Turkish Online Journal of Qualitative Inquiry (TOJQI)

Volume 12, Issue 5, June 2021: 1402 - 1421

Research Article

Parkinson's Disease Classification Using Fuzzy-Based Optimization Approach And Deep Learning Classifier.

B. Sabeena[1], S. Sivakumari[2]

ABSTRACT

Limited care is provided to PD (Parkinson's disease) affected individuals due to inadequate, irregular monitoring of symptoms, occasional care taken, light involvement of clinicians that leads to reduced effective decision and sub-optimal patient health-based results. In the starting period of PD, individuals commonly have vocal impairments. Hence, vocal problem based diagnosis method was the foremost research for PD. The irrelevant and/or redundant features are eliminated in feature selection method. These chosen features provide the best result using the objective function. For most of the cases, it is a NP-hard (Nondeterministic Polynomial-time hard) problem. From last 5 years, the database size has been increased and hence there is need for feature selection before performing any classification method. To solve this problem, Fuzzy Monarch Butterfly Optimization Algorithm (FMBOA) feature selection algorithm is introduced in this work. This algorithm selects most important features from the dataset and increases the PD detection rate. Firstly, KPCA (Kernel based Principal Component Analysis) dimensionality method is introduced for reducing dimension in the dataset. Secondly FBOA based feature selection; weight value is the essential factor that is used for searching optimal features in the PD classification. In the proposed FMBOA algorithm, weight value is computed via the Gaussian fuzzy membership function. A new event is performed in the proposed Fuzzy Monarch Butterfly Optimization Algorithm where the weight value of Butterfly Optimization Algorithm is modified while performing the optimization process to enhance the results. The classification algorithms are used for varied feature set that are obtained from ABOA and each set have different combinations. The FCBi-LSTM (Fuzzy Convolution Bi-Directional Long Short-Term Memory) is developed for PD classification. The introduced framework was evaluated using UCI repository of machine learning and LOPO CV is used for performance validation. The measures that are considered for performance evaluation are MCC, f-measure and accuracy.

[1]Research Scholar, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Education for Women, School of Engineering. E-mail: sabeena.bheeman@gmail.com
[2]Professor, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Education for Women, School of Engineering. E-mail: prof.sivakumari@gmail.com

INDEX TERMS: Health Informatics, PD Classification, KPCA (Kernel based Principal Component Analysis), Dimensionality reduction, deep learning, Fuzzy Monarch Butterfly Optimization Algorithm (FMBOA), FCBi-LSTM (Fuzzy Convolution Bi-Directional Long Short-Term Memory) and vocal features.

1. INTRODUCTION

The next most universal neurodegenerative infection is PD which is a problem that occurs in the central nervous system. Throughout the world, the total number of individuals affected by PD has been increased. Though, the cause for PD is unfamiliar, early identification of the disease can be appreciably lessen the symptoms of PD [1]. By looking into shiver, rigidness, slow movement, asymmetric motor indication and weak posture Parkinson's disease can be categorized [2].

Various studies have indicated that 90% of people who suffer from PD have speech and vocal problems which include dysphonia, monotone and hypophonia [3]. Thus, the degradation of voice is considered to be as the initial symptom of PD. Analysis of voice measurement is simple and non-invasive. Hence, recent Parkinson's disease identification research focus on health methods related vocal problems. Most of the work use many speech signal processing methods to acquire medically related features and those features are used by many artificial learning techniques to attain consistent classification decisions.

Most of the researchers got inspired by the efficiency of speech based diagnostic methods since there is no consistent PD detection system and through which the authors have developed decision supporting tools that can extort features related to dysphonic speech and develop classification method to diagnose PD individuals using speech based feature selection, extraction and classification [4-6].

Most frequently employed machine learning techniques for classifying PD are ANN and SVM [7-8]. The performance the techniques depend on the excellence of selected features from the dataset. Even though selecting relevant features manually is difficult from the inherent properties of audio data, deep learning techniques can learn the dormant properties of data automatically.

Most of the work considered acoustic measurements of dysphonia as dominant feature to differentiate the normal and affected individuals to resolve this issue [9]. When features are redundant, feature selection technique is required before prediction. FS reduces the size of feature set that reduces the training time and less chance for problem of overfitting. Nevertheless none have attempted to compare the performance of diverse feature set with linear and nonlinear classification techniques based on speech data.

The FS technique includes choosing the features that fits best to the problem. Hence, only the related features to the problem are identified instead of considering complete set of features [10]. Feature selection method contains three main levels [11]. Initial process is the generation method. Second, in the assessment stage, features related to the problem are identified by utilizing wrapper or filter technique [12]. Erstwhile, classifier in form of wrapper can also be employed for selecting the features. In generation stage, if complete feasible subsets are generated then the problem become complex and results with increased time for computation (i.e 2*n, n represents the features number) [13]. The metaheuristic technique is the most extensively used AI approach in recent times. This method is mainly applied for feature selection problems [14-15]. Several optimization and feature selection problems acquired reasonable outcome by employing MBO (Monarch Butterfly Optimization) method [16]. The machine learning techniques are also used for various fields of applications like WSN, MANET, which is discussed in [33].

Considering their performance in these fields motivated the researchers to employ Deep Neural Network for classifying PD [17]. This work proposed a novel technique of comparing the performance measures with varied sets of features like Fuzzy Monarch Butterfly Optimization Algorithm (FMBOA) based feature sets and feature extracted through Kernel based Principal Component Analysis (KPCA) feature reduction approach. Finally, selected features were given as input to FCBi-LSTM (Fuzzy Convolution Bi-Directional Long Short-Term Memory) classifier. This is based on CNN which is used to differentiate PD individuals from others. The completion of classification is done by feeding the combined feature representations to successive convolutional layer and fully-connected layers. Leave-One-Person-Out Cross Validation approach is utilized for assessing the performance of the introduced work. Various classification techniques are experimented to analyze the metrics of performance. The outcome of assessment would help the health professionals to distinguish PD individuals from other individuals using their speech data.

2. LITERATURE REVIEW

Chen et al [18] developed an efficient and effective Parkinson’s disease diagnosis model based on FKNN (Fuzzy K-Nearest Neighbor). The developed framework was evaluated against other techniques that are based on SVM. To enhance the accuracy of PD diagnosis system, PCA (Principle Component Analysis) was used to create more discriminative set of new features from which the Fuzzy K-Nearest Neighbor is developed. Hopefully, the developed model will be an efficient tool for PD identification.

Behroozi and Sami [19] introduced a two-fold step approach where the feature selection is performed using PCC (Pearson Correlation Coefficient) along with two methods namely, MCFS and A-MCFS to identify the extremely connected features for the class labelling. The classification structure employs k-NN, SVM, NB and discriminant analysis for the samples of speech data , for instance, for the vowel “a”, a individual classifier is employed instead of using one classifier for complete vocals to identify more representative vocal among all so that the minimum discriminating vocal tests are avoided by combining highest voting for proposed approach.

Gunduz [20] developed two approached depending on CNN (convolutional neural network) for classifying PD using voice feature set. Though both approaches are used for combining varied feature set they differ by combined feature sets. The obtained deep features not only provide better classification of Parkinson’s disease individuals from others, it also boost the discriminative power of classification algorithm.

Aich et al [21] developed GA related set of features and PCA related feature reduction approach for identifying the feature set. SVM along RBF is employed for genetic algorithm-based feature sets. The assessment would help the healthcare professionals to identify Parkinson’s disease patient from other individuals based on speech data. Berus et al [22] proposed a feature selection method depending on PCC, KCC, SOM and PCA (Principal Component Analysis) for enhancing the method’s performance and for reducing the data. Most efficient classification algorithm for Parkinson’s disease diagnosis is M-ANN(Multiple Artificial Neural Networks) where the feature selection is not performed and uses the raw data directly.

Sakar et al [23] utilized TQWT for feature extraction for the voice signals of PD patients that has high resolution than normal discrete wavelet transform. The efficiency of TQWT was analyzed against other modern feature extraction approaches for diagnosing Parkinson's disease through speech problems. The outcome proved that Tunable Q-Factor Wavelet Transform had better performance that other voice signal processing methods that are employed for classifying Parkinson's disease. Hasan and Hasan [24] employed ANOVA (Analysis of Variance) method for feature extraction as the database has complete set of features and the uppermost 50 features are identified or extracted using Analysis of Variance F-score values. various machine learning classification techniques were employed and analyzed against other techniques. The proposed method effectively extracted the important features that are used for identifying Parkinson's disease individuals from normal person and it also enhances the accuracy of classification.

2. PROPOSED METHODOLOGY

Motive of the research is to examine the voice signals, extort feasible features and map them to action i.e. PD with three classes and normal. Figure 1 depicts the steps involved in Parkinson's disease detection model. The major steps of the model are extraction of feature, dimensionality reduction, selection and classification. The model is trained and tested before classification. The success or failure of the model is decided by the performance metrics.

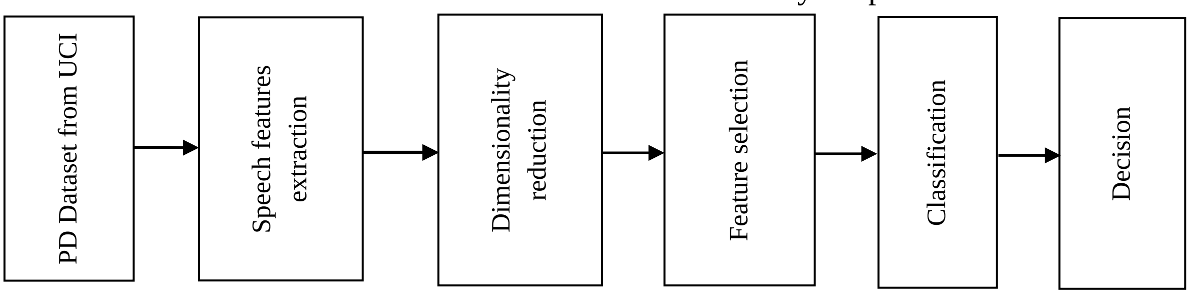

FIGURE 2. STEPS OF THE PROPOSED PARKINSON DETECTION MODEL

Here, a new feature selection with classification method is introduced for the PD classification. The introduced method contains four main steps namely, extracting Vocal (Speech) features from the voice database, dimensionality reduction using Kernel based Principal Component Analysis (KPCA), applying feature selection with Fuzzy Monarch Butterfly Optimization Algorithm (FMBOA), applying a classifier on each subset by Fuzzy

Convolution Bi-Directional Long Short-Term Memory (FCBi-LSTM), finally measuring the results of all classifiers via Accuracy, MCC (Matthews Correlation Coefficient) and f-measure. FCBi-LSTM classifier is proposed by merging multiple speech feature types in the feature-level to identify Parkinson's disease individuals against normal individuals. Leave-One-Person-Out Cross Validation approach is employed for assessing the stability of the introduced method. Figure 2 depicts the flowchart of the introduced approach.

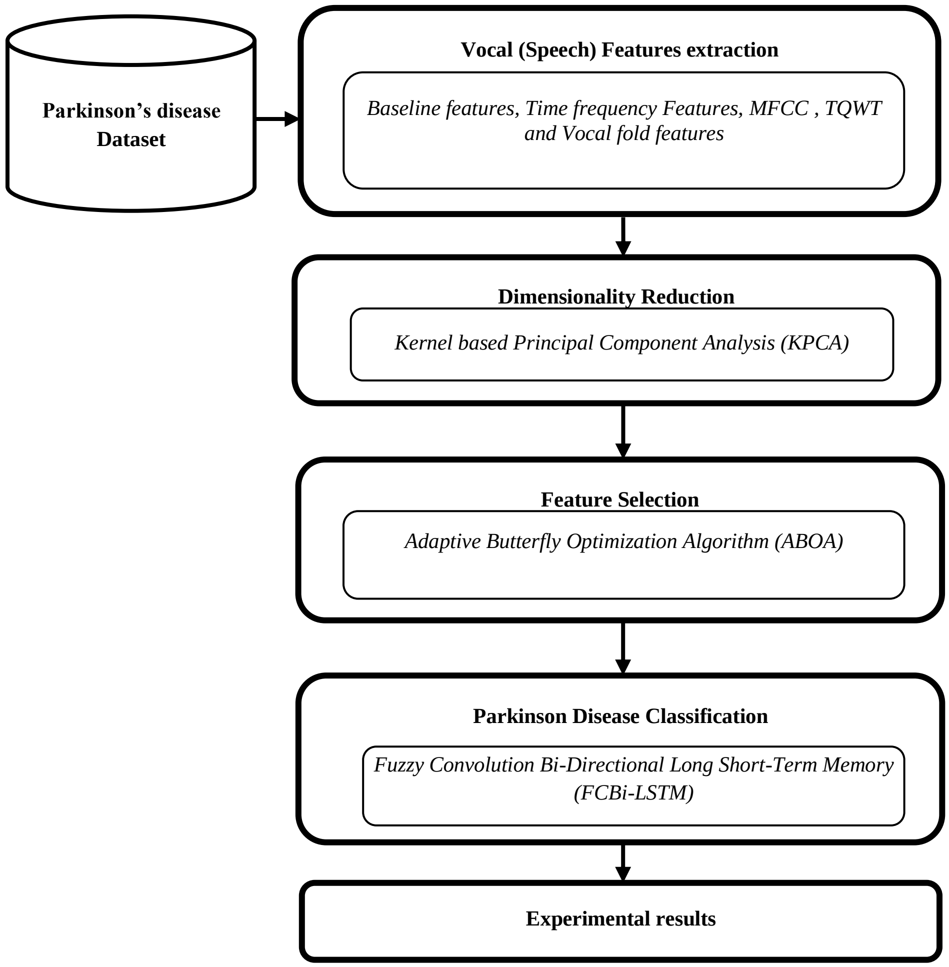

FIGURE 2. WORKING FLOW OF THE PROPOSED METHOD

2.1. Parkinson's disease Dataset

The Parkinson's database containing the features extracted from voice samples are utilized for PD disease diagnosis. Data are obtained from UCI repository [25]. The obtained data is from Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University which contains 188 Parkinson's disease individuals among which 107 men and 81 women, 64 normal person where 23 men and 41 men. PD individual's age varied between 33 to 87 years and normal people age between 41 and 82 years. The frequency output of microphone was initialized to 4.1KHz during acquisition of data and after the review made by the physician, repetitive recurrence of the vowel letter "/a/ " is gathered for ever person three times.

2.2. Feature extraction

The features like Baseline features, Time frequency Features, MFCC, TQWT, and Vocal fold attributes have been obtained from dataset.

i) **Baseline features:** PD affect the voice of the individual in early stage, and hence the voice features are employed for assessing the PD and observe its progress after the treatment of physician. Jitter(#5) and glow based features, fundamental frequency parameters(#5), harmonicity parameters(#2), Recurrence Time Density Entropy (RPDE) (#1), Detrended Fluctuation Analysis (DFA) (#1) and Pitch Period Entropy (PPE) (#1) are most common features of voice that are used for PD diagnosis [4], [39]. The features obtained from the data are termed as baseline features [25].

ii) **Time frequency Features :** The features like Intensity Parameters (#3), Formant Frequencies (#4) and Bandwidth(#4).

iii) **Mel-Frequency Cepstral Coefficients (MFCCs):** this approach employs triangular overlapping filter banks to merge cepstral analysis with spectral domain partitioning. In most of the Parkinson's disease related works, Mel-Frequency Cepstral Coefficients are employed to identify fast deteriorations in articular movement such as lips, tongue etc that are affected by Parkinson's disease [4]. The data contains 84 MFCC-Mel-Frequency Cepstral Coefficients based features that are created using SD and mean of 13 MFCC-Mel-Frequency Cepstral Coefficients in addition to signal log energy and first and second derivatives. To identify the PD effect in vocal tract individually from vocal folds (#84).

iv) **Wavelet Transform (WT):** is a famous tool used for decision making related to signals with little fluctuations in regional scale. Most of the Parkinson's disease work use features acquired from Wavelet Transform using basic frequency of speech signal (F_0). During the collection of data, 10-level discrete wavelet transformation is utilized by voice signals for obtaining WT-based features from raw (F_0) contour and the log transformation of the (F_0) contour. The outcome of the process obtains 182 WT-based features along with Shannon's and the log energy entropy,, energy, Teager-Kaiser energy with approximation and detailed coefficients.

v) **TQWT -Tunable Q-factor Wavelet Transform:** this is an additional technique employed for extracting the features. The benefits of three tunable factors are used by Tunable Q-factor Wavelet Transform namely, Q (Q-factor), r (redundancy) and J (number of levels) to change the quality of signal to better form through the signal's characteristics. Q-factor is related to number of signal oscillations in time domain, j is the levels of decomposition stage. Next to decomposition, there are J + 1 subbands obtained from J high-pass filter and single final output from low-pass filter. Excessive ringing is controlled by r to localize the time wavelet without disturbing shape [40]. From the database, 432 TQWT-related features are obtained through various experiments [37].

vi) **Vocal fold features :** the nose effect on vocal fold are exploited by using features based on vocal fold vibration. For this reason, GQ (#3), GNE (#6), VFER (#7) and EMD features are extracted from data [37].

vii) **Concat features:** these are obtained by combining vocal fold, baseline and time frequency attributes.

2.3. Dimensionality reduction by Kernel Principal Component Analysis (KPCA)

KPCA is a very popular technique for reducing the dimensionality. KPCA finds a lower dimensionality linear subspace than the usual sound recording feature space, where the novel sound recording features of PD acquire the highest variance [28]. Let us believe that the Parkinson Disease (PD) dataset is denoted as $\{a_i\}, i = 1, \dots N$ and each a_i is a D-dimensional sound recorded features vector. At present have to project the PD data into an M-dimensional sound reordered feature subspace, where $< D$. From this vector dimensionality reduced feature vector of the sound recordings are found. Dimensionality reduced feature vector is used for feature selection via Fuzzy Monarch Butterfly Optimization Algorithm (FMBOA).

2.4. Feature selection by Fuzzy Monarch Butterfly Optimization Algorithm (FMBOA)

This work proposes a novel wrapper based feature selection method by FMBOA (Fuzzy Monarch Butterfly Optimization Algorithm) for identifying the optimal feature set. Every sample's feature selection was done by Fuzzy Monarch Butterfly Optimization Algorithm by assessing the feature's influence for Parkinson's disease identification. Identified m samples of features are used by the classifier. Every classifier identifies the class label and assessment measure is used for final output. All the actual features are allocated with feature weights that indicates the significance of the classification problem, and highest weighted features are identified. The MBO is based on the migratory behavior. The relevant feature set was ranked based on importance and fitness value. The proposed Fuzzy Monarch Butterfly Optimization Algorithm resulted in highest classification accuracy when used without any changes and this indicates that there is a stability among local and global search. Here, the attribute of global search of Monarch Butterfly Optimization Algorithm is changed to acquire more accurate outcome and enhance the efficiency to identify exact features from initial stage prior turning to local search. All the butterflies understand the features and correspond to each of them locally, distributing the data between the swarms that resulted in developing ability of the model [29]. It is performed based on two operations such as Migration operator and Butterfly operator adjustment.

2.4.1. Migration operator

Migration method of the Butterfly migration mechanism is explained [30] by equation (1),

$$X_{i,k}^{t+1} = X_{r_1,k}^{t} \qquad (1)$$

where $X_{i,k}^{t+1}$ describe $X_i's$ k^th elements (features of PD) at t +1 production, explaining the ith butterfly's position and $X_{r_1,k}^{t}$ denotes k^th elements from new selected features from PD. r indicates random number described in equation (2),

$$r = rand * peri \qquad (2)$$

here, peri indicates the migration time span. when r> p, equation 3 is used to calculate the new features for the k^th elements of the location,

$$X_{i,k}^{t+1} = X_{r_2,k}^{t} \qquad (3)$$

where $X_{r_2,k}^{t+1}$ defines the k^th elements of X_{r_2} (features of PD) at generation t in r_2, p denotes the ratio of monarch to features . Algorithm 1depicts the steps in butterfly's migration operation

ALGORITHM 1:PSEUDOCODE OF FUZZY MONARCH BUTTERFLY OPTIMIZATION ALGORITHM (FMBOA) FOR MIGRATION OPERATOR

1. Begin
2. for $i = 1\ to\ NP1$ (for each butterfly in sub-population no.1 (samples)) do
3. for $k = 1\ to\ D$ (for each butterfly in i^th *monarch* butterfly(features)) do
4. Compute the r via the equation (2) for optimal selection of features, generate rand via the
   uniform distribution function

5. if $r \le p$ then choose a monarch butterfly (features) in sub-population (samples) 1 randomly say r_1
6. Generate kth elements of X_i by equation (1)
7. else
8. choose a monarch butterfly (features) in sub-population (samples) 2 randomly say r_2
9. create kth elements of X_i by equation (3)
10. end if
11. end for k
12. end for i
13. end

2.4.2. Adjustment of Butterfly operator

Through this approach, the stability is attained by calculating p ratio value among the migration route between feature 1 and 2. If p>, symbolizes the number of feasible butterflies in feature 1has larger fitness value than feature 2 set and vice versa. If rand<=p, location of butterflies are changed. The butterfly position are changed using equation 4,

$$X_{j,k}^{t+1} = X_{best,k}^{t} \tag{4}$$

here $X_{j,k}^{t+1}$ is the kth element of X_j at t + 1 production, that depicts the butterfly location j, and $X_{best,k}^{t}$ describe the kth elements of X_{best} in feature 1 and feature 2 at present production t. Hence, when rand > p, then the equation (5) shifts it,

$$X_{j,k}^{t+1} = X_{r_3,k}^{t} \tag{5}$$

If rand>fitness, the equation (6) will modify the present position,

$$X_{j,k}^{t+1} = X_{j,k}^{t+1} + \alpha * (dx_k - 0.5) * g_w \tag{6}$$

Here fitness imitate the modified feature of butterfly and dx is the walking step for the j feature that is calculated by flying the Lévy as indicated in equation (7),

$$dx = Levy\left(X_j^t\right) \tag{7}$$

And α in equation (7) is a weighted variable measure given in equation (8),

$$\alpha = S_{max}/t^2 \tag{8}$$

And g_w in equation (6) is a Gaussian weighted variable measured given in equation (9),

$$g_w = gaussmf(X, params) = f(X, \sigma, c) = e^{\frac{-(X-c)^2}{2\sigma^2}} \tag{9}$$

where σ standard deviation of the feature and c mean of the feature . Where S_{max} indicate the highest walking butterfly length in one step, t is the present production. Algorithm 2 explains the steps involved in butterfly-adjusting approach.

ALGORITHM 2:PSEUDOCODE OF FUZZY MONARCH BUTTERFLY

OPTIMIZATION ALGORITHM (FMBOA) FOR ADJUSTING OPERATOR

1. Begin
2. for $j = 1\ to\ NP2$ (for each butterfly in sub-population no.2 (samples)) do
3. Get the walking step $(dx)$for the j butterfly(feature) by equation (7)
4. Get the weighted variable (α) & Gaussian weighted variable(g_w) for the j butterfly(feature) by equations (8,9)
5. for $k = 1\ to\ D$ (for each butterfly in ith *monarch* butterfly(features)) do
6. Compute the r via the equation (2) for optimal selection of features, generate rand via the uniform distribution function
7. if $rand \leq p$ then get the kth elements of the $X_{j,k}^{t+1}$ in equation (4)
8. else select a monarch butterfly (features) FMBOA in sub-population (samples) 2 randomly say r_3
9. Generate kth elements of X_i by equation (5)
10. if rand>fitness then
11. Select new feature position of the PD via the equation (6)
12. end if
13. end if
14. end for k
15. end for i
16. end

Algorithm 3 explains the general behavior of the MBO algorithm after learning the behaviour of the butterflies in algorithm 1 and 2[44].

ALGORITHM 3:PSEUDOCODE OF FUZZY MONARCH BUTTERFLY OPTIMIZATION ALGORITHM (FMBOA)

1. Begin
2. While $\left(G < Max_{gen}\right)$
3. Produce sub-population no.1 for each butterfly (features) with $NP1$ better features
4. sub-population no.2 for each butterfly (features) with $NP2$ on the remainders
5. produce the offspring result for sub-population 1
6. produce the offspring result for sub-population 2

7. Select new feature position of the PD via the equation (6)
8. end while
9. find the global best features from PD for classification
10. end

2.5. PD classification via Fuzzy Convolution Bi-Directional Long Short-Term Memory (FCBi-LSTM) classifier

In this work, classification of Parkinson's disease is performed by Fuzzy Convolution Bi-Directional Long Short-Term Memory (FCBi-LSTM) classifier. The FCBi-LSTM classifier calculates the fuzzy weight via the membership function and modify the Parkinson features retort to obtain the feasible features of the identified Parkinson features. FCBi-LSTM classifier employs the features of Parkinson and hidden status for stepwise Parkinson classification and maintains the context data in the internal memory state to extract the association among PD features.

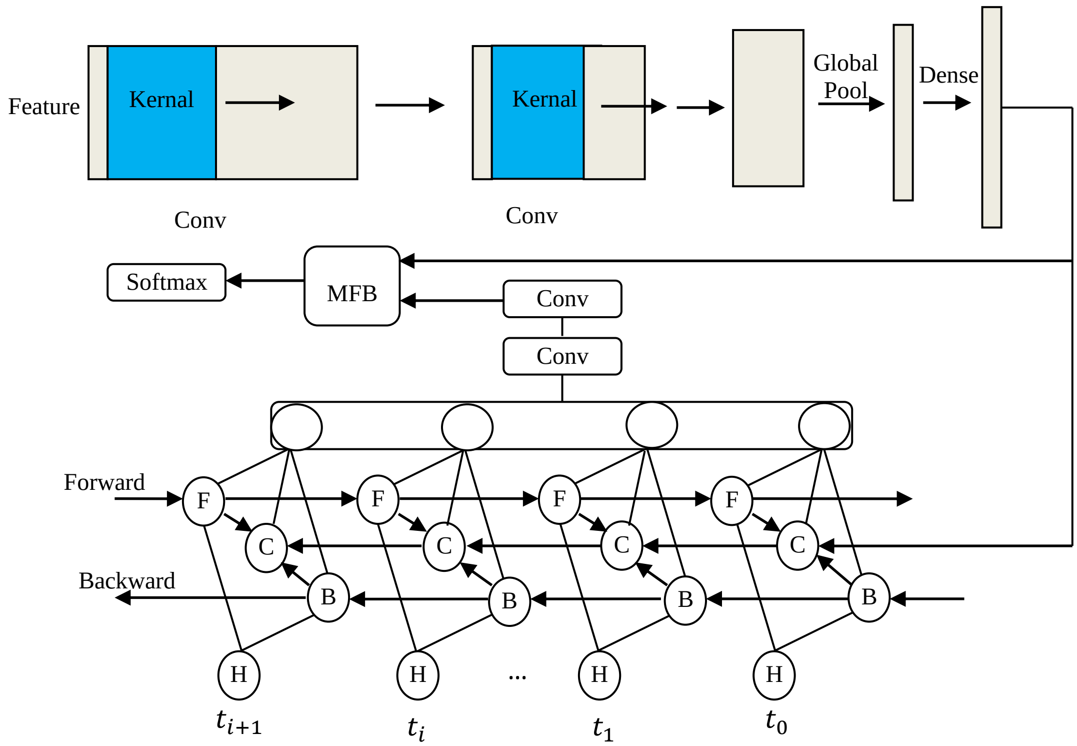

FIGURE 3. PROCESS CHART OF THE PROPOSED FCBi-LSTM CLASSIFIER

The structure of network of FCBi-LSTM is depicted in Figure 3. for FCBi-LSTM, convolutional neural network is employed to process selected features from PD dataset. CNN is a deep learning method employed for solving classification issues [31]. The network contains convolutional and max pooling layers. Convolution followed by pooling is performed and the result is fed to following convolutional layer and etc. CNN characters provides vital benefits for extracting image features. Depending on the information of biological visual cells' local perception of, the convolutional neural network employs partial filters for convolution. Particularly the inner product is obtained from the operation among the input term's local sub-matrix and the filter that is local. Through equation 10, the improved representation of feasible features from Parkinson's disease database, the filters used to produce multiple output matrices from convolutional layer, size of every resultant matrix is (N−m+1), the complete procedure is done.

$$x_i^{l,j} = f\left(b_j + \sum_{a=1}^{m}\sum_{b=1}^{n} fw_{ab}^{j} x_{i+a}^{l-1,j+b}\right) \tag{10}$$

Among $x_i^{l,j}$ there are l convolutional layer, i is the value of the i^{th} convolution output matrix, and j indicates the number of the equivalent output matrix for PD classification. Left and right indices values are 0 to N, and N is the output matrix's number of convolutions. f is sigmoid function.

CNN pooling layer decreases the matrix dimension and does not demolish the inherent links of optimal features. the outcome of the convolutional layer is input for average pooling layer and the outcome of this will be the input for next consecutive layer. Detailed operation procedure by equation (11),

$$x_i^{l,j} = \frac{1}{a.b} \times \sum_{i=1}^{a}\sum_{j=1}^{b} x_{i,j} \tag{11}$$

where $x_i^{l,j}$ indicates an result term for local pairs after pooling. The result of the final convolutional layer is utilized as middle variables to contribute in Bi-LSTM procedure [32]. Bi-LSTM is a type of RNN. Bi-LSTM is employed to encode selected features of PD samples into vectors and code them as h_N . Bi-LSTM can follow the information that are required for analyzing PD samples. Bi-LSTM merge the selected feature vector, h_0, with the second PD sample feature vector to generate new vector h1. Then h_1 continues to combine with the next PD sample feature vector to generate h_2, and etc, till vector h_N . Recurrent neural network is a time series technique. The present outcome is determined by the present input and the prior state. Generally, it is supposed that the provided input selected feature vector of PD samples are represented by the following expression $x = \{x_1, x_2, \cdots, x_t, \cdots, x_T\}$, in which represents the t^{th} PD sample and the sum of PD samples is T . So, can get the following by equation (12),

$$h_t = \sigma_h(FW_{xh}x_t + FW_{hh}h_{t-1} + b_h) \tag{12}$$

where $h_t$denotes the result of the hidden layer in t time, FW_{xh} is the fuzzy weight matrix from the input to hidden layer, FW_{hh} represent the fuzzy weight matrix of hidden layer, and b_h is hidden layer's bias, and σ_h is the function for activation. It is derived using equation (13)

$$y_t = \sigma_y(FW_{ho}h_t + b_o) \tag{13}$$

where y_t is the t^{th} sample's predict label, W_{ho} represent the weight matrices between hidden layer and result, $b_o$is the bias for the result, σ_y is the activation function. Additionally, the outer recurrent neural network cycle contains the inner "LSTM cell" cycle as self loop. Hence, LSTM does not give non-linear element to input transformation and loop cells. Nevertheless, forget gate $f_i^{(t)}$controls the ring weight and sigmoid weight is set to 0 and 1 by equation (14),

$$f_i^{(t)} = \sigma\left(b_i^f + \sum_j FU_{i,j}^f x_j^{(t)} + \sum_j FW_{i,j}^f h_j^{(t-1)}\right) \tag{14}$$

here x_t represent present input vector, h_t represent present vector of hidden layer, and $h_t$contains combined output of LSTM cells b^f , FU^f and FW^f are the bias's fuzzy cyclic weights, input fuzzy weight and forgetting gates correspondingly. Fuzzy weights are computed via the Gaussian Membership function. Hence the LSTM cell's internal state is modified as follows, where there is a conditional self-ring weight $f_i^{(t)}$ by equation (15),

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma\left(b_i + \sum_j FU_{i,j} x_j^{(t)} + \sum_j FW_{i,j} x_j^{(t-1)}\right) \tag{15}$$

here b, FU and FW represent bias, weight of input fuzzy and weight of cyclic fuzzy of the forgetting gates in LSTM cells, correspondingly. $g_i^{(t)}$external input gate unit is alike to the oblivion gate as its has own parameters. $g_i^{(t)}$ external input gate unit is represented by equation (16),

$$g_i^{(t)} = \sigma\left(b_i^g + \sum_j FU_{i,j}^g x_j^{(t)} + \sum_j FW_{i,j}^g h_j^{(t-1)}\right) \tag{16}$$

$h_j^{(t)}$ LSTM cell output is closed by $q_j^{(t)}$, the output gate by equation (17),

$$q_j^{(t)} = \sigma\left(b_i^o + \sum_j FU_{i,j}^o x_j^{(t)} + \sum_j FW_{i,j}^o h_j^{(t-1)}\right) \tag{17}$$

where b^o, FU^o and FW^o are the bias' cyclic weights, weight of the input fuzzy and the forgetting gate, correspondingly. During these deviation, state of cess s(t), additional input with fuzzy weight i and three gates of i can be chosen. It need three extra parameters. Though LSTM captures long-term sequence information, it considers only single direction. That is, LSTM current frame is influenced by current state. In order to build up the relationship, the next Parkinson's disease sample is considered while dealing with present Parkinson's disease sample. Bi-LSTM is most appropriate for current situation. Forward LSTM is the initial layer and next will be the backward LSTM. The last result is computer by equation (18-19),

$$h_t = \alpha h_t^f + \beta h_t^b \tag{18}$$

$$y_t = \sigma(h_t) \tag{19}$$

where h_t^f represent the forward LSTM result that consider PD samples between x_1 and x_T for input, h_t^b indiactes the backward LSTM output that consume PD samples between x_T and x_1, α and β are the significance of forward and backward LSTM ($\alpha + \beta = 1$), h_t indicates the two LSTM's element-wise, y_t represnt classification results of the classifier for PD diagnosis. Bi-LSTMs obtains organization details, to execute superior than single directional LSTMs. Calculation equation of Bi-LSTM's final result is represented through equations (20,21,22),

$$h_t^c = h_{t-1}^c + \tanh\left(FW_1 h_{t-1}^c + FW_2\left(h_{t-1}^f + h_{t-1}^b\right) + b_h\right) \tag{20}$$

$$h_t = \alpha h_t^f + \beta h_t^b + \gamma h_t^c \tag{21}$$

$$y_t = \sigma(h_t) \tag{22}$$

here $h^c{}_t$ indicates selected features contributing in Bi-LSTM for PD diagnosis like $\alpha + \beta + \gamma = 1$. Bi-LSTM final result is analyzed by two convolutional layers to acquire PD diagnosis. Multi-modal factorized bilinear pooling employed to combine features processed by CNN and Bi-LSTM analyzed features. hence, the equation (23) of MFB is described as

$$y_i = 1^T\left(FU_i^T x \circ FV_i^T z\right) + b \tag{23}$$

where FU_i and FV_i weights; b ,bias weight; x, z features that are to be combined by two classifiers, $y_i$is the it h value of the fused feature with classification results.

3. EXPERIMENTAL RESULTS

This part describes the outcomes of the assessment resulted from introduced Fuzzy Convolution Bi-Directional Long Short-Term Memory (FCBi-LSTM) classifier, and those results are compared to the methods like Fuzzy Convolution Long Short-Term Memory based Convolutional Neural Network (FCLSTM-CNN), CNN, and SVM. since there are minimum instances in the database, LOPO CV is employed for performance evaluation where a particular individual instance are considered for testing and other instances are utilized for training. As every person have 3 recordings, the class label is identified by considering the majority class obtained by the recordings. The assessment on the arrhythmia recognition and classification method were performed by MIT-BIH arrhythmia database via Matrix LABoratory R2016 a (MATLAB R2016a). The implementation has been done with following system specifications: Intel(R) Core™i3-4160T CPU@3.10 GHz 3.09 GHz processor, 4.00 GB RAM, Windows 8.1 pro, 64 bit operating system, operation system, and 1 TB hard disk.

3.1. PERFORMANCE METRICS

The prediction performance of the classification method is analyzed using evaluation measures. Though, accuracy is widely used measure, this provide ambiguous results for unbalanced class distributed data. The measures like MCC and f-measure can be used to identify the efficiency of the classifier in differentiating the varied class, even for imbalanced class. Table 1 depicts the confusion matrix that indicates the number of incorrectly and correctly classified cases for binary classification. tp, fp, fn and tn denote true positive (tp), false positive (fp), false negative (fn) and true negative (tn) numbers in confusion matrix correspondingly. Depending on these counts, F-Measure, accuracy and error is computed as by equations (26,27,28),

$$Precision = \frac{tp}{tp + fp} \tag{24}$$

$$recall = \frac{tp}{tp + fn} \tag{25}$$

$$F - measure = \frac{2 * precision * recall}{precision + recall} \tag{26}$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{27}$$

Error=100-Accuracy **(28)**

MCC (Matthews Correlation Coefficient) measure is employed for identifying the binary classifier's quality. MCC uses the number of tp, fp, fn and tn. It is considered to be a balanced metric that handles unbalanced class distribution and it is a correlation coefficient among actual and predicted cases that has value in the range [-1,+1]. When it is +1 then perfect prediction, for -1 there is a difference in actual label and predicted one.

TABLE 1. CONFUSION MATRIX FOR BINARY CLASSIFICATION

ACTUAL/ PREDICTED AS	POSITIVE	NEGATIVE
POSITIVE	tp	fn
NEGATIVE	fp	tn

3.2. RESULTS COMPARISON

Experiment is done with three types of features with classifiers (Proposed FCBi-LSTM classifier, existing classifiers such as CNN, SVM and FCLSTM-CNN correspondingly). The measures employed for analyzing the classifier performance are accuracy, Error, F-Measure and MCC. For final feature-level combination, new analysis are done using triple feature sets. While the combination of TQWT+MFCC +Wavelet features acquired accuracy of 94.3965**%** (for F-Measure 95.3965**%**), the accuracy of TQWT+MFCC+Concat and TQWT+ Wavelet + Concat are 95.7455**%** and 94.2252**%** respectively. Combination excluding TQWT (MFCC+Wavelet+Concat) has light inferior performance against other methods for metrics accuracy, F-Measure and MCC.

TABLE 2. FEATURE-LEVEL COMBINATION RESULTS: TRIPLE FEATURE SETS

FEATURE COMBINATION	F-MEASURE (%)	ACCURACY (%)	ERROR(%)	MCC (%)
TQWT+MFCC+Wavelet	95.3965	94.3965	5.6035	71.3513
TQWT+MFCC+Concat	93.3252	95.7455	5.2545	69.4017
TQWT+ Wavelet + Concat	94.4252	94.2252	7.7748	65.9483
MFCC + Wavelet + Concat	95.4530	96.2865	3.7135	67.9917

TABLE 3. RESULTS OF CLASSIFIERS WITH TRIPLE FEATURE (KPCA+ FMBOA)

SVM CLASSIFIER (%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR**	**MCC**
TQWT+MFCC+Wavelet	81.7152	84.2025	15.7975	54.9000
TQWT+MFCC+Concat	79.5960	82.7640	17.2360	53.3000
TQWT+ Wavelet + Concat	85.1589	86.4662	13.5338	57.5000
MFCC + Wavelet + Concat	86.3510	87.2293	12.7707	58.4000
CNN CLASSIFIER (%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR**	**MCC**
TQWT+MFCC+Wavelet	84.6697	86.6697	13.3303	55.9007
TQWT+MFCC+Concat	89.0315	91.0315	8.9685	60.1656
TQWT+ Wavelet + Concat	85.1695	87.1695	12.8305	62.0993
MFCC + Wavelet + Concat	91.2752	93.2752	6.7248	63.2384
FCLSTM-CNN CLASSIFIER(%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR**	**MCC**
TQWT+MFCC+Wavelet	93.0257	92.1457	7.8543	66.3669

TQWT+MFCC+Concat	90.3252	92.1854	7.8146	66.4060
TQWT+ Wavelet + Concat	92.2252	92.2252	7.7748	64.1457
MFCC + Wavelet + Concat	90.4729	94.2557	5.7443	65.9960
FCBi-LSTM CLASSIFIER (%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR**	**MCC**
TQWT+MFCC+Wavelet	95.3965	94.3965	5.6035	71.3513
TQWT+MFCC+Concat	93.3252	95.7455	4.2545	69.4017
TQWT+ Wavelet + Concat	94.4252	94.2252	5.7748	65.9483
MFCC + Wavelet + Concat	95.4530	96.2865	3.7135	67.9917

The combination of MFCC + Wavelet + Concat features for SVM classifier provides accuracy of 86.3510%. based on feature models, there is increased performance for TQWT+MFCC+Wavelet, TQWT+ Wavelet + Concat, and TQWT+ Wavelet + Concat feature sets (Table 3). For triple feature results (Table 3), MFCC + Wavelet + Concat combination provides enhanced performance against other classifiers for all metrics. though the accuracy of MFCC + Wavelet + Concat is 91.2752% for CNN classifier that. According to feature and FCLSTM-CNN classifier gives the accuracy results of 92.1457%, 92.1854%, 92.2252% and 94.2557% respectively for TQWT+MFCC+Wavelet, TQWT+Wavelet+Concat, TQWT+MFCC+Concat and MFCC+Wavelet+ Concat feature sets. From the table 3 it concludes that MFCC + Wavelet + Concat give improved performance against all classifiers for all measures. MFCC+Wavelet+Concat combination with FCBi-LSTM classifier attains enhanced performance with achieves accuracy of 96.2865% (F-Measure as 95.4530% and 67.9917% for MCC) (See table 3).

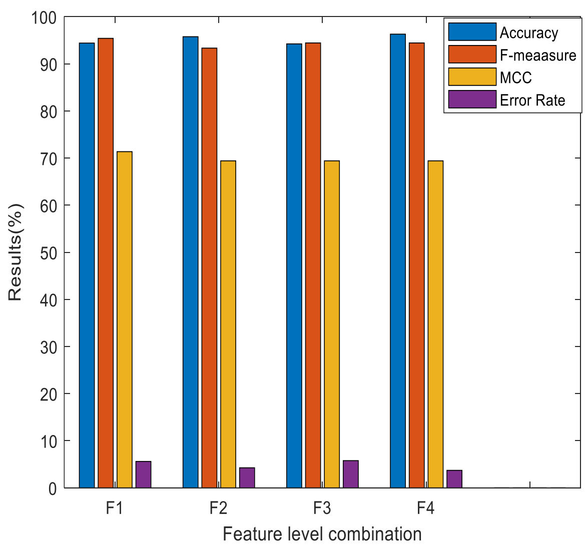

FIGURE 4. FEATURE LEVEL COMBINATION RESULTS VS. METRICS

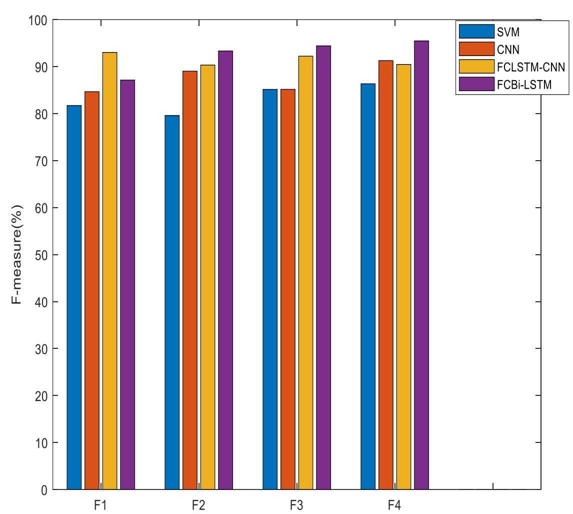

FIGURE 5. F-MEASURE RESULTS OF FEATURE LEVEL COMBINATION VS. CLASSIFIERS

Figure 4 shows outcome for measures such as f-measure, accuracy, MCC and error with combination of the feature set. The results are taken by considering the dimensionality reduction and feature selection. From the results it concludes that the TQWT+MFCC+Wavelet combination of feature set gives higher results of 95.3965%, 94.3965%, 71.3513% and 5.6035% for f-measure, accuracy, MCC and error when compared to other feature combination sets. The proposed work gives higher results for feature combination since the feature selection is performed by using the metaheuristic algorithm. Figure 5 shows the f-measure results comparisons of four different feature level combinations are measured via different classifiers. In the x-axis results are measured via the feature level combination of classifiers such as SVM, CNN, FCLSTM-CNN, FCBi-LSTM and results are shown in y axis as f-measure. The proposed FCBi-LSTM classifier with first feature level combination gives higher f-measure results of 95.3965%, whereas other methods such as SVM, CNN, and FCLSTM-CNN gives f-measure results of 81.7152%, 84.6697% and 93.0257% respectively at the first feature level combination.

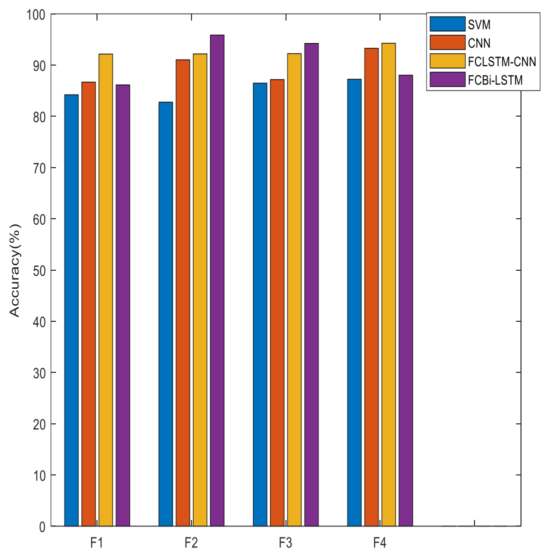

FIGURE 6. ACCURACY RESULTS OF FEATURE LEVEL COMBINATION VS. CLASSIFIERS

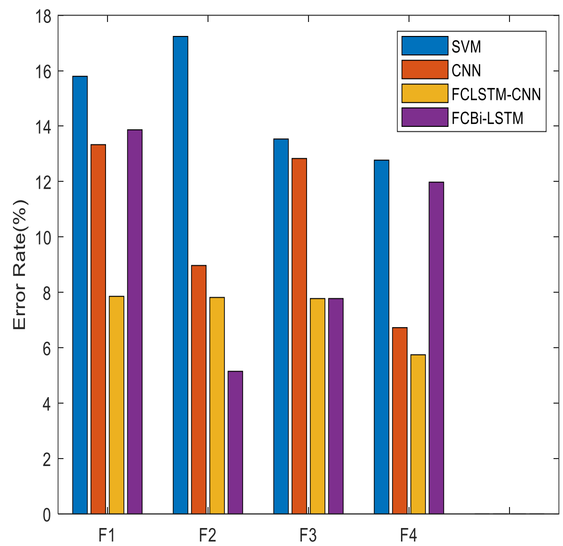

FIGURE 7. ERROR RESULTS OF FEATURE LEVEL COMBINATION VS. CLASSIFIERS

Figure 6 shows the x-axis results are measured via the feature level combination of classifiers such as SVM, CNN, FCLSTM-CNN, FCBi-LSTM and results are shown in y axis as accuracy. The proposed FCBi-LSTM classifier with final feature level combination gives higher f-measure results of 96.2865%, but other methods like SVM, CNN, and FCLSTM-CNN gives accuracy results of 87.2293%, 93.2752% and 94.2557% respectively at the last feature level combination. Error results comparison of classifiers with four different feature level combinations is shown in the figure 7. From the figure 7 it concludes that the proposed FCBi-LSTM classifier with final feature level combination produces lesser error results of 3.7135%, SVM, CNN, FCLSTM-CNN methods gives higher error value of 12.7707%, 6.7248% and 5.7443% respectively at last feature level combination.

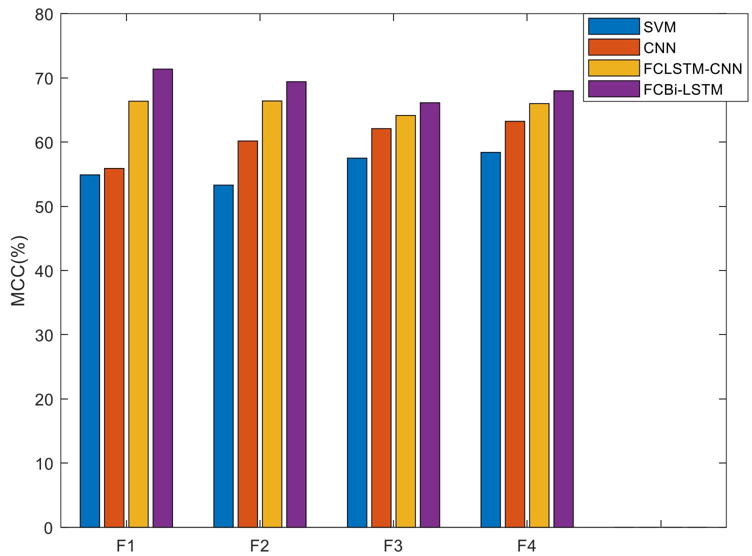

FIGURE 8. MCC RESULTS OF FEATURE LEVEL COMBINATION VS. CLASSIFIERS

Figure 8 shows the results of MCC with combination of the feature set under the four classifiers. From the results it concludes that the TQWT+MFCC + Concat feature set gives higher results of 71.3513% for proposed algorithm, 54.9%, 55.9007% and 66.3669% for SVM, CNN, FCLSTM-CNN classifiers respectively(1^{st} feature level combination). The SVM and CNN classifier gives better results for third feature set when compared to other feature combination sets.

4. CONCLUSION AND FUTURE WORK

Deep Learning (DL) method for clinical database has been developed, and the prospective of Artificial Intelligence (AI) algorithms to solve the feature selection issue, examine the scope of DL based models that are trained with the feasible feature subset of Parkinson's Disease dataset for enhanced classification. In this study, deep FCBi-LSTM classifier is proposed for classifying PD through voice features. starting from input dataset, the types of feature in the database like TQWT, Wavelet, MFCC and Concat have been extracted for classification of PD detection. From extracted features, KPCA based dimensionality reduction is employed for input dataset. Important features of the dataset are selected via Fuzzy Monarch Butterfly Optimization Algorithm (FMBOA) based feature selection algorithm, while eliminating redundant features. Finally FCBi-LSTM classifier is proposed which uses parallel convolution layers related to every feature set to form feature representations automatically and directly. FCBi-LSTM classifier is employed to discover the relationship between features and predict them one after the other. Training of the classifier is performed by the data acquired from UCI data house. From the outcomes it concludes that introduced classifier gives higher accuracy results when compared to FCLSTM-CNN, CNN, and SVM classifier respectively. Based on the instance number in the database, classifier's predicting performance is analyzed using F-Measure, Accuracy, Error and MCC- Matthews Correlation Coefficient measures. All metrics seem to be very promising and effective for analyzing the discriminative power of the introduced classifier. As further work, network can be fed with varied types of data as their input.

REFERENCES

1. Rusz J., R. Cmejla, H. Ruzickova, and E. Ruzicka, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease," The Journal of the Acoustical Society of America, vol. 129, no. 1, pp. 350–367, 2011.
2. Massano J. and K. P. Bhatia, "Clinical approach to Parkinson's disease: features, diagnosis, and principles of management," Cold Spring Harbor Perspectives in Medicine, vol. 2, no. 6, Article ID a008870, 2012.
3. Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., ... & Kursun, O. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics, 17(4), 828-834.
4. Islam MS, Parvez I, Deng H, Goswami P. Performance comparison of heterogeneous classifiers for detection of Parkinson's disease using voice disorder (dysphonia). 2014 International Conference on informatics, electronics and vision (ICIEV); 2014. pp. 1–7.
5. Benba A, Jilbab A, Hammouch A, Sandabad S. Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease. 2015 International Conference on electrical and information technologies (ICEIT), IEEE; 2015.
6. Pérez CJ, Naranjo L, Martín J, Campos-Roca Y. A latent variable-based Bayesian regression to address recording replications in Parkinson's disease. 22nd European signal processing conference; 2014. pp. 1447–51.
7. Sharma A. and R. N. Giri, "Automatic recognition of parkinson's disease via artificial neural network and support vector machine," International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 4, no. 3, pp. 2278–3075, 2014.
8. Shahbakhi M., D. T. Far, E. Tahami et al., "Speech analysis for diagnosis of parkinson's disease using genetic algorithm and support vector machine," Journal of Biomedical Science and Engineering, vol. 7, no. 4, pp. 147–156, 2014.
9. Tsanas A., M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for highaccuracy classification of Parkinsons disease," IEEE Transactions on Biomedical Engineering, vol. 59, no. 5, pp. 1264–1271, 2012.
10. Mohamed AW, Hadi AA, Jambi KM (2019) Novel mutation strategy for enhancing SHADE and LSHADE algorithms for global numerical optimization. Swarm and Evolutionary Computation 50:100455
11. Djellali H, NG Zine, and N Azizi (2016). "Two stages feature selection based on filter ranking methods and SVMRFE on medical applications," In Modelling and Implementation of Complex Systems: Springer, pp. 281–293
12. Apolloni J, Leguizamón G, Alba E (2016) Two hybrid wrapperfilter feature selection algorithms applied to high-dimensional microarray experiments. Appl Soft Comput 38:922–932.
13. Tang, J., Alelyani, S. and Liu, H., 2014. Feature selection for classification: A review. Data classification: Algorithms and applications, pp.1-33.
14. Gokalp O, Tasci E, Ugur A (2020) A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. Expert Syst Appl 146:113176
15. Alweshah M, Abdullah S (2015) Hybridizing firefly algorithms with a probabilistic neural network for solving classification problems. Appl Soft Comput 35:513–524

16. Alweshah M, Al Khalaileh S, Gupta BB, Almomani A, Hammouri AI, Al-Betar MA (2020) The monarch butterfly optimization algorithm for solving feature selection problems. Neural Comput & Applic:1–15
17. Eskofier B. M., S. I. Lee, J.-F. Daneault, F. N. Golabchi, G. Ferreira- Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, T. Kautz et al., "Recent machine learning advancements in sensor-based mobility analysis: Deep learning for parkinson's disease assessment," in 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2016, pp. 655–658.
18. Chen, H.L., Huang, C.C., Yu, X.G., Xu, X., Sun, X., Wang, G. and Wang, S.J., 2013. An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. Expert systems with applications, 40(1), pp.263-271.
19. Behroozi, M. and Sami, A., 2016. A multiple-classifier framework for Parkinson's disease detection based on various vocal tests. International journal of telemedicine and applications, Vol. 2016, no. 6837498, pp.1-9 .
20. Gunduz, H., 2019. Deep learning-based Parkinson's disease classification using vocal feature sets. IEEE Access, 7, pp.115540-115551.
21. Aich, S., Kim, H.C., Hui, K.L., Al-Absi, A.A. and Sain, M., 2019, A supervised machine learning approach using different feature selection techniques on voice datasets for prediction of Parkinson's disease. In 2019 21st International Conference on Advanced Communication Technology (ICACT) ,pp. 1116-1121.
22. Berus, L., Klancnik, S., Brezocnik, M. and Ficko, M., 2019. Classifying Parkinson's disease based on acoustic measures using artificial neural networks. Sensors, 19(1), pp.1-15.
23. Sakar, C.O., Serbes, G., Gunduz, A., Tunc, H.C., Nizam, H., Sakar, B.E., Tutuncu, M., Aydin, T., Isenkul, M.E. and Apaydin, H., 2019. A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. Applied Soft Computing, 74, pp.255-263.
24. Hasan, K.A. and Hasan, M.A.M., 2020, Classification of Parkinson's Disease by Analyzing Multiple Vocal Features Sets. In 2020 IEEE Region 10 Symposium (TENSYMP) ,pp. 758-761.
25. Tsanas A., M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," IEEE Transactions on biomedical engineering, vol. 59, no. 5, pp. 1264–1271, 2012.
26. Selesnick I. W., "Wavelet transform with tunable q-factor," IEEE transactions on signal processing, vol. 59, no. 8, pp. 3560–3575, 2011.
27. Wang, Q., 2012. Kernel principal component analysis and its applications in face recognition and active shape models. arXiv preprint arXiv:1207.3538.
28. Alweshah, M., 2020. Solving feature selection problems by combining mutation and crossover operations with the monarch butterfly optimization algorithm. Applied Intelligence, pp.1-24.
29. G-GWang, X Zhao, and S Deb (2015). "A novel monarch butterfly optimization with greedy strategy and self-adaptive," In Soft Computing and Machine Intelligence (ISCMI), 2015 Second International Conference on, pp. 45–50

30. Feng Y, Wang G-G, Li W, Li N (2018) Multi-strategy monarch butterfly optimization algorithm for discounted {0-1} knapsack problem. Neural Comput & Applic 30(10):3019–3036
31. Chen T., R. Xu, Y. He, and X. Wang, “Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN,” Expert Syst. Appl., vol. 72, pp. 221–230, Apr. 2017
32. Xingjian, S.; Chen, Z.;Wang, H.; Yeung, D.;Wong,W.;Woo,W. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 2015; pp. 802–810.
33. Shajan Joseph, A Rajaram, “A Novel Enhanced SVM cluster based secure and effective routing protocol for node authentication in Mobile Ad Hoc Networks,” International Journal of Computer Technology and Applications (IJCTA), 10(19): 25-39, 2017.