ImCfO_ Attn_EffBNet: Improved Crayfish Optimization based Hybrid Deep Learning for Parkinson Disease Prediction

Abstract - Parkinson's disease prediction plays a crucial role in early diagnosis and intervention, potentially improving patient outcomes. In this research, a novel approach for Parkinson's Disease prediction utilizing publicly available databases is introduced. The proposed method involves initial data preprocessing using a band-pass filter, followed by feature extraction via Empirical Mode Decomposition (EMD). These features are then fed into an Improved Crayfish Optimization based Attention based Efficient Bidirectional Network (ImCfO_Attn_EffBNet) for classification. ImCfO_Attn_EffBNet integrates EfficientNet-B7, BiLSTM, and Attention modules to efficiently collect both temporal and spatial information. Moreover, we employ the Improved Crayfish Optimization (ImCfO) algorithm for optimizing the loss function, enhancing convergence rates, and obtaining global best solutions. ImCfO incorporates a self-adaptation criterion into the conventional crayfish algorithm, leading to improved performance. The resulting solution from ImCfO is utilized to fine-tune the classifier's adjustable parameters, improving overall prediction accuracy. The performance evaluated by the ImCfO_Attn_EffBNet based on various assessments acquired the outcome of accuracy (95.068%), Recall (92.948%), Specificity (92.89%), F-Score (92.89%), Precision (92.89%), and FPR (2.1%) respectively.

Keywords - Parkinson disease prediction, EfficientNet-B7, Improved Crayfish Optimization, Electroencephalography, hybrid classifier, Attention mechanism

1. Introduction

A neurological disorder that affects the central nervous system is Parkinson's disease (PD), is growing more common among the ageing population. In Europe, it impacts 1.2 million individuals, and experts estimate that number will have doubled by 2030 [1]. A precise diagnosis is essential for bettering PD therapy and follow-up, as well as for differentiating the condition from other neurological disorders and healthy individuals [2]. The gold standard for diagnosing PD and tracking symptoms is still clinical examination, despite the fact that a number of criteria and suggestions have been proposed to aid in the process [3, 4]. The accuracy range for clinical evaluation, which takes into account several subjective aspects, is 75% to 82%. The diagnosis of Parkinson's disease depends on the presence of resting tremor, stiffness, and bradykinesia. Consequently, PD presents with a gradual development of symptoms, and by the time it is identified, brain damage has advanced considerably [6]. In this context, clinical features that predate motor symptoms may be useful. Apart from coexisting with, but often preceding, the onset of motor characteristics, non-motor indicators such as depressive

symptoms, visual impairment, cognitive decline, sleep disorders, olfactory dysfunction, orautonomic symptoms are increasingly recognised [7]. An increasing number of people are interested in utilising this range of premotor symptoms to identify illness.

It might be challenging to determine the clinical diagnosis of illnessby considering the symptoms. Clinicians have utilised many neuroimaging measuring approaches to diagnose PD [8, 9]. Magnetic resonance imaging or Computed tomography types are the bases for some of these procedures, while physiological signals like electromyography and EEG are the basis for others [10]. With a high temporal resolution and no invasiveness, electroencephalography (EEG) captures the electrical activity of the brain's pyramidal neurons to provide an inferred understanding of their function. It is a widely accessible, low-cost approach that has been extensively utilised to research epileptic diseases [11].Although information processing techniques enable the extraction of several properties that might be useful for characterising neurological illnesses, visual EEG analysis remains challenging [12, 13]. Dynamic information on electrical brain activity and connections may be obtained from EEG data because of its high temporal resolution [14]. EEG has therefore been monitored in a variety of physiological settings, including basic wakefulness, sleep, particular sensitive input, cognitive processes, and, lastly, resting state. Different networks should engage under each of these circumstances, and in the resting state, spontaneous connection should occur [15].Two key features characterise the EEG data and complicate any further analysis in addition to the widely recognised inter- and intra-subject variability. Machine learning techniques may be used to assess the EEG signals and overcome these challenges since they are powerful methods that enable dealing with raw EEG data and conducting non-linear investigations [16].

The development of automated diagnostic systems mostly uses machine learning techniques on EEG data. Non-stationary EEG waves can be challenging to evaluate due to their intricate nature. Applying features derived from different nonlinear feature extraction techniques makes it simple to classify PD -healthy control (HC) groups in order to facilitate these studies [17]. Furthermore, by breaking down EEG signals into smaller bands using a variety of techniques, comprehensive data may be acquired at distinct frequencies. This approach can lead to success in research using machine learning for the automated identification of neurological diseases [18]. These investigations enable physicians to make decisions in a more methodical manner. The signal processing techniques used aids in the automated identification of PD [19] through deep learning. However, a number of issues still impact how well ML-based algorithms can predict [20]. Therefore, this research introduces an optimised deep learning model to address the problem. The major contributions of the research are:

- ***Design of ImCfO Algorithm***: The proposed ImCfO algorithm is designed by integrating the self-adaption criterion with the conventional crayfish algorithm for enhancing the convergence rate and global best solution. The ImCfO algorithm is utilized for fine-tuning the parameters of classifier for optimizing the loss function.
- ***Design of Attn_EffBNet Model***: In the Attn_EffBNet, the EfficientNet-B7, BiLSTM and Attention modules are integrated for enhancing the classification accuracy through capturing spatial and temporal features. Besides, the adjustable parameters of the classifier are fine-tuned using the proposed ImCfO algorithm.

The research is structured as follows: Section 2 lists relevant works together with the problem description, and Section 3 has a complete proposed approach. Section 4 presents the results of the experiment, and Section 5 presents the conclusion.

2. Related Works

The existing methods concerning the PD classification is reviewed in this section. A technique for classifying PD in patients using deep learning methods was designed by [21] based on ensemble of Convolutional Neural Networks (CNNs), specifically ResNet models, for analyzing spiral and wave form data. In this, Pre-trained ResNet models are fine-tuned on the PD data to leverage existing knowledge from larger datasets. The model's performance was evaluated, with a reported accuracy of 96.67% using the ResNet50 model for spiral waveform. The study likely involves a limited dataset, raising concerns about generalizability, which was a challenging aspect of the model.

In order to increase the precision of diagnosing Parkinson's illness, [22] created an optimized deep learning model. Relevant information such as kinematic features from hand motions or acoustic features from voice recordings, were retrieved in this case once the pre-processing procedures were completed. The Quantum Mayfly Optimization (QMO) method was then used to choose the best attributes. For the purpose of classifying diseases, a deep learning model was created especially for examining data with spatial patterns. In this instance, the classifier used was a recurrent neural network, which is beneficial for analysing speech or voice patterns since it can detect temporal relationships in data. In this, QMO-based feature selection helps reduce the data dimensionality, potentially improving model training efficiency. Still, the time complexity of the model was higher that limits the applicability of the designed model in real world applicability.

Electroencephalography (EEG) data were used by [23] to create a machine learning model for the early and automated identification of PD. Here, EEG waves were broken down into smaller

bands using a signal processing method called Kalman filtering, which may have the ability to uncover hidden patterns linked to PD. Relevant traits were retrieved from the decomposed sub-bands that might help distinguish between individuals with Parkinson's disease and those in good health. Statistically significant features were found using the Chi-squared test, which may have reduced data complexity and enhanced model performance.

A hybrid model based on XGBoost with Random Forest was designed by [24] for disease prediction and classification. The imbalance of data was solved through the data sampling approach for the reduction of biased outcome. The designed approach could potentially provide a tool to aid healthcare professionals in early and more accurate identification. In this, the use of a relatively small, publicly available dataset raises concerns about generalizability to real-world populations. Training XGBoost models on large-scale datasets or with complex parameter configurations requires significant computational resources and time, which was considered as limitation in resource-constrained environments or when dealing with real-time applications.

For the purpose of early PD diagnosis, [25] created an advanced deep learning model. The study introduces a novel method using affine transformation to improve the alignment of MRI scans, which ensures images were in the same orientation, crucial for accurate analysis by deep learning models. A deep learning architecture specifically designed for image segmentation was used to identify and isolate potentially Parkinson's disease-affected brain regions within the MRI scans. Finally, using the DenseNet Model disease classification was devised. Here, it analyzes the segmented brain regions from the U-Net model to differentiate between PD and non- PD cases. In this, improved image alignment reduces artifacts and improves the effectiveness of deep learning models. Higher computational complexity of the model was considered as the challenging aspect.

2.1. Problem Statement

PD prediction holds immense importance for early diagnosis and intervention, crucial in managing the condition effectively. Its application domains span various fields, including healthcare, where accurate prediction aids in timely treatment planning and monitoring of disease progression. Existing methods for PD prediction encompass diverse approaches, such as deep learning models, optimized deep learning models employing Quantum Mayfly Optimization (QMO) for feature selection, machine learning models utilizing Electroencephalography (EEG) signals with techniques like Kalman filtering and algorithms like Support Vector Machine (SVM), and hybrid models combining XGBoost with Random Forest for classification. Challenges faced by these methods include limited dataset sizes

affecting generalizability, high time complexity hindering real-world applicability, data imbalance issues, and concerns regarding model scalability and computational resources.

The proposed novel Improved Attention based Efficient Bidirectional Network (Attn_EffBNet) model with Improved Crayfish Optimization Algorithm (ImCfO) based loss function optimization addresses several challenges encountered in existing methods for PD prediction. Firstly, by incorporating attention mechanisms and bidirectional network architecture, the model effectively captures and prioritizes relevant features while considering temporal dependencies within the data, thus enhancing its ability to generalize across different patient populations. Additionally, the utilization of ImCfO for loss function optimization enables efficient navigation of high-dimensional parameter spaces, overcoming the curse of dimensionality prevalent in PD classification tasks. Moreover, the model's optimization approach enhances training efficiency and effectiveness. Furthermore, the Attn_EffBNet model's capability to accurately analyze and classify data, coupled with its efficient computational architecture, ensures improved diagnostic accuracy and practical applicability in clinical settings for early and accurate PD diagnosis.

3. Proposed Methodology

The proposed Parkinson Disease Prediction acquires the data for processing the input from publically available database and then pre-processed using the band-pass filter. Then, the essential features are acquired through empirical mode decomposition (EMD) technique, which is fed into the proposed Improved Crayfish Optimization based Attention based Efficient Bidirectional Network (ImCfO_Attn_EffBNet) technique. In this, Attn_EffBNet is designed by integrating EfficientNet-B7, BiLSTM and Attention modules are integrated for enhancing the classification accuracy through capturing spatial and temporal features. Furthermore, the ImCfO method is utilized for the optimization of the loss function. In order to improve the global best solution and convergence rate, the self-adaption criteria is integrated with the traditional crayfish method to create the suggested ImCfO algorithm. Figure 1 illustrates the suggested method's process.

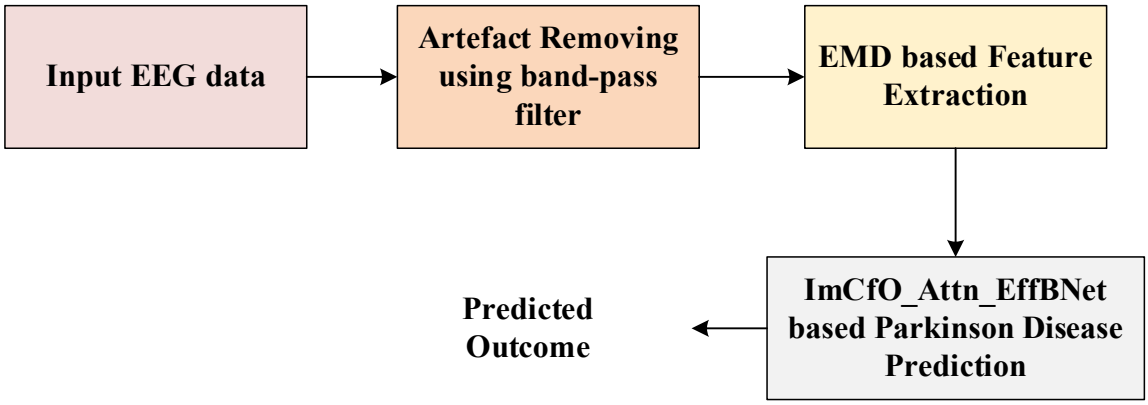


Figure 1. Workflow of proposed Parkinson Disease Prediction

3.1. Data Acquisition

The EEG data was gathered from the publicly accessible Kaggle Database [26], which contains the EEG signals of 31 patients with PD as well as normal people.

3.2. Pre-Processing using band-pass filter

The essential features from the EEG signals are removed through pre-processing technique based on filtering process. When a signal is processed through a band-pass filter, frequencies outside of the pass-band are attenuated and only a specific range of frequencies, called the pass-band, are allowed to pass through. While considering the EEG signal processing, a band-pass filter is used to isolate specific frequency bands of interest, such as delta, theta, alpha, beta, or gamma waves.EEG signals are composed of various frequency components that represent different brain activities. Filtering involves separating these frequency components into distinct bands to extract relevant information.

Delta Waves (0.5-4 Hz):The slowest brain waves, or delta waves, are generally linked to deep sleep stages, such non-rapid eye movement sleep stages 3 and 4. Additionally, they exist in some unconscious states. Characterized by a slow frequency and large amplitude, delta waves are believed to contribute to the restorative effects of sleep, such as memory consolidation and physical restoration.

Theta Waves (4-8 Hz): These bands are slightly faster than delta waves and are often observed during states of drowsiness, relaxation, or deep meditation. They are also prominent during REM (rapid eye movement) sleep, which is associated with dreaming. Theta waves are implicated in memory formation, emotional processing, and creative thinking. Increased theta activity has been linked to cognitive deficits in conditions such as Alzheimer's disease.

Alpha Waves (8-13 Hz): These bands are prominent during wakeful relaxation with closed eyes and are typically observed over the occipital region of the brain. They are associated with

a state of calmness, relaxation, and mental alertness. Alpha waves are often attenuated by opening the eyes or engaging in cognitive tasks. Abnormalities in alpha wave activity have been observed in conditions such as attention deficit hyperactivity disorder and anxiety disorders.

Beta Waves (13-30 Hz): These bands are faster and lower in amplitude compared to alpha waves. They are typically observed during wakefulness and are associated with active cognitive processing, focused attention, and alertness. Beta waves are most prominent during periods of concentration, mental tasks, and sensory stimulation. Abnormal beta wave activity has been implicated in conditions such as Parkinson's disease, schizophrenia, and anxiety disorders.

Gamma Waves (30-100 Hz): These bands are the fastest brain waves and are involved in high-level cognitive functions such as perception, consciousness, and memory encoding. They are often observed during active cognitive tasks, sensory processing, and states of heightened attention. Gamma wave activity is hypothesized to be a reflection of synchronous neuronal firing and is involved in the binding and integration of information between different areas of the brain.

Thus, the filtered delta, theta, alpha, beta, and gamma features are fed into the feature extraction module for deriving the essential attributes.

3.3. Feature Extraction

Empirical Mode Decomposition (EMD) is a signal processing technique used for feature extraction from EEG signals, particularly in the context of PD classification. EMD decomposes a signal $l(t)$into a finite set of intrinsic mode functions (IMFs) $p_x(t)$along with residual signal $b(t)$based on the local characteristic scales of the signal. Each IMF represents a different oscillatory mode within the signal, capturing local oscillations or patterns. While considering the EEG signal processing for PD classification, EMD is applied to decompose the EEG signal into IMFs. These IMFs are then analyzed to extract features that are relevant for distinguishing illness among individuals.The EMD process is defined as:

$$l(t) = \sum_{i=1}^{Q} p_x(t) + b(t) \tag{1}$$

where, the total IMF derived is denoted as Q. Each IMF $p_x(t)$is characterized by its local oscillatory behavior and represents a different frequency component of the original signal. To extract features from the IMFs for PD classification, various statistical and spectral measures can be computed and are detailed below.

- Mean (IMF_1): Represents the average value of the IMF over a specific time interval and is calculated as:

$$IMF_1 = \frac{1}{E}\int_0^E p_x(t)dt \tag{2}$$

where, E signifies the

- Standard Deviation (IMF_2): Measures the dispersion of data points around the mean and is computed as:

$$IMF_2 = \sqrt{\frac{1}{E}\int_0^E [p_x(t) - IMF_1]^2 dt} \tag{3}$$

- Skewness (IMF_3): Describes the asymmetry of the IMF's probability distribution and is given by:

$$IMF_3 = \frac{1}{E}\int_0^E \left[\frac{p_x(t) - IMF_1}{IMF_2}\right]^3 dt \tag{4}$$

- Kurtosis (IMF_4): Measures the peakedness of the IMF's probability distribution and is defined as:

$$IMF_4 = \frac{1}{E}\int_0^E \left[\frac{p_x(t) - IMF_1}{IMF_2}\right]^4 dt - 3 \tag{5}$$

Using the extracted features, the Parkinson disease prediction is employed using the ImCfO_Attn_EffBNet technique.

3.4. Parkinson Disease Classification Using Improved Attn_EffBNet

The Parkinson Disease classification is employed using the novel Improved Attention based Efficient Bidirectional Network (Attn_EffBNet) model. In the Attn_EffBNet, the EfficientNet-B7, BiLSTM and Attention modules are integrated for enhancing the classification accuracy through capturing spatial and temporal features. Besides, the adjustable parameters of the classifier are fine-tuned using the proposed ImCfO algorithm.

3.4.1. Design of Attn_EffBNet

Well-known for its ability to extract spatial information from input data, EfficientNet is potent convolutional neural network (CNN) architecture. By taking into account the EEG data, EfficientNet is utilized to extract pertinent spatial features from the signal representations, identifying significant patterns and structures from the data. For capturing temporal relationships in sequential data, bidirectional long short-term memory (BiLSTM) networks are

an effective alternative. Considering the intrinsic temporal nature of EEG data, activity patterns throughout time might yield vital information for the categorization of various diseases. BiLSTM layers enable the model to concurrently learn from past and future contexts by efficiently capturing both forward and backward temporal relationships in the EEG data. Additionally, the attention processes allow the model to ignore noisy or irrelevant information and concentrate on pertinent portions of the input data. Attention processes can dynamically weight the value of various temporal and spatial variables retrieved by EfficientNet and BiLSTM while analyzing the EEG data for the categorization of PD. By focusing on prominent aspects in the EEG signals that are most instructive for precise classification, this attention mechanism helps the model improve its discriminative capacity. Therefore, BiLSTM and Attention modules are incorporated for improved classification accuracy in the proposed EfficientNet-B7 Parkinson disease classification model. Figure 2 shows the structure of the Attn_EffBNet Parkinson Disease classification model.

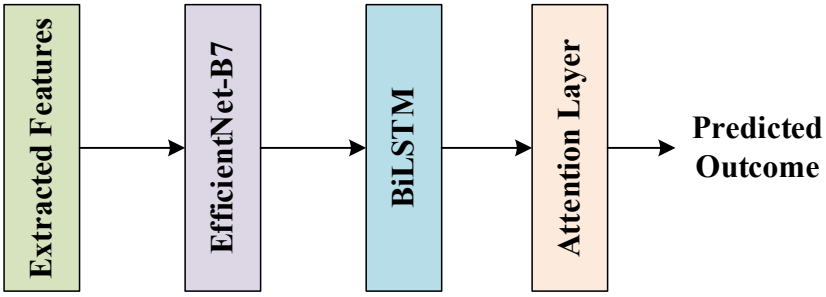


Figure 2. Structure of Attn_EffBNet

EfficientNet Module: The electrical activity of the brain throughout time is represented by EEG signals, which frequently include spatial patterns characteristic of neurological disorders. EfficientNet, with its deep architecture and effective convolutional layers, can extract spatial features from EEG signals. These features capture patterns and structures present in different regions of the brain, which may vary between individuals with and without disease. The structure of the EfficientNet utilized for capturing the spatial patterns is portrayed in Figure 3.

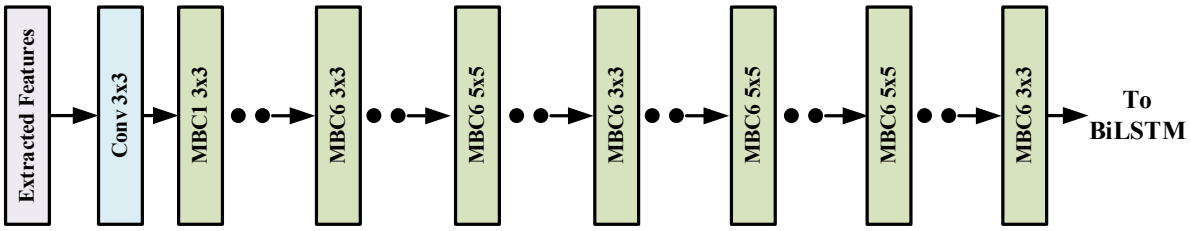


Figure 3. Structure of EfficientNet

EfficientNet utilizes three various parameters like resolution, width and height with the compound scaling factor for providing the enhanced efficiency with minimal parameters.

Depth: Depth of EfficientNet refers to the number of layers or the overall network's depth. Deeper networks often have more capacity to capture complex patterns and representations from input data. However, increasing depth also leads to higher computational costs and the risk of overfitting, especially with limited training data. In EfficientNet, the compound scaling method ensures that depth is balanced with other factors such as width and resolution to achieve optimal performance without excessive complexity. It is defined as:

$$F = \chi^{\omega} \tag{6}$$

where, depth is indicated as F, compound coefficient is indicated as ω and χ refers to the scaling factor.

Width: The quantity of channels or neurons in each network layer is referred to as width. Increasing the width of the network allows for more features to be learned at each layer, potentially improving the model's representational capacity. However, widening the network also increases computational requirements and memory usage. EfficientNet uses compound scaling to carefully balance width with other architectural parameters, ensuring that the network achieves a good trade-off between model capacity and efficiency. It is defined as:

$$G = \beta^{\omega} \tag{7}$$

where, width is indicated as G, and β refers to the scaling factor.

Resolution: Resolution refers to the spatial dimensions of the input data. In the case of image classification tasks, resolution corresponds to the height and width of the input images. Higher resolution images contain more detailed information, which can be beneficial for capturing fine-grained features. However, processing high-resolution images requires more computational resources and memory. EfficientNet addresses this by scaling down the input resolution while maintaining the network's efficiency. This scaling process ensures that the model can handle inputs of varying resolutions without sacrificing performance. It is defined as:

$$K = \gamma^{\omega} \tag{8}$$

where, resolution is indicated as K, and γ refers to the scaling factor.

These architectural aspects are critical in defining the model's capacity to effectively extract pertinent features from EEG data for PD classification using EfficientNet. EfficientNet can efficiently capture spatial patterns and temporal dynamics in EEG signals by carefully

balancing depth, breadth, and resolution using compound scaling. This results in accurate classification outcomes while minimizing processing overhead.

BiLSTM Module: BiLSTM is well-suited for capturing long-range temporal dependencies in sequential data by maintaining an internal state that evolves over time. It can learn from past and future contexts simultaneously, allowing it to capture complex temporal dynamics present in EEG signals. BiLSTM analyses input sequences in both forward and backward directions, in contrast to conventional LSTM networks that solely take historical data into account when generating predictions. The model can include data from both past and future time steps because to its bidirectional processing, giving it a more thorough knowledge of the temporal environment. In the context of PD classification, this bidirectional processing can help capture subtle patterns that may span across different parts of the EEG signal.

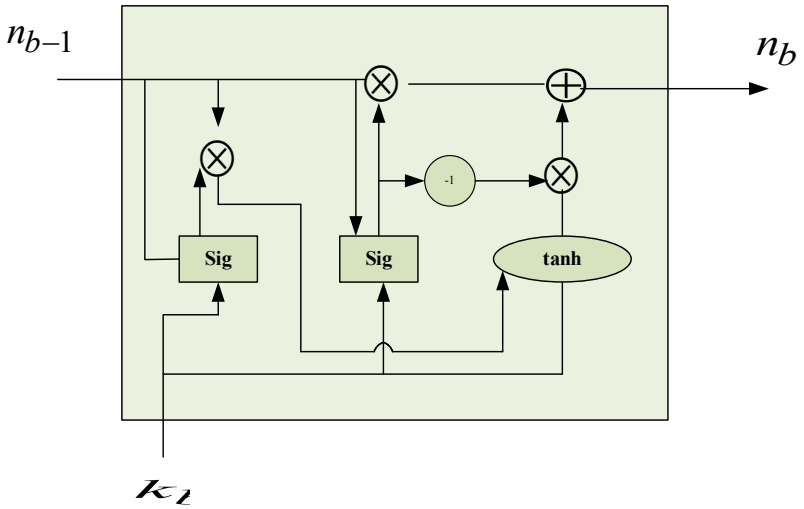


Figure 4. Structure of LSTM

Several Bidirectional LSTM layers make up the central component of the BiLSTM architecture. A BiLSTM layer consists of two LSTM sub-layers, one for forward processing of the input sequence and another for backward processing. Due to its bidirectional processing, the network may concurrently collect temporal dependencies from contexts in the past and the future. Within each LSTM sub-layer, LSTM cells process the sequential input data. An LSTM cell consists of several components like input, forget, cell state and output gate.

Input Gate: Controls the flow of information entering the cell. It is defined as:

$$r_b = \lambda\left(c_r \cdot \left[n_{b-1}, k_b\right] + l_r\right) \tag{9}$$

Forget Gate: Determines which information from the previous state should be discarded. It is defined as:

$$g_b = \lambda\left(c_g \cdot \left[n_{b-1}, k_b\right] + l_g\right) \tag{10}$$

Cell State: Represents the memory of the cell, which can be updated or maintained over time. It is defined as:

$$\widetilde{w}_b = \tanh\left(c_w \cdot \left[n_{b-1}, k_b\right] + l_w\right) \tag{11}$$

Output Gate: Controls the data that is sent to the output layer or subsequent time step. It is defined as:

$$q_b = \lambda\left(c_q \cdot \left[n_{b-1}, k_b\right] + l_q\right) \tag{12}$$

Hidden State and cell state: The LSTM cells maintain hidden states and cell states, which store information about the temporal context of the input sequence. These states evolve over time as the network processes the input data, capturing relevant features and patterns. It is defined as:

$$w_b = g_b \cdot w_{b-1} + r_b \cdot \widetilde{w}_b \tag{13}$$

$$n_b = q_b \cdot \tanh\left(w_b\right) \tag{14}$$

where, the input is denoted as k_b for the time step b, the hidden state outcome concerning the past and present time step is signified as n_{b-1} and n_b, and the sigmoid function is defined as λ. The weight is denoted as l, and the bias is signified as c. w_b denotes the cell state and candidate state is denoted as $\widetilde{w}_b$. The output gate is defined as q_b, the input gate is defined as r_b and the forget gate is described as g_b respectively.As a result, the LSTM cells mitigate the vanishing gradient issue that traditional RNNs frequently face while efficiently capturing and remembering long-range relationships in sequential data. The LSTM is taken into account by the BiLSTM in both forward and backward directions, and the results are concatenated to take into account both past and present characteristics. The outcome of the BiLSTM is fed into the attention module for assigning weights to the extracted features.

Attention Module: An attention mechanism can be integrated into the proposed architecture to selectively focus on informative features extracted from the EEG data. By assigning higher weights to relevant channels or time points in the EEG signal, the attention mechanism helps the model identify patterns and abnormalities indicative of PD while filtering out noise and irrelevant information. It is defined as:

$$A_l = soft\max\left(R_q N_l + M_l\right) \tag{15}$$

where, the attention layer outcome is denoted as A_l, weight and bias is denoted as R_q and M_l respectively and the outcome of the GhostNet is denoted as N_l.Here, the adjustable parameters like the weights and bias are modified using ImCfO algorithm, which is detailed below.

3.4.2. Improved Crayfish Optimization Algorithm (ImCfO)

The ImCfO algorithm is designed for fine-tuning the parameters of the classifier for enhancing the classification accuracy of Parkinson disease.In order to improve the global best solution and convergence rate, the self-adaption criteria is integrated with the traditional crayfish method to create the suggested ImCfO algorithm. The ImCfO algorithm's solution, which is thus achieved, is used to alter the classifier's configurable parameters.

Initialization: Within the bounds of the problem, a population of crayfish (possible solutions) is created at random manner. It is described as:

$$H=\left[H_1,H_2,....H_D\right]=\begin{bmatrix} H_{1,1} & \cdots & H_{1,m} & \cdots & H_{1,E} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ H_{l,1} & \cdots & H_{l,m} & \cdots & H_{l,E} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ H_{D,1} & \cdots & H_{D,m} & \cdots & H_{D,E} \end{bmatrix} \tag{16}$$

where, the population of the algorithm is indicated as D, the initial solution of the algorithm is indicated as H, and E refers the dimension of solution. $H_{l,m}$ denotes the solution of the search agent l with the dimension m and the formulation for evaluating the solution acquired by the search agent is described as:

$$H_{l,m}=L_m+\left(M_m-L_m\right)\times A \tag{17}$$

where, the limits of the search agent is indicated as lower and upper bounds and is represented as L_m and M_m respectively and the arbitrary value is referred as A.

The behavior of the search agent varies based on the temperature of the surrounding and hence the definition of the temperature is expressed as:

$$N=A\times 15+20 \tag{18}$$

where, the temperature is signified as N, and the feeding quantity of the search agent is defined as:

$$q=f_1\times\left(\frac{1}{\sqrt{2\pi}\times SD}\times\exp\left(-\frac{\left(N-Avg\right)^2}{2\cdot SD^2}\right)\right) \tag{19}$$

where, the feeding intake of the search agent is controlled by the parameter SD and f_1, the most appropriate temperature for the search agent is indicated as Avg, and the feeding quantity is signified as q.

Randomization: During hot weather, crayfish seek refuge in caves or cooler areas, which is considered as the exploration stage of the algorithm. In this stage, the algorithm broadly searches the problem space for potential solutions, similar to how crayfish explore their

environment for suitable shelters. The appropriate temperature for the favorable randomization is greater than 30^0 Celsius. The formulation for the search area identified in this phase is expressed as:

$$H_C = \frac{H_P + H_Q}{2} \tag{20}$$

where, the best solution identified by the search agent until the previous iteration is denoted as H_P and the solution identified in the present iteration is denoted as H_Q. The solution arrived by the crawfish in the randomization phase is formulated as:

$$H_{l,m}^{\tau+1} = H_{l,m}^{\tau} + f_2 \times A \times \left(H_C - H_{l,m}^{\tau}\right) \tag{21}$$

Here, the parameter f_2 diminishes throughout the iterations and is defined as:

$$f_2 = 2 - \left(\tau \big/ \tau_{\max}\right) \tag{22}$$

where, the maximal iteration is indicated as $\tau_{\max}$ and the current iteration is defined as τ. Here for enhancing the randomization criteria further self-adaption concept is incorporated within the solution updation phase of the crayfish. The solution updation based on the self-adaption is formulated as:

$$Self_{ada} = H_{l,m}^{\tau+1} + Z\left(H_Q - H_{l,m}^{\tau}\right) \tag{23}$$

where, the control factor is denoted as Z, the self-adaption concept is denoted as $Self_{ada}$, and the solution updation by incorporating the self-adaption is expressed as:

$$\left(H_{l,m}^{\tau+1}\right)_{\text{ImCfO}} = Self_{ada} * \left(H_{l,m}^{\tau+1}\right)_{Crayfish} \tag{24}$$

$$\left(H_{l,m}^{\tau+1}\right)_{\text{ImCfO}} = Self_{ada} * H_{l,m}^{\tau} + f_2 \times A \times \left(H_C - H_{l,m}^{\tau}\right) \tag{25}$$

Thus, using the novel solution updation, the randomization criteria of the algorithm is enhanced for obtaining global best solution by solving the issue of local optimum trapping.

Local Search 1: Crayfish often compete for hiding spots, especially during limited availability, which is considered in the competition stage of the algorithm. Here, individuals (potential solutions) are adjusted based on the positions of other individuals, mimicking the jostling for better hiding spots. This stage helps the algorithm explore new areas within promising regions of the search space. The temperature considered in this stage is $N > 30$ and the solution updated in this phase is described as:

$$H_{l,m}^{\tau+1} = H_{l,m}^{\tau} - H_{r,m}^{\tau} + H_C \tag{26}$$

Here, the crayfish chosen arbitrarily is denoted as r, which is defined as:

$$r = round\left(A \times \left(D - 1\right)\right) + 1 \tag{27}$$

Local Search 2: Crayfish actively forage for food, adapting their strategy based on the food size, which is considered into the foraging stage. When encountering large food that representing a significant difference from the optimal solution, the algorithm employs a mathematical operation to adjust the search direction, similar to how a crayfish might tear up large food before consuming it. This enhances exploration in areas with less optimal solutions. Conversely, for small food that representing a solution closer to the optimal one, the algorithm uses a sine-cosine combination to simulate the alternating movement of crayfish claws during feeding, effectively refining the search around promising areas. The temperature considered in this stage is $N \le 30$ and the position of target identified is described as:

$$H_{tar} = H_P \tag{28}$$

The size of the target identified in this phase is expressed as:

$$B = f_3 \times A \times \left(Fit_l / Fit_{tar}\right) \tag{29}$$

where, thetarget parameter is signified as f_3, the target fitness is denoted as Fit_{tar} and the fitness evaluated by the l^{th} crayfish is signified as Fit_l. When $B > \left(f_3 + 1\right)/2$, the identified food reserve is larger and is defined as:

$$H_{tar} = \exp\left(-\frac{1}{B}\right) \times H_{tar} \tag{30}$$

Based on the identified target, the solution updation in this phase is defined as:

$$H_{l,m}^{\tau+1} = H_{l,m}^{\tau} + H_{tar} \times q \times \cos\left(2\pi A\right) - \sin\left(2\pi A\right) \tag{31}$$

If $B \le \left(f_3 + 1\right)/2$, then the solution accomplished in the second local search criteria is expressed as:

$$H_{l,m}^{\tau+1} = \left(H_{l,m}^{\tau} - H_{tar}\right) \times q + q \times A \times H_{l,m}^{\tau} \tag{32}$$

The identification of the target tar is the solution utilized by the ImCfO algorithm for fine-tuning the optimal parameters of the classifier for enhancing the accuracy of Parkinson disease prediction. The pseudo-code for the proposed ImCfO algorithm is presented in Algorithm 1.

Algorithm 1: Pseudo-code for ImCfO algorithm

Pseudo-code for ImCfO algorithm
Initialize D, E and H
Evaluate the solution using $H_{l,m} = L_m + \left(M_m - L_m\right) \times A$

{

while

{

$\tau < \tau_{\max}$

Evaluate the randomization using $\left(H_{l,m}^{\tau+1}\right)_{\text{ImCfO}} = Self_{ada} * H_{l,m}^{\tau} + f_2 \times A \times \left(H_C - H_{l,m}^{\tau}\right)$

Evaluate the local search 1 using $H_{l,m}^{\tau+1} = H_{l,m}^{\tau} - H_{r,m}^{\tau} + H_C$

Evaluate the local search 2

{

if $B > \left(f_3 + 1\right)/2$

{

Evaluate the solution using $H_{l,m}^{\tau+1} = H_{l,m}^{\tau} + H_{tar} \times q \times \cos\left(2\pi A\right) - \sin\left(2\pi A\right)$

else

{

Evaluate the solution using $H_{l,m}^{\tau+1} = \left(H_{l,m}^{\tau} - H_{tar}\right) \times q + q \times A \times H_{l,m}^{\tau}$

}

}

end

$\tau = \tau + +$

}

stop

Consequently, the ImCfO algorithm's answer is used to modify the classifier's ideal parameters. The local optimum trapping problem is resolved and the global best solution is given by the algorithm with expanded exploration. Thus, the ImCfO algorithm improves the prediction accuracy of Parkinson disease.

4. Result and Discussion

The proposed ImCfO_ Attn_EffBNet is implemented in PYTHON programming language and is evaluated using Parkinson EEG dataset [26]. To demonstrate the superiority of the suggested approach, the ImCfO_ Attn_EffBNet method is compared with other existing methods such as CNN, QMO_DL, and DenseNet.

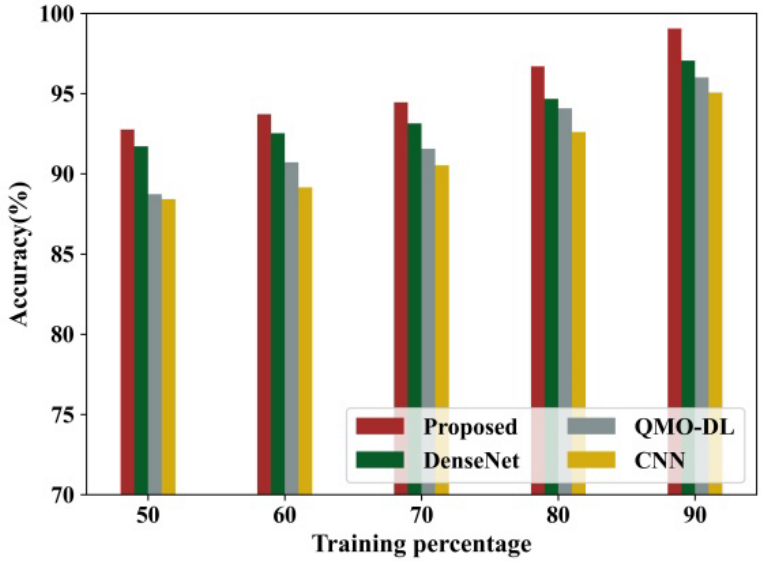
Dataset Description: The Parkinson Disease Dataset is taken from [26], which comprises of EEG signal taken from patients with Parkinson and normal individual. Various attributes of the patients are recorded for making the disease prediction more effective.

4.1. Comparative Analysis

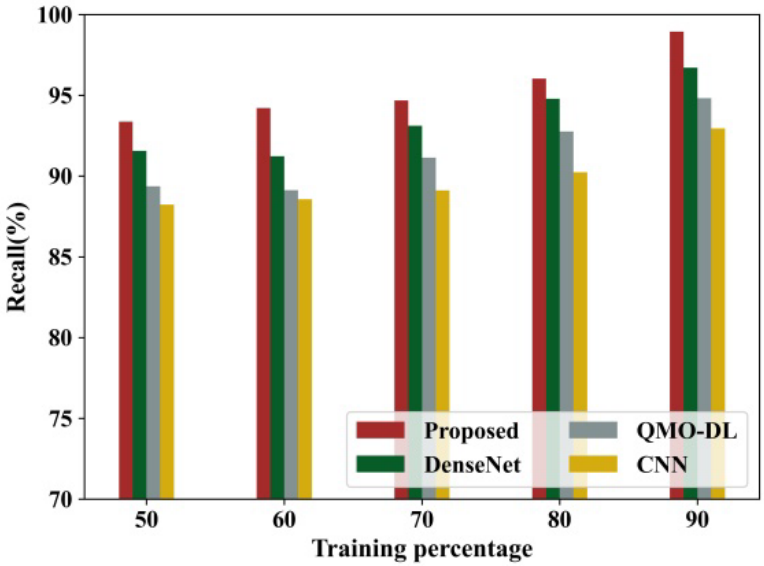
The proposed ImCfO_ Attn_EffBNet is assessed using accuracy, recall, specificity, F-Score, FPR, and precision. In this, the analysis is devised based on training percentage and K-Fold data.

4.1.1. Training Data

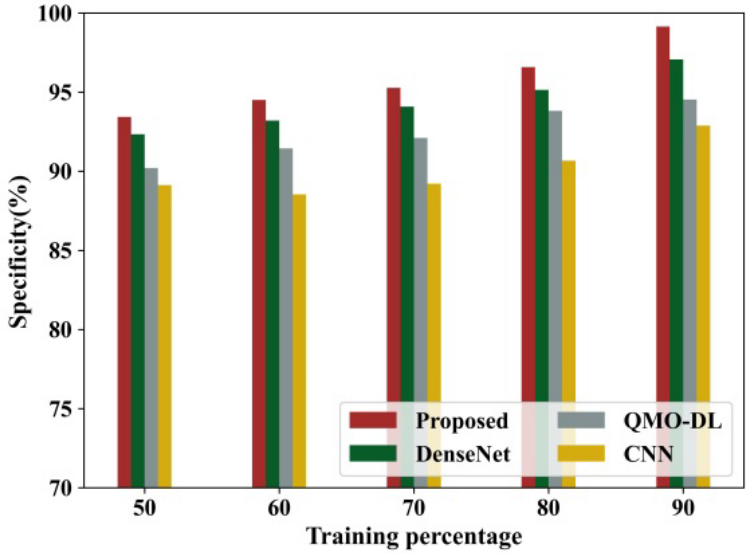
Various training data amount based analysis of ImCfO_ Attn_EffBNetrefers to the investigation of how the size of the training dataset affects the performance of predictive models. The analysis is portrayed in Figure 5 and its detailed outcome is presented in Table 1. For example, by training models using only 50% of the available data and evaluate their performance on a separate 50% test dataset. They then repeat the process with larger training datasets like 60%, 70%, 80%, and 90% of the data to observe how the model performance changes with increasing amounts of training data.The examination showed a better result with more training data, indicating that the ImCfO algorithm of Attn_EffBNet's loss function optimization improves the prediction accuracy by providing more generalization capacity.



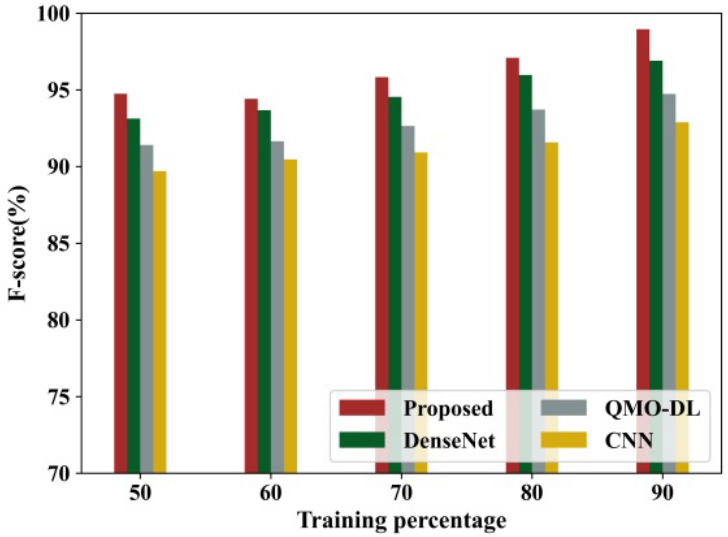
(a)


100
95
90
85
80
75
70
Recall(%)
50
60
70
80
90
Training percentage
Proposed
DenseNet
QMO-DL
CNN

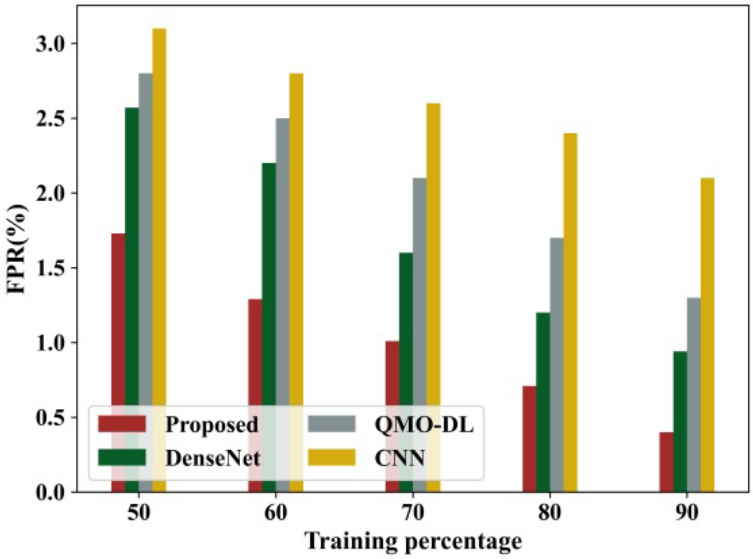
(b)


100
95
90
85
80
75
70
Specificity(%)
50
60
70
80
90
Training percentage
Proposed
DenseNet
QMO-DL
CNN

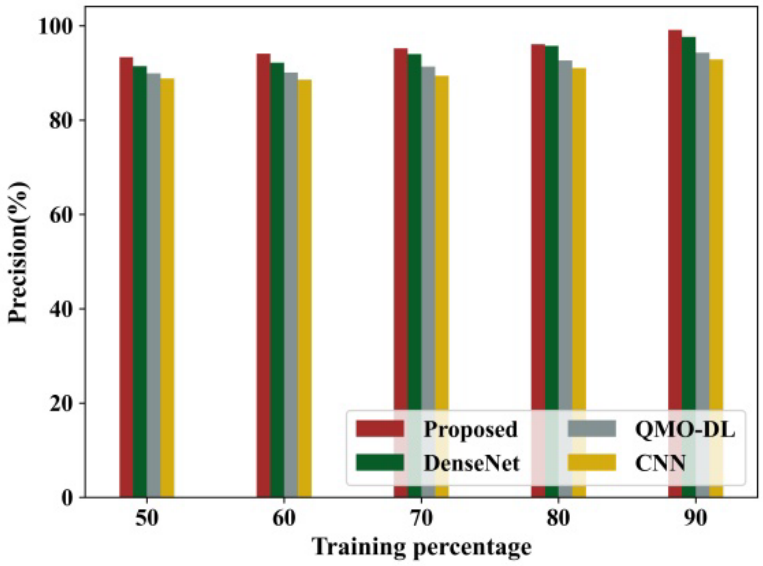
(c)


100
95
90
85
80
75
70
F-score(%)
Proposed
DenseNet
QMO-DL
CNN
50
60
70
80
90
Training percentage

(d)


3.0
2.5
2.0
1.5
1.0
0.5
0.0
FPR(%)
Proposed
DenseNet
QMO-DL
CNN
50
60
70
80
90
Training percentage

(e)



(f)

Figure 5. Training Data based analysis (a) Accuracy, (b) Recall, (c) Specificity, (d) F-Score, (e) FPR and (f) Precision

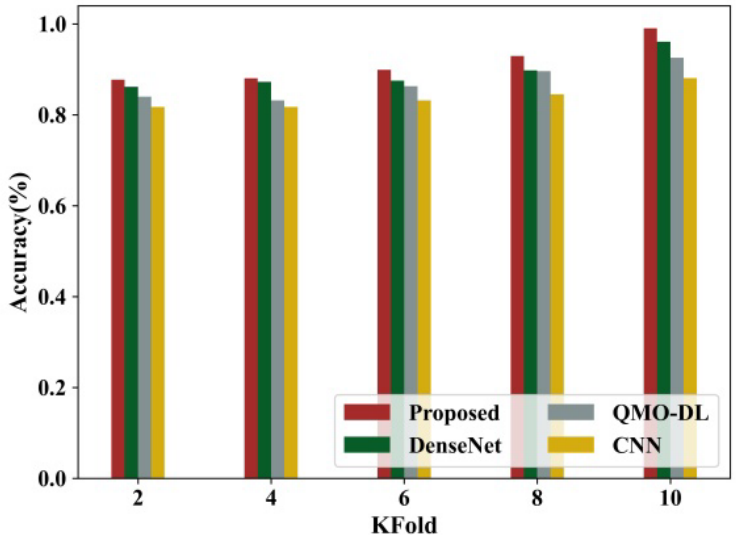
Table 1. Training Data based analysis

Training Percentage/ Methods	Proposed	DenseNet	QMO-DL	CNN
Accuracy				
50	92.753	91.712	88.727	88.420
60	93.712	92.534	90.714	89.150
70	94.455	93.145	91.562	90.523
80	96.701	94.667	94.090	92.607
90	99.058	97.061	96.006	95.068
Recall				
50	93.370	91.560	89.370	88.240
60	94.210	91.230	89.120	88.560
70	94.680	93.110	91.145	89.110
80	96.030	94.780	92.760	90.232
90	98.946	96.714	94.823	92.948
Specificity				
50	93.430	92.340	90.210	89.120
60	94.512	93.210	91.450	88.546
70	95.270	94.090	92.110	89.210

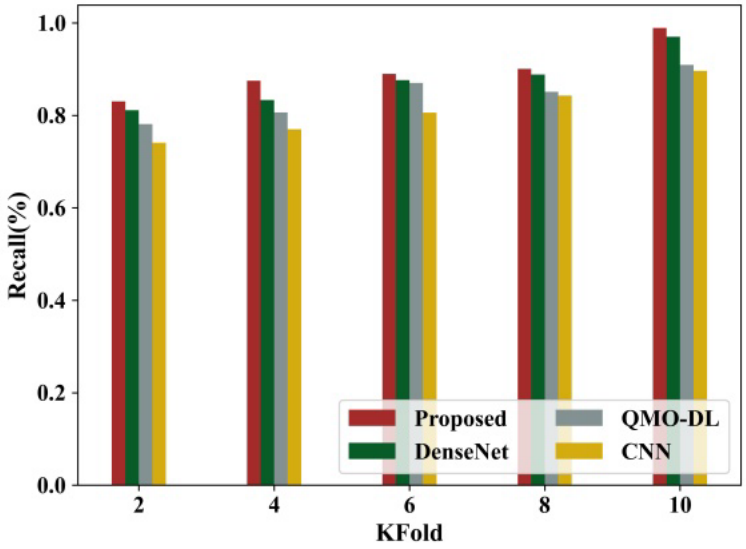
80	96.588	95.145	93.818	90.673
90	99.152	97.065	94.532	92.890
F-score				
50	94.743	93.134	91.402	89.701
60	94.412	93.671	91.645	90.460
70	95.827	94.539	92.650	90.920
80	97.088	95.970	93.718	91.573
90	98.960	96.900	94.730	92.890
Precision				
50	93.370	91.430	89.924	88.812
60	94.110	92.160	90.110	88.560
70	95.230	94.020	91.370	89.410
80	96.090	95.740	92.650	91.030
90	99.145	97.650	94.320	92.890
FPR				
50	1.730	2.570	2.800	3.100
60	1.290	2.200	2.500	2.800
70	1.010	1.600	2.100	2.600
80	0.710	1.200	1.700	2.400
90	0.400	0.940	1.300	2.100

4.1.2. K-Fold Data

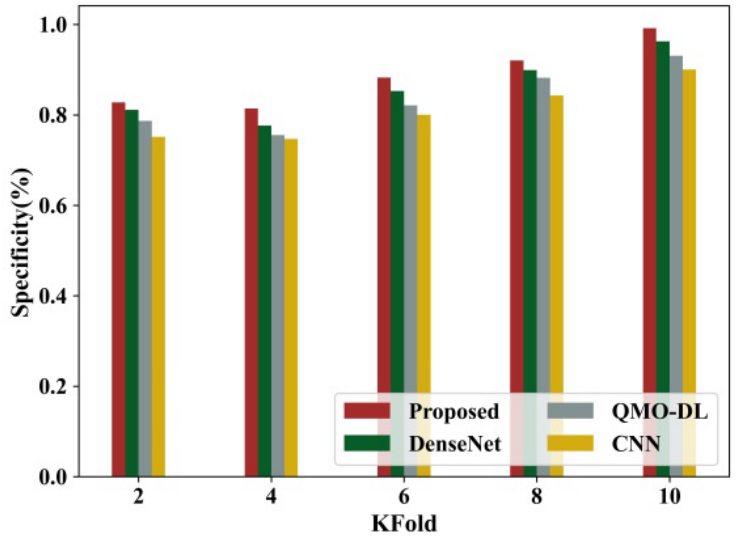
Analysis of PD prediction using different K-fold values, such as 2, 4, 6, 8, and 10, focuses on how the number of folds in K-fold cross-validation influences the predictive models' performance assessment. A popular resampling method in machine learning for evaluating the effectiveness of prediction models is K-fold cross-validation. Figure 6 depicts the K-Fold based examination, and Table 2 provides the entire findings. The dataset is split into K subsets, or folds, of about similar size for K-fold cross-validation. K-1 folds are used for training each time the model is trained, and the leftover fold is used for testing. Each fold is used as a testing set once, and this procedure is repeated K times. Smaller training and validation sets may arise from using a greater value of K, such as K=10, which can provide a reliable assessment of the model's generalization performance.


1.0
0.8
0.6
0.4
0.2
0.0
Accuracy(%)
2
4
6
8
10
KFold
Proposed
DenseNet
QMO-DL
CNN

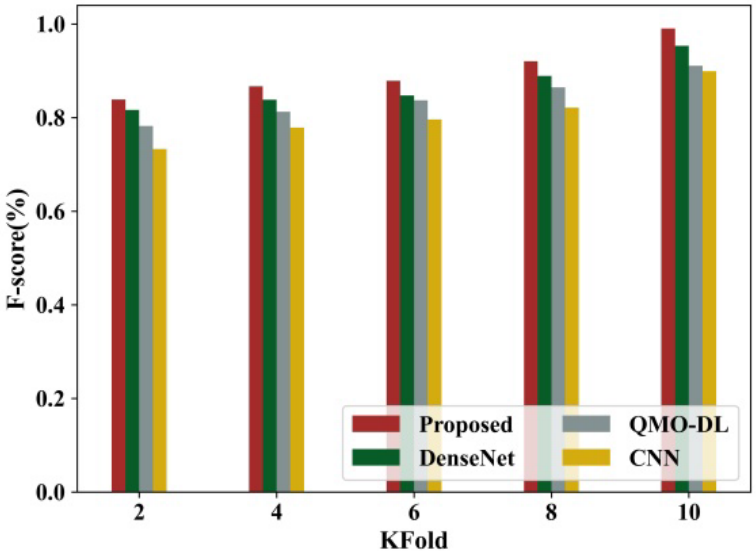
(a)


1.0
0.8
0.6
0.4
0.2
0.0
Recall(%)
2
4
6
8
10
KFold
Proposed
DenseNet
QMO-DL
CNN

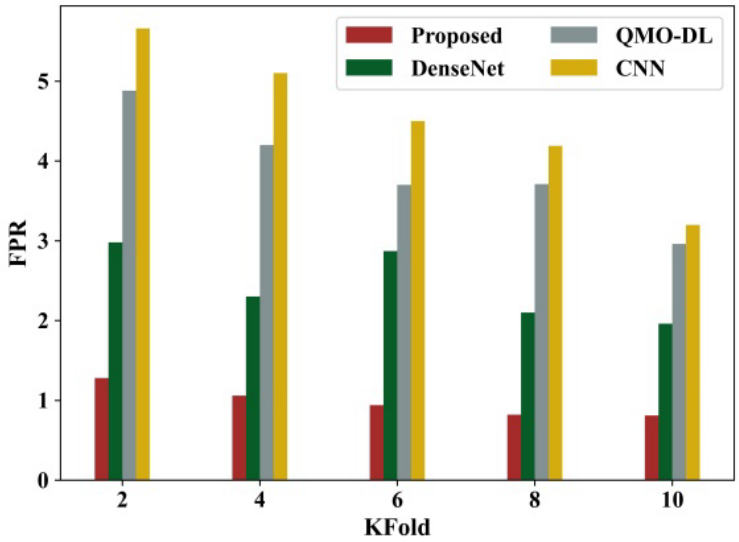
(b)


1.0
0.8
0.6
0.4
0.2
0.0
Specificity(%)
2
4
6
8
10
KFold
Proposed
DenseNet
QMO-DL
CNN

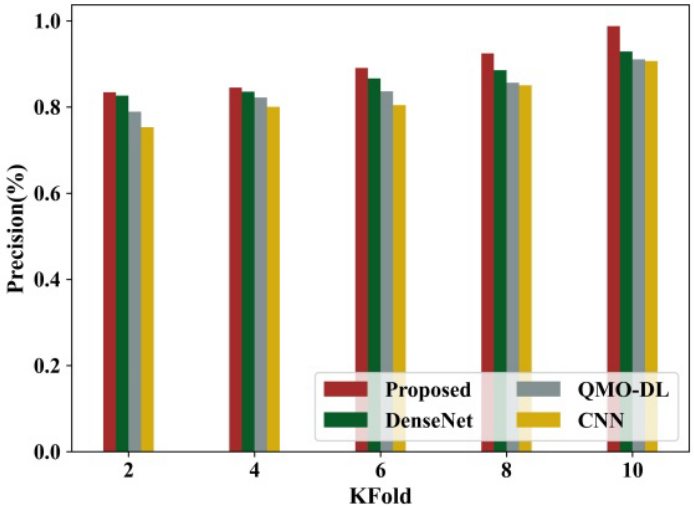
(c)


1.0
0.8
0.6
0.4
0.2
0.0
F-score(%)
2
4
6
8
10
KFold
Proposed
DenseNet
QMO-DL
CNN

(d)


5
4
3
2
1
0
FPR
2
4
6
8
10
KFold
Proposed
DenseNet
QMO-DL
CNN

(e)



(f)

Figure 6. K-Fold Data based analysis (a) Accuracy, (b) Recall, (c) Specificity, (d) F-Score, (e) FPR and (f) Precision

Table 2. K-Fold Data based analysis

K-Fold/ Methods	Proposed	DenseNet	QMO-DL	CNN
Accuracy				
2	0.877	0.862	0.840	0.818
4	0.880	0.873	0.832	0.817
6	0.899	0.875	0.863	0.832
8	0.930	0.898	0.896	0.845
10	0.991	0.961	0.926	0.881
Recall				
2	0.830	0.811	0.781	0.741
4	0.875	0.833	0.806	0.770
6	0.890	0.876	0.870	0.806
8	0.901	0.888	0.851	0.843
10	0.989	0.970	0.909	0.896
F-Score				
2	0.839	0.817	0.782	0.733
4	0.867	0.838	0.813	0.779
6	0.879	0.847	0.837	0.796
8	0.920	0.889	0.865	0.822
10	0.991	0.954	0.911	0.899

Precision				
2	0.835	0.827	0.789	0.753
4	0.845	0.836	0.822	0.801
6	0.891	0.866	0.837	0.804
8	0.925	0.885	0.856	0.850
10	0.988	0.929	0.911	0.907
Specificity				
2	0.828	0.811	0.787	0.751
4	0.814	0.776	0.756	0.747
6	0.882	0.853	0.821	0.800
8	0.920	0.899	0.882	0.843
10	0.992	0.963	0.931	0.900
FPR				
2	1.280	2.980	4.880	5.660
4	1.060	2.300	4.200	5.100
6	0.940	2.870	3.700	4.500
8	0.820	2.100	3.710	4.190
10	0.810	1.960	2.960	3.198

4.2. ROC Analysis

A prediction model's capacity to distinguish between those with PD and those without is evaluated using ROC analysis. Plotting the real positive rate versus the false positive rate across various threshold values used to categorize people as positive (having Parkinson's disease) or negative (not having Parkinson's disease) is estimated by ROC. The ROC analysis is illustrated in Figure 7.

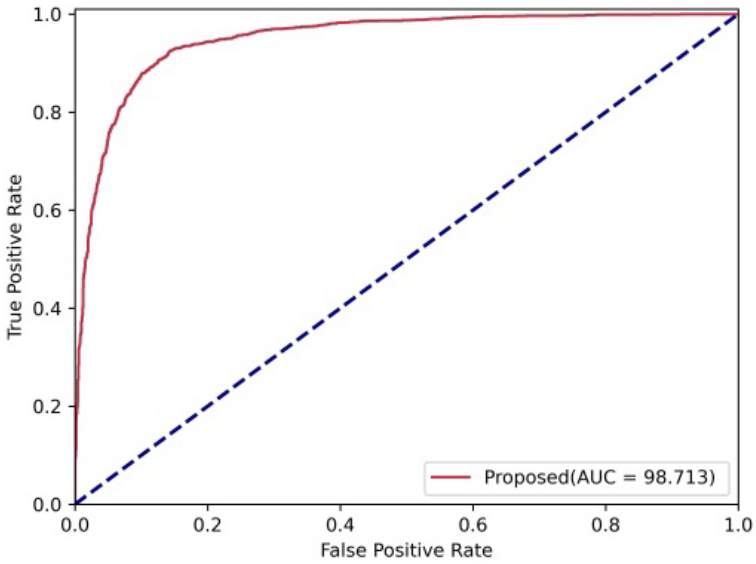
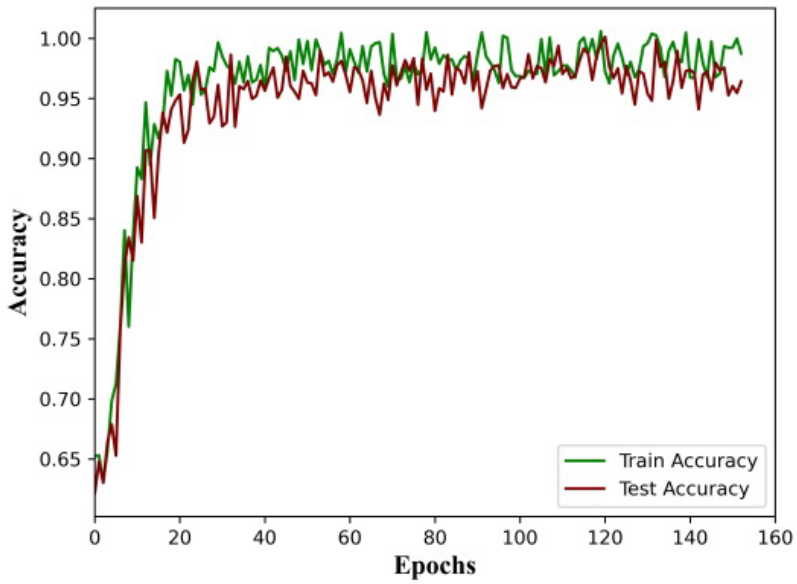


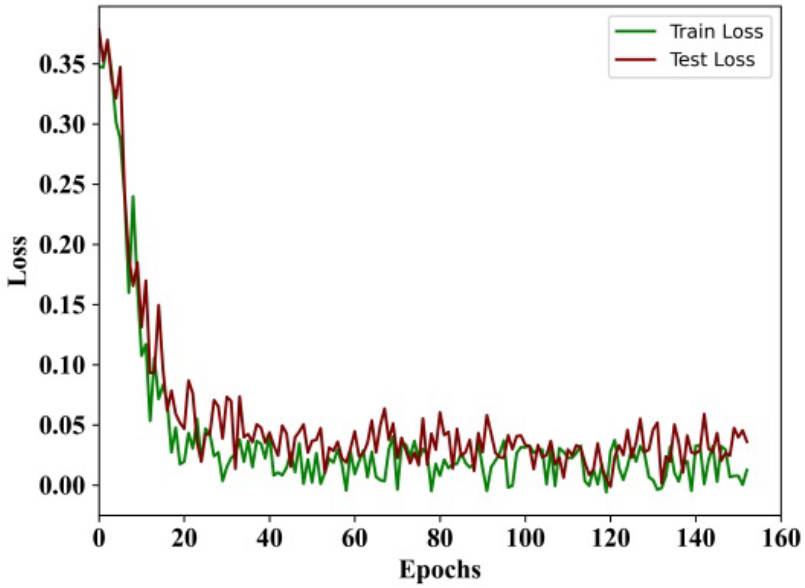
Figure 7. ROC Analysis

4.3. Accuracy-Loss Analysis

When predicting PD, accuracy-loss analysis refers to analyzing the correlation between the predictive models' accuracy and the loss function applied during model training. In Figure 8, the accuracy-loss analysis is shown.



(a)



(b)

Figure 8. Accuracy-Loss Analysis

4.4. Convergence Analysis

Assessing how quickly the optimization algorithm progresses towards the optimal solution. A faster rate of convergence indicates efficient optimization, while slower convergence may suggest potential issues such as suboptimal parameter settings or convergence to local minima. Proposed ImCfO algorithm converges faster compared to the traditional Crayfish optimization (CFO) algorithm, which is portrayed in Figure 9.

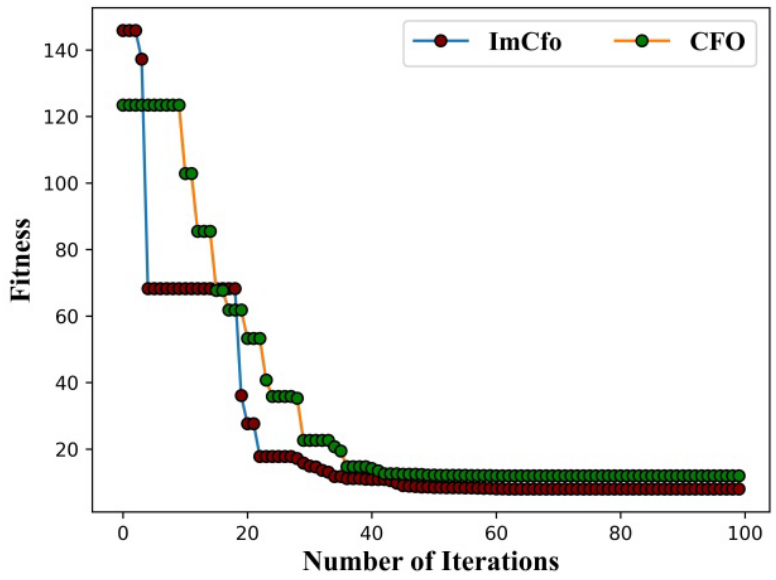


Figure 9. Convergence Analysis

4.5 Comparative Discussion

The best performance evaluated by the ImCfO_Attn_EffBNet based on various assessment are accuracy (95.068%), Recall (92.948%), Specificity (92.89%), F-Score (92.89%), Precision (92.89%), and FPR (2.1%). Here, the combination of Efficientnet-B7 and Bidirectional LSTM networks in the Attn_EffBNet model allows for highly efficient feature extraction from the input data. Sequential data can have temporal relationships captured by Bidirectional LSTM, and Efficientnet-B7 is renowned for its adeptness in extracting hierarchical features from input. This makes it possible for the model to extract temporal and geographical information from the input data, which is essential for precise PD prediction. The use of the Improved Crayfish Optimization Algorithm to optimize the loss function of the Attn_EffBNet model is another significant advantage. This optimization technique helps the model converge to better solutions faster, resulting in improved performance and efficiency during training. By effectively optimizing the loss function, the model can better learn from the training data and generalize to unseen data, leading to enhanced prediction accuracy for PD.

Thus, in the comparative discussion, various metrics depicts the superior performance of the Attn_EffBNet model with the Improved Crayfish Optimization Algorithm in predicting PD. High values for accuracy, recall, specificity, F-Score, and precision indicate a strong predictive model with good performance in identifying both positive and negative instances of PD. Conversely, a low False Positive Rate indicates that the model is effectively minimizing the misclassification of negative instances as positive. Thus, the overall effectiveness of the proposed predictive model based on the assessment measures depicts its potential utility in clinical practice for PD diagnosis and prediction.

5. Conclusion

In this research, we introduced a comprehensive framework for PD prediction that leverages advanced machine learning techniques and optimization algorithms. Through experimentation, we demonstrated the effectiveness of our proposed method in accurately predicting PD from publicly available datasets. By integrating Empirical Mode Decomposition for feature extraction and ImCfO_Attn_EffBNet for classification, we achieved notable improvements in prediction accuracy compared to existing approaches. The incorporation of the ImCfO algorithm further enhanced convergence rates and global solution quality. The performance evaluated by the ImCfO_Attn_EffBNet based on various assessments acquired the outcome of accuracy (95.068%), Recall (92.948%), Specificity (92.89%), F-Score (92.89%), Precision (92.89%), and FPR (2.1%) respectively. The findings highlight the potential of combining state-of-the-art machine learning models with innovative optimization techniques for

advancing disease prediction and healthcare applications. Further research could explore the application of our approach in clinical settings and investigate its potential for early detection and personalized treatment of PD.

References

[1]Govindu, A., &Palwe, S. (2023). Early detection of Parkinson's disease using machine learning. *Procedia Computer Science*, *218*, 249-261.

[2]Elshewey, A. M., Shams, M. Y., El-Rashidy, N., Elhady, A. M., Shohieb, S. M., &Tarek, Z. (2023). Bayesian optimization with support vector machine model for parkinson disease classification. *Sensors*, *23*(4), 2085.

[3] Nancy Noella, R. S., &Priyadarshini, J. (2023).Machine learning algorithms for the diagnosis of Alzheimer and Parkinson disease. *Journal of Medical Engineering & Technology*, *47*(1), 35-43.

[4]Shcherbak, A., Kovalenko, E., &Somov, A. (2023). Detection and classification of early stages of Parkinson's disease through wearable sensors and machine learning. *IEEE Transactions on Instrumentation and Measurement*.

[5]Yousif, N. R., Balaha, H. M., Haikal, A. Y., & El-Gendy, E. M. (2023). A generic optimization and learning framework for Parkinson disease via speech and handwritten records. *Journal of Ambient Intelligence and Humanized Computing*, *14*(8), 10673-10693.

[6]Nijhawan, R., Kumar, M., Arya, S., Mendirtta, N., Kumar, S., Towfek, S. K., ...&Abdelhamid, A. A. (2023). A Novel Artificial-Intelligence-Based Approach for Classification of Parkinson's Disease Using Complex and Large Vocal Features. *Biomimetics*, *8*(4), 351.

[7]Suuronen, I., Airola, A., Pahikkala, T., Murtojärvi, M., Kaasinen, V., &Railo, H. (2023). Budget-based classification of Parkinson's disease from resting state EEG. *IEEE Journal of Biomedical and Health Informatics*.

[8]Gulay, B. K., Demirel, N., Vahaplar, A., &Guducu, C. (2023). A novel feature extraction method using chemosensory EEG for Parkinson's disease classification. *Biomedical Signal Processing and Control*, *79*, 104147.

[9]Avvaru, S., &Parhi, K. K. (2023). Effective Brain Connectivity Extraction by Frequency-Domain Convergent Cross-Mapping (FDCCM) and its Application in Parkinson's Disease Classification. *IEEE Transactions on Biomedical Engineering*.

[10]Sugden, R. J., &Diamandis, P. (2023). Generalizable electroencephalographic classification of Parkinson's disease using deep learning. *Informatics in Medicine Unlocked*, *42*, 101352.

[11]Lilhore, U. K., Dalal, S., Faujdar, N., Margala, M., Chakrabarti, P., Chakrabarti, T., ...&Velmurugan, H. (2023). Hybrid CNN-LSTM model with efficient hyperparameter tuning for prediction of Parkinson's disease. *Scientific Reports*, *13*(1), 14605.

[12]Khaskhoussy, R., &Ayed, Y. B. (2023). Improving Parkinson's disease recognition through voice analysis using deep learning. *Pattern Recognition Letters*, *168*, 64-70.

[13]Rahman, S., Hasan, M., Sarkar, A. K., & Khan, F. (2023). Classification of Parkinson's disease using speech signal with machine learning and deep learning approaches. *European Journal of Electrical Engineering and Computer Science*, *7*(2), 20-27.

[14] Li, K., Ao, B., Wu, X., Wen, Q., UlHaq, E., & Yin, J. (2023). Parkinson's disease detection and classification using EEG based on deep CNN-LSTM model. *Biotechnology and Genetic Engineering Reviews*, 1-20.

[15]Shehadeh, H. A., Jebril, I. H., Jaradat, G. M., Ibrahim, D., Sihwail, R., Al Hamad, H., ...& Alia, M. A. (2023). Intelligent Diagnostic Prediction and Classification System for Parkinson's Disease by Incorporating Sperm Swarm Optimization (SSO) and Density-Based Feature Selection Methods. *International Journal of Advances in Soft Computing & Its Applications*, *15*(3).

[16]Mahmood, A., Mehroz Khan, M., Imran, M., Alhajlah, O., Dhahri, H., &Karamat, T. (2023). End-to-end deep learning method for detection of invasive Parkinson's disease. *Diagnostics*, *13*(6), 1088.

[17] Bhatt, K., Jayanthi, N., & Kumar, M. (2023). High-resolution superlet transform based techniques for Parkinson's disease detection using speech signal. *Applied Acoustics*, *214*, 109657.

[18]Faragó, P., Ștefănigă, S. A., Cordoș, C. G., Mihăilă, L. I., Hintea, S., Peștean, A. S., ... &Ileșan, R. R. (2023). CNN-Based Identification of Parkinson's Disease from Continuous Speech in Noisy Environments. *Bioengineering*, *10*(5), 531.

[19]Parajuli, M., Amara, A. W., &Shaban, M. (2023). Deep-learning detection of mild cognitive impairment from sleep electroencephalography for patients with Parkinson's disease. *PLoS One*, *18*(8), e0286506.

[20]Keles, A., Keles, A., Keles, M. B., &Okatan, A. (2023).PARNet: Deep neural network for the diagnosis of parkinson's disease. *Multimedia Tools and Applications*, 1-13.

[21]Mahendran, M., &Visalakshi, R. (2024). An ensemble of ResNet model for classification of Parkinson disease. *International Journal of Nutrition, Pharmacology, Neurological Diseases*, *14*(1), 9-14.

[22] Mansour, R. F. (2024). Quantum mayfly optimization based feature subset selection with hybrid CNN for biomedical Parkinson's disease diagnosis. *Neural Computing and Applications*, 1-14.

[23]Latifoğlu, F., Penekli, S., Orhanbulucu, F., &Chowdhury, M. E. (2024). A novel approach for Parkinson's disease detection using Vold-Kalman order filtering and machine learning algorithms. *Neural Computing and Applications*, 1-15.

[24] Mahesh, T. R., Bhardwaj, R., Khan, S. B., Alkhaldi, N. A., Victor, N., &Verma, A. (2024). An artificial intelligence-based decision support system for early and accurate diagnosis of Parkinson's Disease. *Decision Analytics Journal*, *10*, 100381.

[25]Kanagaraj, S., Hema, M. S., &Guptha, M. N. (2024). An improved approach for early diagnosis of Parkinson's disease using advanced DL models and image alignment. *Automatika*, *65*(3), 911-924.

[26] Parkinson disease dataset, https://www.kaggle.com/datasets/s3programmer/parkison-diseases-eeg-dataset