%, 82%, and 97.5%, the proposed is higher when compared to the EOT-VRT-CNN, EOT-VRT-F-D2QLN, and EOT-VRT-TDNN, respectively.

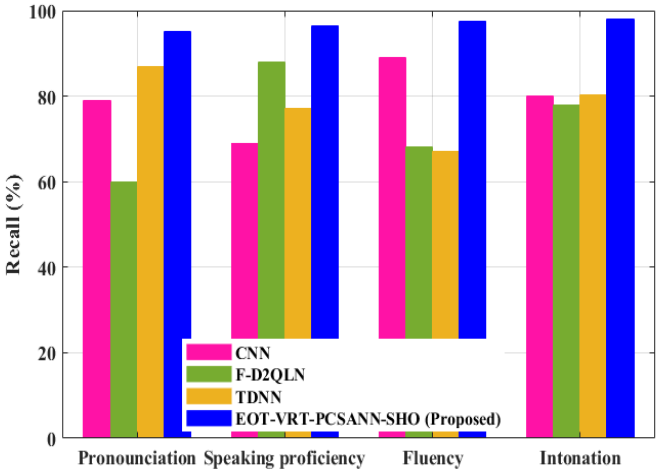

Fig 7: Performance of recall analyses

Fig 7 shows the performance of recall analyses. The pronunciation attains 79%, 60%, 83%, and 96%, the speaking proficiency attains 68%, 86%, 78%, and 98%, the fluency attains 88%, 65%, 64%, and 99%, and the intonation attains 80%, 79%, 70.5%, and 99%, the proposed is higher when compared to the EOT-VRT-CNN, EOT-VRT-F-D2QLN, and EOT-VRT-TDNN, respectively.

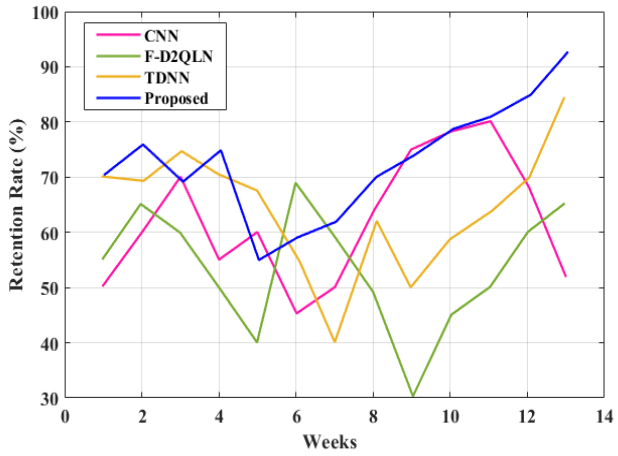

Fig 8: Performance of retention rate analysis

Fig 8 shows the performance of retention rate analysis. The retention rate is attained the proposed, EOT-VRT-CNN, EOT-VRT-F-D2QLN, and EOT-VRT-TDNN, as 93%, 53%, 66%, and 85%, respectively. The proposed is higher than the existing methods.

V. CONCLUSION

In this section, English oral teaching with virtual reality system using Dual-discriminator conditional generative adversarial network optimized with pyramidal convolution shuffle attention Neural Network with sea-horse optimizer was successfully implemented for classifying pronunciation, speaking proficiency, fluency, and intonation, of the English oral teaching (EOT-VRT-PCSANN-SHO). The proposed EOT-VRT-PCSANN-SHO approach is implemented in MATLAB utilizing the dataset of English oral teaching dataset. The performance of the proposed EOT-VRT-PCSANN-SHO approach contains EOT-VRT-PCSANN-SHO method attains 99%, 98%, 97.5%, and 97%, as high accuracy, 98%, 98.5%, 95%, and 99% in F1 score, and 98.7%, 98%, 99%, and 97.5%, in precision, are high, when compared with existing methods.

REFERENCES

[1] Bi, S. (2020). Intelligent system for English translation using automated knowledge base. *Journal of Intelligent & Fuzzy Systems*, *39*(4), 5057-5066.

[2] Chen, Y. L., & Hsu, C. C. (2020). Self-regulated mobile game-based English learning in a virtual reality environment. *Computers & Education*, *154*, 103910.

[3] Sun, R., Zhang, H., Li, J., Zhao, J., & Dong, P. (2020). Assessment-for-learning teaching mode based on interactive teaching approach in college English. *International Journal of Emerging Technologies in Learning (Online)*, *15*(21), 24.

[4] Xu, Z., Chen, Z., Eutsler, L., Geng, Z., & Kogut, A. (2020). A scoping review of digital game-based technology on English language learning. *Educational Technology Research and Development*, *68*, 877-904.

[5] Asongu, S. A., & Odhiambo, N. M. (2019). Basic formal education quality, information technology, and inclusive human development in sub-Saharan Africa. *Sustainable Development*, *27*(3), 419-428.

[6] Qiu, L. (2019). Computer-Aided English Teaching Platform Based on Secure Shell Framework. *International Journal of Emerging Technologies in Learning*, *14*(16).

[7] Ahmed, M. K. (2018). Multimedia aided language teaching: an ideal pedagogy in the English language teaching of Bangladesh. *American International Journal of Social Science Research*, *3*(1), 39-47.

[8] Geng, L. (2021). Evaluation model of college english multimedia teaching effect based on deep convolutional neural networks. *Mobile Information Systems*, *2021*, 1-8.

[9] Kormos, J., & Préfontaine, Y. (2017). Affective factors influencing fluent performance: French learners' appraisals of second language speech tasks. *Language Teaching Research*, *21*(6), 699-716.

[10] Ockey, G. J., Gu, L., & Keehner, M. (2017). Web-based virtual environments for facilitating assessment of L2 oral communication ability. *Language Assessment Quarterly*, *14*(4), 346-359.

[11] Wang, J., An, N., & Wright, C. (2018). Enhancing beginner learners' oral proficiency in a flipped Chinese foreign language classroom. *Computer Assisted Language Learning*, *31*(5-6), 490-521.

[12] Klimova, B. (2021). Use of virtual reality in non-native language learning and teaching. *Procedia Computer Science*, *192*, 1385-1392.

[13] Wong, Y. R., Wong, P. L., Wong, P. W., & Goh, C. P. (2020). The Implementation of Virtual Reality (VR) in Tertiary Education in Malaysia. In *International Conference on Digital Transformation and Applications (ICDXA)*.

[14] Adokorach, M., & Isingoma, B. (2022). Homogeneity and heterogeneity in the pronunciation of English among Ugandans: A preliminary study. *English Today*, *38*(1), 15-26.

[15] Kim, N. Y., Cha, Y., & Kim, H. S. (2019). Future english learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning*, *22*(3).

[16] Chen, C. Y. (2022). Immersive virtual reality to train preservice teachers in managing students' challenging behaviours: A pilot study. *British Journal of Educational Technology*, *53*(4), 998-1024.

[17] Ban, H., & Ning, J. (2021). Online English teaching based on artificial intelligence internet technology embedded system. *Mobile Information Systems*, *2021*, 1-9.

[18] Divekar*, R. R., Drozdal*, J., Chabot*, S., Zhou, Y., Su, H., Chen, Y., ... & Braasch, J. (2022). Foreign language acquisition via artificial intelligence and extended reality: design and evaluation. *Computer Assisted Language Learning*, *35*(9), 2332-2360.

[19] Kozo, J., Wooten, W., Porter, H., & Gaida, E. (2020). The partner relay communication network: sharing information during emergencies with limited english proficient populations. *Health security*, *18*(1), 49-56.

[20] Sun, H. (2023). Enhancing Higher Education English Learning through Virtual Reality and Game-Based Approaches Using the Fuzzy Deep Model.

[21] Liu, H. (2021). College oral English teaching reform driven by big data and deep neural network technology. *Wireless Communications and Mobile Computing*, *2021*, 1-8.

[22] Zhou, Y. (2020). VR technology in English teaching from the perspective of knowledge visualization. *IEEE Access*.

[23] Xie, Y., Chen, Y., & Ryder, L. H. (2021). Effects of using mobile-based virtual reality on Chinese L2 students' oral proficiency. *Computer Assisted Language Learning*, *34*(3), 225-245.

[24] Muhammad, R. (2023). Enhancing English Oral Skills among Malaysian Rural School Students through the Implementation of Virtual Reality (VR). *International Journal on E-Learning Practices (IJELP)*, *6*(1).

[25] Luo, X. (2022). Practice of artificial intelligence and virtual reality technology in college English dialogue scene simulation. *Wireless Communications and Mobile Computing*, *2022*.

[26] Zhai, H. (2023). Virtual Reality-Enabled Deep Learning and Communication Technology for English Teaching through Webcast and Short Video.

[27] Wang, X. (2022). Research on Open Oral English Scoring System Based on Neural Network. *Computational Intelligence and Neuroscience*, *2022*.

[28] Perifanis, V., & Efraimidis, P. S. (2022). Federated neural collaborative filtering. *Knowledge-Based Systems*, *242*, 108441.

[29] Zhang, K., Ma, C., Xu, Y., Chen, P., & Du, J. (2021). Feature extraction method based on adaptive and concise empirical wavelet transform and its applications in bearing fault diagnosis. *Measurement*, *172*, 108976.

[30] Chen, K., Wang, X., & Zhang, S. (2022). Thorax disease classification based on pyramidal convolution shuffle attention neural network. *IEEE Access*, *10*, 85571-85581.

[31] Zhao, S., Zhang, T., Ma, S., & Wang, M. (2023). Sea-horse optimizer: a novel nature-inspired meta-heuristic for global optimization problems. *Applied Intelligence*, *53*(10), 11833-11860.