

¹Xiaoli Hu*

English Speaking Assessment Algorithm Based on Deep Learning



Abstract: - English as a Foreign Language (EFL) students perform when speaking in public. An increasingly globalized world, effective public speaking is critical, but EFL students struggle to perform it, despite importance of qualities such as eye contact, speech pauses, there is presently no objective examination of such elements. A summative assessment has historically been the predominant form of evaluation in college English speaking assessments. Exam-centric teaching has considerable negative effect on foreign language training. In this research work, English Speaking Assessment Algorithm Based on Deep Learning (ESA-NEGCN-NBOA) is proposed. Initially, input video data are gathered from the multiple video dataset (MVD). The input video data is then pre-processed using Deep Attentional Guided Image Filtering (DAGIF) to remove presence of signal-dependent noise and improve lack of pixels from the regions and enhanced the video data. The data that has been pre-processed is utilized to Feature extraction using New General Double Integral Transform (NGDIT), which extract the significant attributes such as mel-frequency cepstral coefficients, energy, speech rate and pitch. Then NEGCN is proposed to improve students spoken English performance by assessing the English speakers. In general, NEGCN doesn't express some adaption of optimization approaches for determining optimal parameters to promise exact improvement of assessment. Therefore, NBOA is proposed to enhance weight parameter of NEGCN for English speaking assessments, which precisely assess the English speaking. Performance measures such as accuracy, assessment error, evaluation time, pretest and posttest are examined when the proposed ESA-NEGCN-NBOA method is put into practice. The proposed ESA-NEGCN-NBOA method attains 21.36%, 23.42% and 19.29% higher accuracy, 23.36%, 18.42% and 28.27% lower evaluation error, 20.36%, 27.42%, 28.17% lesser evaluation time analysed with existing techniques, likes innovative strategy towards oral English assessment utilizing machine learning, data mining, blockchain methods (IST-OEA-ML), machine learning assessment system for spoken English depend on linear predictive coding (AS-SE-LPC-ML), multimodal transfer learning for oral presentation assessment (MM-TL-OPA) respectively.

Keywords: Deep Attentional Guided Image Filtering, English Speaking Assessment, New General Double Integral Transform, Neighbor Enhanced Graph Convolutional Networks, Namib Beetle Optimization Algorithm.

I. INTRODUCTION

According to English speaking assessments, the capacity to efficiently communicate using verbal, nonverbal clues, term "oral English speaking assessment" denotes to process of assessing someone's oral English speaking, comprehension abilities. Oral test allows to demonstrated presentation, addressing, and two-way communication abilities, as well as knowledge, understanding of topic utilizing a SN. Sensor networks principal job is to acquire data [1, 2]. Although fact that sensor technology is constantly evolving further effective at collecting information from physical world, SNs are classically small, low-cost devices with restricted computing, storing, communication abilities to handle scenarios connecting widespread usage [3, 4]. English is widely spoken, vital language around the world. Accordingly, it is widely spoken throughout country. One of the organization's primary goals is to unite Malaysia's diverse population. In school uses a sensor system to monitor each student's English learning. To become proficient in English, students must complete all of the course materials [5, 6]. Machine learning have a significant impact on a range of important aspects of daily life, including education and healthcare, Fluency in English and diverse skill sets are more important than machine learning. Education, worldwide medical care, business are some examples of SN applications ML has significant influence on WSNs [7, 8]. Machine learning research in education has the potential to profoundly impact instructional technology, despite its young age. Machine learning and technological breakthroughs have simplified and improved the process of learning and sharing new information via sensor networks. Reading and writing are essential for developing fluency in English, broad knowledge base [9, 10]. Students may benefit from ML method based on scientific results, teaching technique that excludes and selects spiritual material. Sensor networks and machine learning-based channels have helped students enhance their English skills [11, 12]. Internationalization has increased demand for public speaking abilities in English. As role of academic requirements, "English as Second Language", EFL students exhibit public speaking abilities. Speakers of English as ESL have higher challenges in public speaking than native speakers [13, 14].

¹ School of International Education, Jiangsu Maritime Institute, Nanjing, Jiangsu, China, 211170

*Corresponding author e-mail: 20070872@jmi.edu.cn

Copyright © JES 2024 on-line: journal.esrgroups.org

Numerous English public speaking textbooks for EFL or ESL students focus on speaking, writing speeches in front of audience [15, 16]. "Data mining" is development of uncovering knowledge, information from important amount of imperfect, fuzzily, noisy, randomly made data that unknown prior but may be useful. The phrase "mining knowledge from data" might well explanation of DM [17, 18]. Lesser conceptions are alike to this one, likes data fusion, data analysis, skill detection. It is a term utilised in AI, DM is term common in database industry [19, 20].

In contrast, the present distance education network infrastructure can only handle preset instructional modes, not generative instructional modes. Teachers develop fixed instructional content and resources for all students, leading to concerns such as strict classroom objectives, course content, and instruction processes. Quantifying instructional quality assessment is challenging due to the fuzziness of data, making it a complex undertaking. English speaking assessment is a crucial component of college education, but evaluating its quality can be challenging, these drawbacks in the existing approaches motivated to do this work.

This study aims to improve an English speaking assessment approach using video data, this study will address how to optimize and improve evaluation and coding in the analysis of English speaking behavior using deep learning technology.

Major contribution of this work is brief as below,

- In this research work, ESA-NEGCN-NBOA is proposed.
- During preprocessing step, DAGIF technique is utilized for the spoken English video data.
- The pre-processed data is then used to extract features using New General Double Integral Transform (NGDIT).
- The obtained results of proposed ESA-NEGCN-NBOA algorithm is comparing to the existing models such as IST-OEA-ML, AS-SE-LPC-ML and MM-TL-OPA methods respectively.

The remaining manuscript is arranged as below: Part 2 outlines literature review, Part 3 displays proposed method, Part 4 presents outcomes with discussions, Part 5 concludes the manuscript.

II. LITERATURE REVIEW

The literature presents research projects on deep learning-based English speaking assessments; this section evaluated some of the most recent studies.

Zhao [21] has presented IST-OEA utilizing ML, DM, blockchain methods. Here, well EFL students learn orally utilizing video data samples, ML approach known as "discrete three-layered fuzzy logic ANN". The integrity system architecture for college students aligns with Blockchain Technology's decentralization, safety, trustworthiness, traceability, was widely employed in various fields. It collected and filtered spoken English data using normalization, median filtering, "Local Contrast Stretching" techniques. The feature extraction step uses multi-manifold discriminant analysis to extract important features from pre-processed data. It provides lower assessment error and higher evaluation time.

Wang [22] has presented ML-AS-SE depend on LPC. Here, analyses the idea of linear predictive coding, decoding, suggests employing hybrid excitation rather than binary excitation to improve existing algorithm. The machine learning assessment systems were organized into four modules likes acoustic method acquisition, speech recognition, standard pronunciation transcription, decision. The speech recognition module utilizes an updated linear predictive speech coding algorithm to extract speech signal features and build a feature vector. Acoustic method acquisition module was implemented by training speech features using the convolutional neural network algorithm. It provides lower evaluation time and higher assessment error.

Tun et al [23] has presented MM-TL-OPA. Here, how transfer learning between linked multimodal datasets can improve assessment of oral performance ability. These use knowledge from job interview dataset as pretraining material, apply learned knowledge from pre-trained method to modest quantity of performance data to develop learning of performance evaluation task. This strategy improves inference performance on tiny datasets, and publishes findings. Furthermore, compare the suggested TL methodology to a typical TL method using large-scale pre-trained method. It provides higher accuracy and higher assessment error.

Tong [24] has presented automatic assessment of dysarthric severity level utilizing audio-video cross-modal method in DL. Here, audio-video cross-modal framework were suggested for the first time utilizing a DL process in which network uses audio, video data as input to find dysarthria cruelty levels. DL technology includes two network topologies that use audio-only/video-only input to automatically find dysarthria cruelty levels. Compared to present one-modality schemes, DL framework produces satisfactory outcomes. It suggests

an audio-video deep-learning cross modal system for automatically assessing dysarthria severity levels, which can improve training speed compared to systems that simply use audio input. It provides lower assessment error and lower accuracy.

Liu et al. [25] has presented non-intrusive speech quality assessment depend on DNN for speech communication. Here, describes DL-depend strategy uses large-scale intrusive simulated data to increase accurateness, generalizability of non-intrusive techniques. Important contributions were listed below. Initially, introduces data simulation technique that makes degraded speech signals, assigns speech quality using perceptual objective listening quality assessment. The created data has been shown to valuable for pre-training DL methods. Secondly, suggests using adversarial speaker classifier to mitigate effect of speaker-dependent information in speech quality judgment. It provides higher accuracy, evaluation time.

Migliorelli et al. [26] has presented store-and-forward cloud-depend telemonitoring scheme for automatic evaluating dysarthria evolution in neurological illnesses from video-recording analysis. Here, describes a store-and-forward self-service telemonitoring system that incorporates CNN for interpreting video records captured by dysarthria patients into its cloud architecture. Facial landmark mask RCNN architecture identifies facial landmarks to assess or facial speech capabilities and study dysarthria evolution in neurological illnesses. The suggested CNN performed well on Toronto NeuroFace dataset, which includes annotated video records from ALS and stroke patients, by NMSE of 1.79 in identifying facial landmarks. These pilot studies were an important step toward utilize of remote apparatuses to help doctors track the progression of dysarthria. It provides higher posttest and higher assessment error.

Zhao et al. [27] has presented automatic evaluation of depression from speech via hierarchical attention transfer network with attention auto encoders. Here, quantitative mental health research was gaining interest in using behavioural signals, such as speech patterns, to accurately diagnose depression severity. Although machine learning approaches were widely used in depression analysis, a shortage of labeled data has hindered their wider implementation, including deep learning. Here present DL strategy uses unsupervised learning, knowledge transfer, and hierarchical attention to measure depression severity through speech. Then unique approach, HATN, employs hierarchical attention auto encoders to learn attention from source task. It provides higher retest and lower accuracy.

III. PROPOSED METHODOLOGY

In this sector, English Speaking Assessment Algorithm Based on Deep Learning (ESA-NEGCN-NBOA) deliberated. Block diagram of proposed ESA-NEGCN-NBOA method is in Figure 1. It covers such stages as Deep Attentional Guided Image Filtering, New General Double Integral Transform, Neighbor Enhanced Graph Convolutional Networks, Namib beetle optimization algorithm. Thus, detailed explanation about every steps given below,

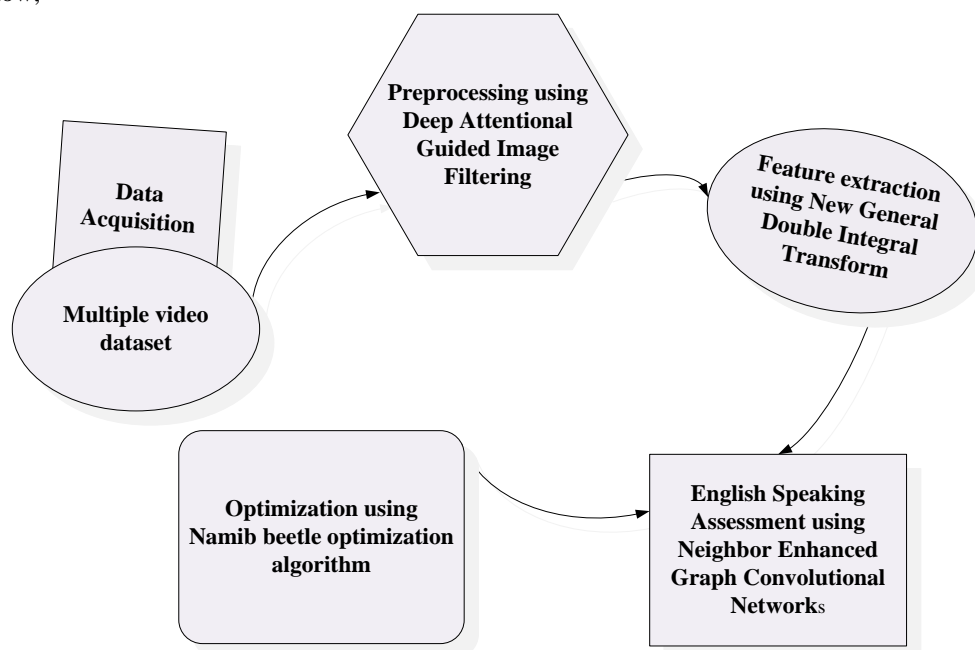


Figure 1: Block diagram of ESA-NEGCN-NBOA method

A. Data Acquisition

The input video data are gathered from multiple video dataset [28]. Video-Dataset allows loading video datasets in PyTorch quickly and efficiently. It makes working with video datasets simple and convenient. It simply demands that have video dataset on disk in a specific format; all else is handled automatically. There are no complicated dependencies, and it enables native Torch vision video data augmentation. Video-Dataset-Loading-Pytorch is the lowest entry barrier for creating deep learning training loops on video data.

B. Pre-Processing using Deep Attentional Guided Image Filtering

In this section, the pre-processing using Deep Attentional Guided Image Filtering (DAGIF) [29] is discussed. The input video data then pre-processed using DAGIF to remove presence of signal-dependent noise and improve lack of pixels from the regions and enhanced the video data. Attention methods enable method to selectively emphasis on exact regions in the input, which is useful for tasks in which some aspects are more important than others. The purpose of including attentional mechanisms may be to improve the model's sensitivity to important input features crucial to English speaking assessment, such as pronunciation patterns or intonation. The attention mechanism enables the model to focus on certain portions of data, emphasizing significant features and details while disregarding irrelevant information. Depending on the model's performance, it may be appropriate for real-time data processing applications that demand quick and adaptive filtering, resulting in cleaner and more effectively converted datasets for subsequent analysis and applications. Then the guidance video filtering is given in equation (1)

$$e_i = \sum_j W_{i,j}(g,t)t_j \quad (1)$$

Where, guidance image as g target image as t , output e of guided filtering, i and j denotes pixel coordinates; $W_{i,j}$ signifies filter kernel weight. In classical bilateral filter with guided image filter, $W_{i,j}$ is dependent on guidance g . The filter weight in bilateral filter is given in equation (2)

$$e_i = \sum_{j \in N_i} K(g_i, g_j)W[p_i - p_j]t_j + b \quad (2)$$

Where W denotes spatially invariant kernel; $K(.,.)$ denotes varying filter kernel function has fixed form likes Gaussian, $[p_i, p_j]$ signifies index offset of kernel weights. DAGIF has cleaned and transformed the data and it is shown in equation (3)

$$W(g,t)=K(g)\Theta K(t) \quad (3)$$

Where $K(g)$, $K(t)$ denotes kernel weights learned from guidance with target, Θ signifies element wise multiplication. This is resolute routinely by analyzing guide, target information, here removing signal-dependent noises, improving the pixels in regions and enhancing the video data is calculating in equation (4)

$$e_i = \sum_{j \in N_i} D_{i,j} W_{i,j}^g t_j + \sum_{j \in N_i} (1 - D_{i,j}) W_{i,j}^t t_j \quad (4)$$

Where $W_{i,j}^g$ and $W_{i,j}^t$ denotes filter kernels generated from guide, target. $D_{i,j}$, denotes pixel-wise dependability weight of guidance video. Finally, DAGIF removes the existence of signal-dependent disturbances, improves the absence of pixels in regions, and enhances the video data before subjecting the pre-processed data to feature extraction.

C. Feature Extraction Using New General Double Integral Transform

The feature extraction using NGDIT [30] is discussed. The feature extraction process proposes the NGDIT to extract the significant attributes such as mel-frequency cepstral coefficients, energy, speech rate and pitch; from the pre-processed data. The transform may provide a more effective method for representing acoustic data collected from spoken English, thereby capturing relevant patterns more precisely. The transform may be designed to increase sensitivity to tiny differences in speech, enabling a more nuanced evaluation of speaking proficiency. The NGDIT has increased processing speed, making it more resource-friendly and effective at solving mathematical problems. This improvement allows for the refinement of prediction models related to English speaking assessments, resulting in enhanced accuracy for anticipating upcoming academic influence

and English speaking exams. Inverse general transformation defines complex integral formula given in equation (5)

$$e(t) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \frac{1}{p(s)} e^{\varphi(s)t} F_D(s) \varphi^i(s) ds \tag{5}$$

Where, $e(t)$ denotes integrable function let $F_D(s)$ be general transform of $e(t)$. The function is integrable a denotes real constant, complex functions $p(s)$ and $q(r)$, NGDIT has extracted the features that extracted features are as follows: Mel-Frequency Cepstral Coefficients are calculated using a number of signal processing processes, involving the Fourier transform, filter financial institutions, and the logarithmic conversion. The general formula for MFCC computation is given in equation (6)

$$MFCC_i = DCT(\log(\sum_{l=1}^N M_j \cos[i(j - \frac{1}{2})\frac{\pi}{N}])) \tag{6}$$

Where M_j the magnitude spectrum of the signal, then DCT the Discrete Cosine Transform, \log to sum of magnitude spectrum components weighted by cosine functions, i and j are represents the Cepstral coefficients in a way that emphasizes the most relevant features. Pitch characteristics like as Prosodic and energy can be retrieved utilizing various signal processing methods. For pitch, one typical method is autocorrelation. The pitch is derived in equation (7)

$$f_0 = \frac{1}{\arg \max(Rxx)} \tag{7}$$

Where Rxx represents the autocorrelation function of each signal, f_0 fundamental frequency of speech signal. $\arg \max$, is the reciprocal of the lag corresponding to the maximum value in the autocorrelation function. Energy protein's total energy is calculated by adding the energies of all building units. Energy is calculated as the total of squared signal values over a frame. The energy formulated in equation (8)

$$Energy = \sum_{i=1}^N x(i)^2 \tag{8}$$

Where N represents the number of total signal strength and $x(i)^2$ computed as the sum of squared signal values. It indicates the overall strength of the spoken signal. Speech rate is commonly calculated as number of words spoken per minute. It entails counting the amount of words and dividing by the total speaking duration. The speech rate is given in equation (9)

$$Speech Rate = \frac{Number\ of\ words}{Speaking\ time\ (in\ minutes)} \tag{9}$$

It defines number of words spoken per unit of time. It gives information regarding the rate of speech. Thus, the above features are extracted by NGDIT and that output is fed in to English speaking assessment stage.

D. English Speaking Assessment using Neighbor Enhanced Graph Convolutional Networks

The English speaking assessment using NEGCN [31] is discussed. NEGCN is proposed to improve the student's spoken English performance by assessing the English speakers. NEGCNs can record localized associations between nodes in a graph, allowing the model to factor in the influence of surrounding nodes when making predictions. This might be useful for situations where the context of adjacent pieces is important. The purpose of employing NEGCN might be increase representation of features related to nodes in a graph, capturing nuanced patterns necessary for English speaking assessment. Then the pattern of assessment is given in equation (10)

$$h^{(l+1)} = \sigma \left(FC\theta_{(l)} \left(\sum_{u \in N_v^+} w_{v,u}^{(l)} + \sum_{u \in N_v^-} w_{v,u}^{(l)} h_u^{(l)} \right) \right) \tag{10}$$

Where, central node (v), positive neighbour set N_v^+ , negative neighbour set (N_v^-) and σ is the sigmoid function. Simple Graph convolution (SGC) does not require nonlinearity between NEGCN layers. By reducing non-linearity, NEGCN can achieve performance comparable to SGC and 10 times quicker. Non-linearity, rewrite updating function of node has given in equation (11)

$$h_v^{(l)} = \frac{1}{n_v} \left(\sum_{u \in N_v^+} h_u^{(l-1)} + \sum_{u \in N_v^-} h_u^{(l-1)} \right) \tag{11}$$

Where, $h_v^{(l)}$ and $h_u^{(i)}$ are the function node, and n_v is the node of v . Using linear updating function, embedding's from different convolution times are located in similar hidden space, Similar to one vs. all classification in multi-class forecast, multi-class node classification issue to binary classification. It is shown in equation (12)

$$F(h_v^{(l)}) = \frac{1}{n_v} \left(\sum_{u \in N_v^+} F(h_u^{(l-1)}) + \sum_{u \in N_v^-} F(h_u^{(l-1)}) \right) \tag{12}$$

Where, $F(h_v^{(l)})$ denotes as mapping, Using the expectation formula, It can increase the probability of correctly classifying node v by increasing the positive ratio r_v . This observation leads to consider the following equation (13)

$$E_{origin} = \frac{1}{n_v} (n_v^+ \mu^+ + n_v^- \mu^-) = r_v \mu^+ + (1 - r_v) \mu^- \tag{13}$$

Where, E_{origin} denotes as the origin graph structure, node v by increasing the positive ratio r_v , here NEGCN is improving the English speaking assessment. It is shown in equation (14)

$$E_{filter} = \frac{p.n_v^+ \mu^- + q.n_v^- \mu^-}{p.n_v^+ + q.n_v^-} \tag{14}$$

Where, E_{filter} denotes as the filtering process, predicted positive neighbours that indeed negative $p.n_v^+ \mu^-$. Finally NEGCN assessment the English speaking in high performance. In this work, NBOA is employed to optimize the NEGCN optimum parameters F and C . Here NBOA is employed for turning the weight and bias parameter of NEGCN.

E. Optimization using Namib Beetle Optimization Algorithm

The optimization using NBOA [32] is discussed. Here, NEGCNs weight parameters F and C are optimized using NBOA. Optimization methods assist in efficiently tweaking the parameters of deep learning models. This is critical for reaching peak model performance in tasks like speech recognition and natural language processing (NLP), which are critical components of the English speaking evaluation. NBOA, like similar optimization algorithms, can be used to identify optimal neural network topologies. This includes deciding on layers, amount of neurons in all layer, other structural factors that can improve the model's performance in English speaking tests; it provides a general overview of how optimization algorithms often use parameters and weights. Especially, the absence of a transfer parameter during transition from exploration phase to exploitation phase directly influences the algorithm's performance. The initiation of NBO involves the initialization step.

1) Stepwise process of NBO

Here, stepwise process defined to get ideal value of NEGCN based on NBO. Initially, NBO makes the equally distributing populace to optimize parameters F and C of NEGCN. Ideal solution promoted using NBO algorithm.

Step 1: Initialization

All solution to issue is regarded beetle, may be encoded in D dimensions. It is given in equation (15)

$$Pop = \begin{bmatrix} NB_{1,1} & NB_{1,2} & \dots & NB_{1,D} \\ NB_{2,1} & NB_{2,2} & \dots & NB_{2,D} \\ \dots & \dots & \dots & \dots \\ NB_{N,1} & NB_{N,1} & \dots & NB_{N,D} \end{bmatrix} \tag{15}$$

Where, NB is denoted as beetle, D is a decision variable, x is the initial population of beetles.

Step 2: Random generation

Input parameters made at randomly. Best fitness value selection is depending upon obvious hyper parameter condition.

Step 3: Fitness function

The outcome is determined by initialized judgments and random responses. The fitness is then computed using the equation (16)

$$Fitness\ Function = Optimizing [F\ and\ C] \tag{16}$$

Step 4: Exploration Phase for suitability of all area for water collection

All beetle is initially placed in randomly generated issue area, assessed utilizing objective function. Beetles with higher objective function values exhibit greater moisture and water gathering capabilities. As a result, placing a beetle in an optimal region makes it more successful, and this optimal area may act as a magnet for other beetles. This magnetism can be calculated using equation (17)

$$P_j = P_{max} \cdot \sin\left(\frac{s(NB_j) - s_{min}}{s_{max} - s_{min}} \cdot \frac{\pi}{2}\right) \tag{17}$$

Where, P_j represents the capacity of the area hosting the beetle, NB_j denoting the number of beetles, P_{max} indicates the maximum capacity of beetles in a single area, P_{min} denotes the minimum capacity of beetles in single region. The competence of beetle NB_j is represented by $f(NB_j)$ with s_{min} , s_{max} representing the minimum and maximum skills within the population of beetles. Flowchart of NBOA for optimizing NEGCN parameter is shown in figure 2.

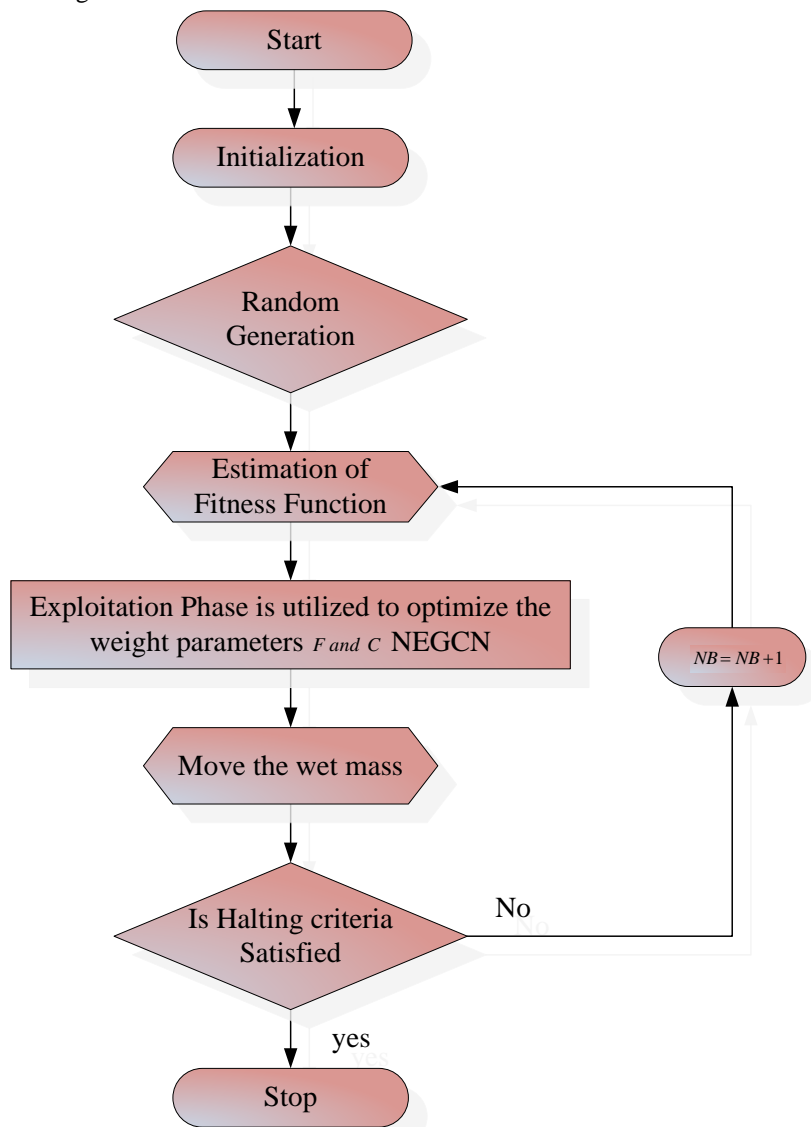


Figure 2: Flowchart for Namib Beetle Optimizing NEGCN

Step 5: Exploitation phase for optimizing F and C

The exploitation process, moving the wet mass and the Beetles uses their sense of smell to identify areas with higher moisture levels. To replicate this behaviour, the model includes the beetles' center of gravity as well as the possibility of meeting in damp places. Furthermore, the beetles explore space among gravity center, most ideal solution, described in equation (18)

$$NB_i^{new} = NB_i^{old} + rand.(NB^* - \overline{NB}) + levy \tag{18}$$

Where, NB^* denotes as location, \overline{NB} represents the position of the water gravity, current with novel locations of beetle means to move NB_i^{old} and NB_i^{new} , the position of a beetle that attracts other beetles indicates NB_i and $levy$ denotes random vector.

Step 6: Termination Condition

The weight parameter values F and C of generator from Neighbour Enhanced Graph Convolutional Network is optimized by support of NBO, iteratively repeat the step 3 until fulfil halting conditions $NB = NB + 1$ is met. Then NEGCN has assessed the English speaking with greater accuracy by lessening evaluation time with error.

IV. RESULT AND DISCUSSION

Experimental results of ESA-NEGCN-NBOA approach have been Assessed the English speaking. The proposed work is performed in Python on an Intel(R) core(TM) i7 CPU M60 @ 2.80 GHz and evaluated using a variety of performance metrics such as accuracy, evaluation time, pretest, posttest, and assessment error. The proposed ESA-NEGCN-NBOA methodology is compared to other methods like IST-OEA-ML, AS-SE-LPC-ML and MM-TL-OPA.

A. Performance measures

This is a crucial step for determining the exploration of optimization algorithm. Performance measures to evaluate to access performance such as accuracy, evaluation time, pretest, posttest, and assessment error.

1) Accuracy

Accuracy refers to the ability to measure a precise value. A statistic known as accuracy can be utilized to assess method's performance across all classes. It is quantified by the following equation (19)

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{19}$$

Where TP denotes true positive, TN represents true negative, FP and FN denotes false positive and false negative.

2) Evaluation Time

The evaluation time is the time required to examine and score a specific English speaking challenge using deep learning models. It can be quantified in a variety of ways, including average time per assessment and total time for a batch of assessments. The Evaluation time is given in equation (20)

$$EvaluationTime = End\ Time - Start\ Time \tag{20}$$

3) Pretest

Pretesting is an assessment measure given to volunteers prior to having undergone any form of counselling as part of a research project. The Pretest formula is given as equation (21)

$$pretest = Mean \frac{\sum Scores}{Number\ of\ Assessments} \tag{21}$$

4) Posttest

A posttest is an assessment measurement administered to participants after counselling as part of a research study. The Posttest formula is given as equation (22)

$$Posttest = Mean \frac{\sum Scores}{Number\ of\ Assessments} \tag{22}$$

5) Assessment Error

English speaking assessments involve continuous rating, i.e., assessing fluency on a scale of 1 to 5, as well as the use of MSE as a measure of assessment error. The Assessment error is given in equation (23)

$$Assessment\ Error = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \tag{23}$$

B. Performance analysis

The simulation results of ESA-NEGCN-NBOA technique are shown in Figure 3 to 7. The ESA-NEGCN-NBOA techniques analysed with existing techniques likes IST-OEA-ML, AS-SE-LPC-ML, MM-TL-OPA.

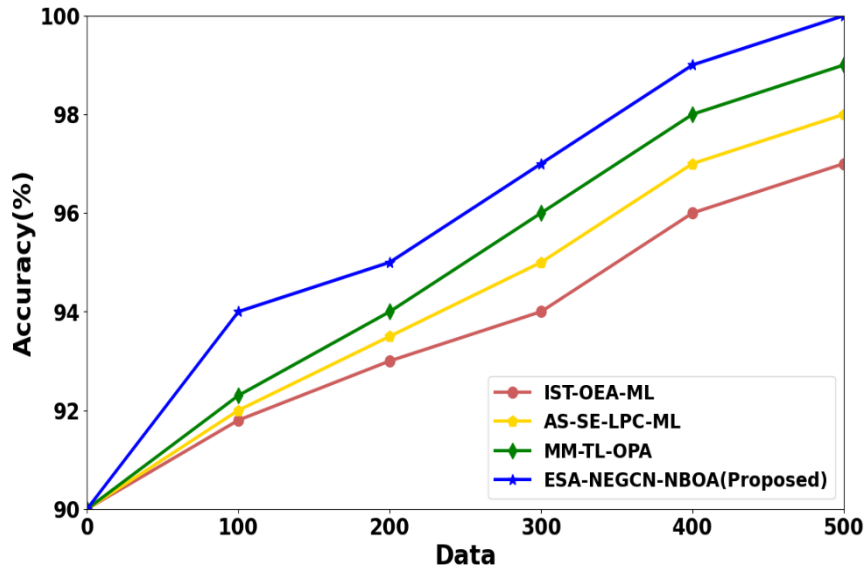


Figure 3: Accuracy analysis

Figure 3 shows accuracy analysis. The ESA-NEGCN-NBOA technique reaches in the range of 22.36%, 25.42% and 18.27% higher accuracy at data 100; 27.26%, 20.41% and 23.26% higher accuracy at data 200; 18.20%, 20.19%, 30.02% higher accuracy at data 300; 22.06%, 15.26%, and 21.19% higher accuracy at data 400 20.06%, 17.26%, and 25.19% higher accuracy at data500; analysed with existing techniques likes IST-OEA-ML, AS-SE-LPC-ML and MM-TL-OPA respectively.

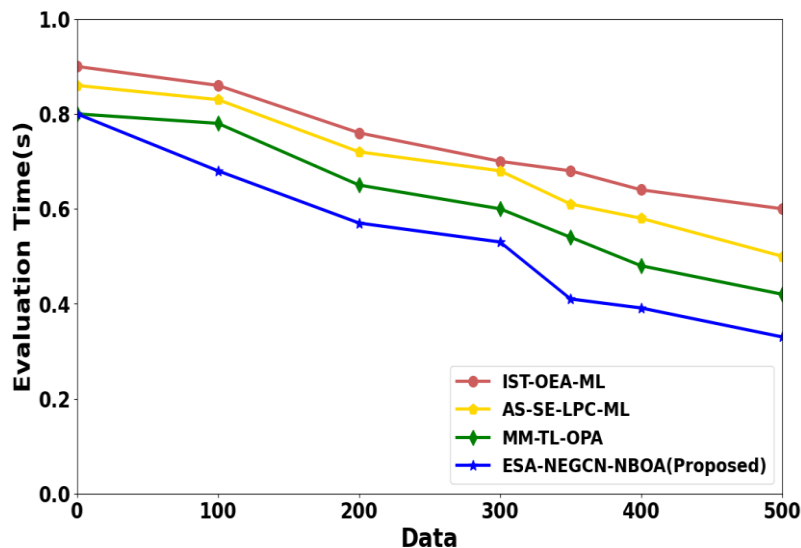


Figure 4: Evaluation time analysis

Figure 4 shows evaluation time analysis. The ESA-NEGCN-NBOA technique reaches in the range of 23.36%, 18.42% and 29.25% lower evaluation time at data 100; 29.26%, 15.41% and 26.26% lower evaluation time at data 200; 28.20%, 30.19%, 17.02% lower evaluation time at data 300; 25.06%, 19.26%, and 23.19% lower evaluation time at data 400; 32.06%, 17.26%, and 28.19% lower evaluation time at data 500; analysed with existing techniques likes IST-OEA-ML, AS-SE-LPC-ML and MM-TL-OPA respectively.

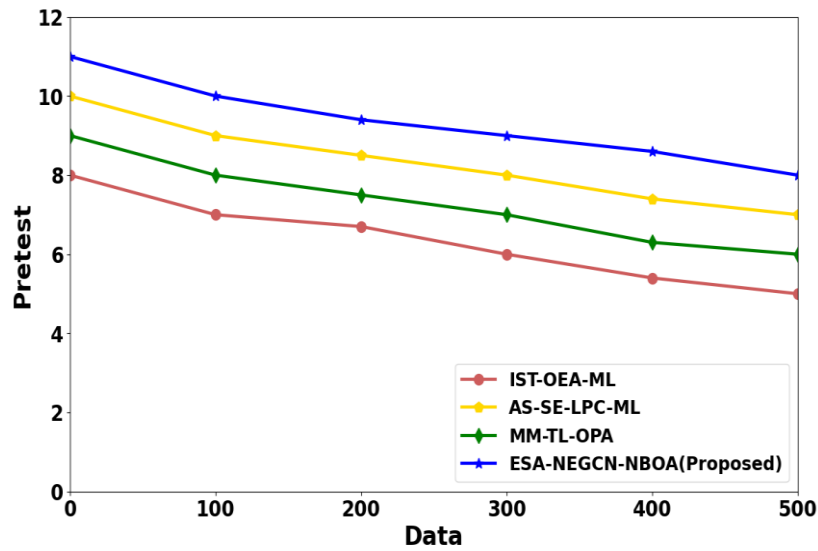


Figure 5: Pre-test analysis

Figure 5 shows pre-test analysis. The ESA-NEGCN-NBOA technique reaches in the range of 33.36%, 28.42% and 19.25% higher Pretest at data 100; 22.26%, 15.41% and 25.26% higher Pretest at data 200; 28.20%, 20.19%, 19.02% higher Pretest at data 300; 22.06%, 20.26%, and 21.19% higher Pretest at data 400; 31.06%, 18.26%, and 24.19% higher Pretest at data 500; analysed with existing techniques likes IST-OEA-ML, AS-SE-LPC-ML and MM-TL-OPA respectively.

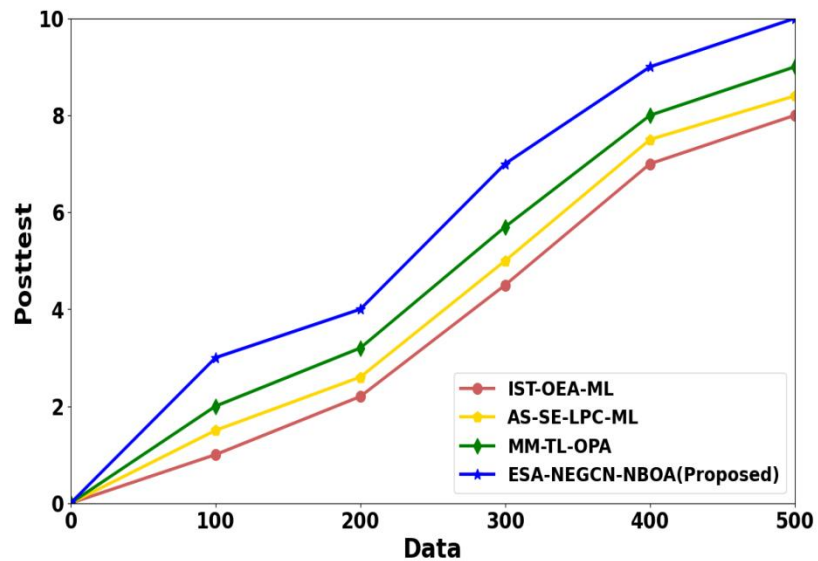


Figure 6: Post-test analysis

Figure 6 shows post-test analysis. The ESA-NEGCN-NBOA technique reaches in the range of 20.36%, 28.42% and 19.25% higher Posttest at data 100; 21.26%, 20.41% and 27.26% higher Posttest at data 200; 18.20%, 30.19%, 20.02% higher Posttest at data 300; 22.06%, 15.26%, and 21.19% higher Posttest at data 400; 29.06%, 17.26%, and 25.19% higher Posttest at data 500; analysed with existing techniques likes IST-OEA-ML, AS-SE-LPC-ML and MM-TL-OPA respectively.

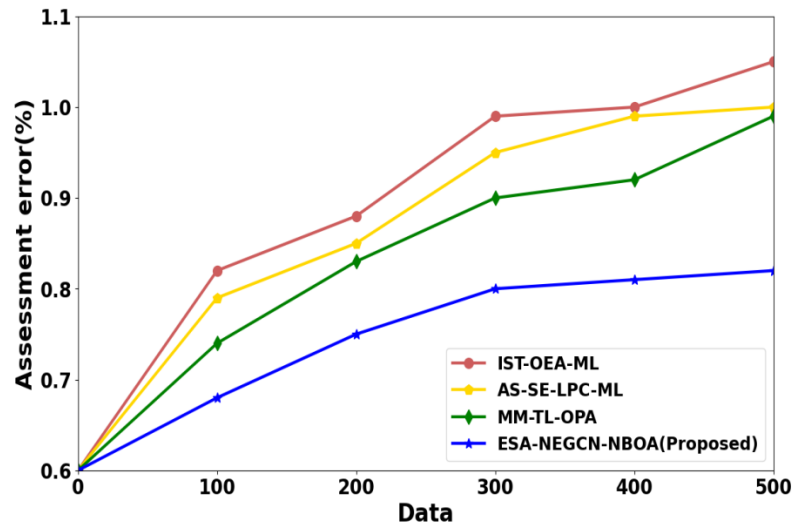


Figure 7: Assessment error analysis

Figure 7 shows assessment error analysis. The ESA-NEGCN-NBOA technique reaches in the range of 23.36%, 18.42% and 29.25% lower assessment error at data 100; 26.26%, 25.41% and 27.26% lower assessment error at data 200; 18.20%, 31.19%, 27.02% lower assessment error at data 300; 22.06%, 15.26%, and 21.19% lower assessment error at data 400; 33.06%, 27.26%, and 24.19% lower assessment error at data 500; analysed with existing techniques likes IST-OEA-ML, AS-SE-LPC-ML and MM-TL-OPA respectively.

C. Discussion

In this paper, proposed a NEGCN technique as a first stage for English speaking assessment. Training data doesn't comprise information about object's location. In a NEGCN network, each cell must include four multilayer perceptron. While computing linear layers, system necessitates significant amount of memory bandwidth. However, this bandwidth is insufficient to power all compute units. Because there is inadequate data to train NEGCN, it can't be utilized density deterioration has a major impact on system performance. The NGDIT has only one notable downside. Pre-processed blocks frequently have integer input video data values and real output video data values. Education is further just "learning information" as NEGCN technology makes it tougher to build students moral character, increase total level of proficiency in abilities such as self-assurance, communication, thinking critically, sharing. Teachers, on the other hand, may utilize DL as sensor-kind educational instrument to additional properly measure their student's comprehension of the subject. For example, using oral evaluation software mentioned above, instructors can build further individualized learning plans for students depend on understanding of their student's weak parts in spoken English learning. Then look how to use video data, DL to evaluate capable spoken English utilizing NEGCN. Investigate relies on a deep learning knowledge base. MVD collects video data. NGDIT and the proposed and NEGCN both examine accuracy, assessment error, evaluation time, pre-test, post-test of self-learning. Such results are represented as graphs utilizing origin tool.

V. CONCLUSION

In this paper, English Speaking Assessment Algorithm Based on Deep Learning (ESA-NEGCN-NBOA) was successfully implemented. Here, multi video dataset (MVD) were used in thorough evaluation tests to assess the presented technique. The proposed ESA-NEGCN-NBOA method is executed in Python. The presentation of proposed ESA-NEGCN-NBOA method covers 22.36%, 25.42% and 18.27% higher accuracy, 22.36%, 25.42% and 28.27% lower evaluation error, and 22.36%, 15.42% and 18.27% lower evaluation time compared with existing IST-OEA-ML, AS-SE-LPC-ML and MM-TL-OPA methods. Furthermore, many improvements will be made in the future, this proposed ESA-NEGCN-NBOA model, future of English speaking assessments further focused due to usage of English speaking evaluation technologies, considerably increase student learning, testing efficacy.

Acknowledgement

1. Research and Practice on the Path of Improving the Ideological-political Teaching Ability of Higher Vocational English Teachers (Jiangsu Maritime Institute Key Project, Project No.: xjzd2020-4).
2. "Qing Lan Project" in Jiangsu Province

REFERENCE

- [1] Cui, Y., & Li, H. (2022). Evaluation System of Mobile English Learning Platform by Using Deep Learning Algorithm. *Mobile Information Systems*, 2022.
- [2] Kiriakopoulos, K. (2022). *Deep learning for automatic assessment and feedback of spoken English* (Doctoral dissertation, University of Cambridge).
- [3] Song, C. (2022). An Evaluation Method of English Teaching Ability Based on Deep Learning. *Security and Communication Networks*, 2022, 1-9.
- [4] Wang, C. (2022). English Process Assessment Based on Deep Learning and Cloud Computing. *Wireless Communications and Mobile Computing*, 2022.
- [5] Xu, Y. (2022). English speech recognition and evaluation of pronunciation quality using deep learning. *Mobile Information Systems*, 2022, 1-12.
- [6] Shi, J. (2022). Deep Learning for College English Education Evaluation. *Mobile Information Systems*, 2022.
- [7] Zhang, F., & Zhou, Z. (2023). Design of English pronunciation quality evaluation system based on the deep learning model. *Applied Mathematics and Nonlinear Sciences*, 8(2), 2805-2816.
- [8] Cheng, J., Chen, X., & Metallinou, A. (2015). Deep neural network acoustic models for spoken assessment applications. *Speech Communication*, 73, 14-27.
- [9] Wang, Y. (2021). Detecting pronunciation errors in spoken English tests based on multifeature fusion algorithm. *Complexity*, 2021, 1-11.
- [10] Zhao, X., & Jin, X. (2022). Standardized evaluation method of pronunciation teaching based on deep learning. *Security and Communication Networks*, 2022.
- [11] Qi, S., Liu, L., Kumar, B.S., & Prathik, A. (2022). An English teaching quality evaluation model based on Gaussian process machine learning. *Expert Systems*, 39(6), e12861.
- [12] Li, Y. (2022). Deep Learning-Based Correlation Analysis between the Evaluation Score of English Teaching Quality and the Knowledge Points. *Computational Intelligence and Neuroscience*, 2022.
- [13] Han, X., & Huwan, T. (2022). The modular design of an English pronunciation level evaluation system based on machine learning. *Security and Communication Networks*, 2022.
- [14] Wang, Y. (2022). Construction of Intelligent Evaluation Model of English Composition Based on Machine Learning. *Mobile Information Systems*, 2022.
- [15] Zafari, M., Sadeghi-Niaraki, A., Choi, S.M., & Esmaeily, A. (2021). A practical model for the evaluation of high school student performance based on machine learning. *Applied Sciences*, 11(23), 11534.
- [16] Mahmoud, S.S., Kumar, A., Li, Y., Tang, Y., & Fang, Q. (2021). Performance evaluation of machine learning frameworks for aphasia assessment. *Sensors*, 21(8), 2582.
- [17] Zhao, S. (2022). Evaluation of English Subjective Questions Based on Deep Neural Networks. *Scientific Programming*, 2022.
- [18] Jin, Y. (2022). Sentiment analysis in british and american literature teaching under formative assessment and machine learning. *Wireless Communications and Mobile Computing*, 2022.
- [19] Vásquez-Correa, J.C., Arias-Vergara, T., Orozco-Arroyave, J.R., Eskofier, B., Klucken, J., & Nöth, E. (2018). Multimodal assessment of Parkinson's disease: a deep learning approach. *IEEE journal of biomedical and health informatics*, 23(4), 1618-1630.
- [20] Xu, W. & Zhang, W. (2023). Assessment Algorithm based on Deep Learning for Psychological Capital and Job Performance of Kindergarten Teachers.
- [21] Zhao, C. (2023). An Innovative Strategy Towards Oral English Assessment Using Machine Learning, Data Mining, and Blockchain Techniques.
- [22] Wang, L. (2022). A Machine Learning Assessment System for Spoken English Based on Linear Predictive Coding. *Mobile Information Systems*, 2022.
- [23] Tun, S.S.Y., Okada, S., Huang, H.H., & Leong, C.W. (2023). Multimodal Transfer Learning for Oral Presentation Assessment. *IEEE Access*.
- [24] Tong, H. (2020). Automatic assessment of dysarthric severity level using audio-video cross-modal approach in deep learning (Master's thesis).

- [25] Liu, M., Wang, J., Wang, F., Xiang, F., & Chen, J. (2023). Non-Intrusive Speech Quality Assessment Based on Deep Neural Networks for Speech Communication. *IEEE Transactions on Neural Networks and Learning Systems*.
- [26] Migliorelli, L., Berardini, D., Cela, K., Coccia, M., Villani, L., Frontoni, E., & Moccia, S. (2023). A store-and-forward cloud-based telemonitoring system for automatic assessing dysarthria evolution in neurological diseases from video-recording analysis. *Computers in Biology and Medicine*, 163, 107194.
- [27] Zhao, Z., Bao, Z., Zhang, Z., Deng, J., Cummins, N., Wang, H., Tao, J., & Schuller, B. (2019). Automatic assessment of depression from speech via a hierarchical attention transfer network and attention autoencoders. *IEEE Journal of Selected Topics in Signal Processing*, 14(2), 423-434.
- [28] <https://www.kaggle.com/discussions/general/275710>
- [29] Zhong, Z., Liu, X., Jiang, J., Zhao, D., & Ji, X. (2023). Deep attentional guided image filtering. *IEEE Transactions on Neural Networks and Learning Systems*.
- [30] Meddahi, M., Jafari, H., & Yang, X.J. (2022). Towards new general double integral transform and its applications to differential equations. *Mathematical Methods in the Applied Sciences*, 45(4), 1916-1933.
- [31] Chen, H., Huang, Z., Xu, Y., Deng, Z., Huang, F., He, P., & Li, Z. (2022). Neighbor enhanced graph convolutional networks for node classification and recommendation. *Knowledge-Based Systems*, 246, 108594.
- [32] Chahardoli, M., Eraghi, N.O., & Nazari, S. (2022). Namib beetle optimization algorithm: a new meta-heuristic method for feature selection and dimension reduction. *Concurrency and Computation: Practice and Experience*, 34(1), e6524.s