Journal of Theoretical and Applied Information Technology
15th May 2023. Vol.101. No 9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ENSEMBLE FEATURE SELECTION AND ENSEMBLE DEEP LEARNING (EDL) CLASSIFIER FOR PARKINSON'S DISEASE

B. SABEENA, S. SIVAKUMARI,

Research Scholar, Department of Computer Science and Engineering, Avinashilingam Institute forHome Science and Education for Women, School of Engineering

Professor, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Education for Women, School of Engineering.

E-mail: sabeenar09@gmail.com

ABSTRACT

PDs (Parkinson's diseases) are chronic neuro-degenerative conditions that impact humans in their day to day lives. Diagnosis and monitoring of these conditions based on limited physical symptoms are painstaking evaluations for medical professionals and clinicians may miss early prodromal phases. Though many DMTs (data mining techniques) for automated assessments of PDs have recently been presented, their performances get reduced due to dataset's irrelevant features. EFSs (Ensemble Feature Selections) have more benefits than single FSAs (Feature Selection algorithms) as they address drawbacks by mixing different models and improve outcomes of MLTs (machine learning techniques). This work uses OBEFSs (Optimization Based Ensemble Feature Selections) including FMBOAS (Fuzzy Monarch Butterfly Optimization Algorithms), LFCSAs (Levy Flight Cuckoo Search Algorithms), and AFAs (Adaptive Firefly Algorithms) for selection of features based on their correlations. On selection of features, EDL (Ensemble Deep Learning) classifiers classify PD datasets. EDLs include FCBi-LSTMs (Fuzzy Convolution Bi-Directional Long Short-Term Memories), CAEs (Contractive Auto-encoders), and SAEs (Sparse Auto-encoders). CAEs are robust variant of standards of auto encoders which learn representations with reduced sensitiveness to small variations of data. SAEs are used to train classifiers using NNs (neural networks) for identifying PDs from datasets. Stacked generalization is used to combine the results of DL classifiers. When compared to a single model, EDL techniques offer improved predictive performances. The datasets used for this study were obtain ed from machine learning repositories of UCI (University of California-Irvine). The performance of the classifiers were measured using accuracies, F-measures, MCCs (Matthews Correlation Coefficients), and errors.

Index Terms: *Parkinson's Disease (PDs), Optimization Based Ensemble Feature Selection (OBEFS), Levy Flight Cuckoo Search Algorithm (LFCSA), and Adaptive Firefly Algorithm (AFA), Ensemble Deep Learning (EDL) Classifier, Fuzzy Convolution Bi-Directional Long Short-Term Memory (FCBi-LSTM), Contractive Autoencoder (CAE), and Sparse Autoencoder (SAE).*

1. INTRODUCTION

PDs are degenerative neurological illnesses characterised by low dopamine levels in the brain [1]. Inadequate, intermittent symptom monitoring, infrequent access to care, and few interactions with healthcare experts restrict PDs patient care, resulting in poor medical decision making and suboptimal patient health-related outcomes. PDs are the second most common neurological illnesses behind Alzheimer's [2-3]. Tremors, stiffness, bradykinesia, and postural instabilities are the four main signs of PDs [4]. These symptoms are persistent and degenerative and worsen over a period of time, however they appear in varying degrees and combinations based on individuals. PDs may include both motor and non-motor symptoms that can make it difficult for affected patients to live normally [5]. Over ninety percent of patients with PDs have vocal impairments including dysphonic (difficulty in producing sounds vocally). Recent studies have also discovered relationships between risk alleles counts and health and the probability of developing PDs [6]. Around 1–2 percent of adults over the age of 60 globally have been impacted by the condition.

Systems detecting PDs use many sensors for determining symptom's severities. A major symptom that is evident in patients with PDs is vocal presentations which they suffer in early stages of diseases. Thus, detecting impairments in vocal presentations are primary to investigations of PDs [7–8]. The patient's vocal recordings have been exploited by algorithms for generating relevant features and input into many systems that learnt from these characteristics and used it for classifications. MLTs including ANNs (Artificial Neural Networks) and SVMs (Support Vector Machines) [9] have been common in studies which classified PDs in addition to RFs (Random Forests) [10] and KNNs (K-Nearest Neighbors) [11] which are simple and convenient. The quality of the data attributes are intricately related to success of aforementioned algorithms. Although manually identifying relevant features to characterize inherent attributes of speech (audio) data is complex, data's latent properties can be determined automatically using a DLTs (deep learning techniques).

FSAs play crucial roles in reducing training time by their reductions of features and by their selections of relevant features and eliminations of redundant features with the aim of improving classifier performances. There are three forms of FSAs (selecting subsets of features without assessments) and hybrid techniques. They include wrappers which are search algorithms to find and estimate relevant subsets of characteristics and filters which are combinations of previous methods. Though there are several FSAs, there are no tools or solutions that can objectively determine which algorithms work best with given datasets. Hence, trial-and-error strategies are used in various inquiries. Studies explored variety of FSAs with one or more classifiers before selecting ones that performed best in tests [11,12,13]. Alternatively, ensemble learning algorithms have been described for picking features based on consensus or aggregates of many FSAs [14,15].

In ensemble learning, SIs (swarm intelligences) are decentralised techniques that are self-organizing with collective behaviours . They usually comprise of basic agent groups that interact with one another and their environments at local levels. There are no centralised control structures to decide single agent's behaviours, and agents follow basic principles. On the other hand, interactions of these entities result in emergence of "intelligent" global behaviours . SIs can also use indirect communications to exchange and coordinate information. The increase in communication overhead is minor as the number of people increases. As a result, it is also scalable. Recent works have employed rigorous FSAs based on EFSs where results suggested that the study's proposed approach had significant potential in selecting features from datasets with more features and lesser sample counts. Many research on EFSs have recently been undertaken; some use classifiers, while others do not. Population-based optimization techniques such as ACOs (Ant Colony Optimizations) . GAs (Genetic Algorithms), SAs (Simulated Annealing) . TSs (taboo searches) [20], and PSOs (Particle Swarm Optimizations) have recently been employed as FSAs. Hybrid search strategies that merge wrappers and filters have also been used.

This work's OBEFSs framework aims to assist in the creation of EFSs mechanisms that combine benefits of several FSAs while avoiding biases and compensating for downsides. SIs have been extensively used in feature selections because of advantages namely, their combinations with MLTs produce outstanding results. DNNs (Deep Neural Networks) with their hierarchical layers produce deep abstract representations, which are used as inputs in many MLTs. These performances have prompted researchers to apply DNNs in classification of PDs. DNNs is a potential classifier for classifications of PDs since it can represent intricate and non-linear relationships from data.

Second, the class of provided data samples predicted in classifications are a type of supervised learning. Ensemble learning evaluates a variety of approaches instead of single classification algorithms, and final results are generated by merging outputs of classifiers. Main objective of ensemble classifiers is to combine advantages of numerous classifiers and integrate their outputs, such that Base classifiers are individual classifiers. Ensemble classifiers have two main problems: (1) selecting basic classifiers and (2) aggregating outputs of base classifiers . It is critical to ensure that basic classifiers are sufficiently diversified while forming successful ensembles . Ensemble learning is a powerful approach that combines numerous learning algorithms to increase overall prediction accuracy and may exceed any single smart classifier.

In this research, OBEFSs method has been suggested to choose features based on the PD datasets. The OBEFS algorithm, which seeks to merge numerous feature selection approaches via FMBOAs, LFCSAs, and AFAs, is introduced. The optimal features selected through OBEFS are used to train an EDL (Ensemble Deep Learning) classifier. Stacking generalisation is used to integrate EDL classifiers such as FCBi-LSTMs, CAEs, and SAEs. Proposed method is trained with a dataset taken from machine learning repositories of UCI and their performances. Evaluation metrics such as F-Measure, MCCs, accuracies, and errors are used for the assessment of classification.

2. LITERATURE REVIEW

MLTs for recognizing PDs. Their scheme used hybrid FSAs by combining ACOs and Relief networks. The feature outputs FSAs were used for classifications by SVMs for maximum classification accuracies. The study's schemes were evaluated using K-fold cross validations for justifying hyper

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

parameters. Experimental results of the scheme from real world PD datasets, showed that their proposed system outperformed baseline techniques in recognising PDs from specified attributes. The performances suggested that their approach was highly recommendable for identification of PDs.

a new class of OCSAs (Optimized version of Crow Search Algorithms). Their recommended OCSAs could be used to predict PDs and help people receive appropriate treatments at early stages. The performance of OCSAs were evaluated for 20 benchmark datasets and the results compared to original CCSAs (Chaotic Crow Search Algorithms). The proposed nature-inspired algorithm discovered ideal subsets of characteristics, maximized accuracy while minimizing selected features according to their findings.

an improved FKNNs (Fuzzy KNNs) approach based on voice measures for the early diagnosis of PDs. The suggested method called CBFO-FKNNs used evolutionary instance based learning strategies where CBFOs (Chaotic Bacterial Foraging Optimizations) with Gauss mutations and FKNNs were used. CBFOs examined parameters before tuning them for use by FKNNs. The scheme's obtained values of accuracies, sensitivities, specificities, and AUC (Area Under the Curves) were compared with other methods using ten fold cross validations for PD datasets. The study's suggested scheme outperformed other FKNN models based (Five) on BFOs (Bacterial Foraging Optimizations), PSOs, GAs, FFOs (Fruit Fly Optimizations) and FAs (Firefly Algorithms) and MLTs (three) including SVMs, local learning feature selections based SVMs, and KELMs (kernel Extreme Learning Machines). The proposed approach provided work has a very strong possibility of bringing tremendous ease to physicians in making better clinical diagnosis decisions.

To choose the best features from the speech collection, suggested MAFTs (Multi-Agent Feature Filters). MAFTs aimed at selecting sets of characteristics to improve overall prediction model's performances and reduce over-fits which might occur due to the extreme reduction of features. Furthermore, this approach minimizes prediction complexities, expedites training, and develops robust training models. MAFTs were then combined with ten different MLTs to build a sophisticated voice-based detection model for PDs. To increase classification accuracies for detecting PDs, a Hybrid Model composed of binary CNNs (Convolution Neural Networks) and three FSAa namely, GAs, Adam optimizers, and mini-batch gradient descents was presented. Their results suggested that combination of MAFTs with hybrid modela greatly increased diagnostic outcomes for PDs.

MLTs for comparison of voice measurement features in patient datasets to determine if patients had PDs. The performances of basic classifiers like DTs (Decision Trees), LRs (Logistic Regressions), and KNNs were compared to Ensemble learning classifiers including Bagging, RFs, and Boosting. The accuracies of classifiers were found to be more accurate in predictions of sicknesses. Moreover, important traits needed for classifications were found based on feature's importance. The major purpose of the study was to detect dysphonia and distinguish healthy individuals from patients affected with PDs.

two frameworks based on CNNs for classifying PDs utilising sets of vocal (speech) characteristics. Before being integrated in layers, deep features from parallel branches were collected at the same time. The study's MLTs were trained on machine learning repositories of UCI while outcomes were verified using the metrics of LOPO CVs (Leave-One-Person-Out Cross Validations), accuracies, F-scores and , MCCs as the inputs had unequal class distributions.

Mutual Information Gains, extra trees, GAs, classifiers including NBs (Naive Bayes), KNNs, and RFs as FSAs. The study evaluated performances of various combinations on the speech dataset obtained from machine learning repositories of UCI. The study used SMOTEs (Synthetic Minority Oversampling Techniques) to handle class imbalances as the dataset was substantially uneven in instances. Their results from experiments were very useful.

model-based logics like LRs, KNNs, SVCs (Support Vector Classifiers), GBCs (Gradient Boost Classifiers) and RFs. The study evaluated reliability using five-fold cross-validations similar to curves of ROCs (Receiver Operating Characteristics) and confusion matrices. The study used majority voting, weighted averages, bagging, Ada boost, and GBCs in their ensembles. Their model also identified confusion matrices, five-fold cross-validations, precisions, recall rates, and F1 scores. The study's correlation matrices were also constructed to indicate if these characteristics were connected. Their findings implies that MLTs provided more reliable identifications of PDs in patients when compared to conventional techniques.

created a three-stage ensembles from DLTs for prognosis of PDs. To extract characteristics, the study employed DaTscan and clinical evaluations of motor complaints. Their ensembles of DNNs generated subsets of information gathered from patients four years after the onset of PDs to estimate PDs. Their proposed method was evaluated using MAPEs (Mean

Absolute Percentage Errors), MAEs (Mean Absolute Errors), PCCs (Pearson's Correlation Coefficients) and biases between predicted and observed motor outcome scores. The study compared different data subsets as inputs to individual networks in evaluations.

dual layered stacking ensembles for accurately identifying PDs from healthy controls where multi-modal information were used. The first layer of their stacking ensemble architecture evaluated the advantages of four fundamental classifiers SVMs, RFs, KNNs, and ANNs while their second layer used LRs for categorizations. The performance of their proposed model was compared with standard ensembles.

a strategy that combined CARTs (Classification and Regression Trees) with ensembles. CARTs iteratively chose optimal training speech samples resulting in samples with high separability. Subsequently, these outputs were optimised by ensembles including RFs, SVMs, and ELMs. Their test data were classified using the trained ensembles. The study's recommended strategy was validated with other comparable methods using most recent datasets.

an ensemble-based technique for class label prediction based on voice frequency features for classifying sick and healthy individuals. Their scheme was divided into three stages namely data preparations, internal and final classifications. The findings of the study's tests showed that by using ensembles medical diagnostics were enhanced and provided comparative analyses of many MLTs. Internal classifier results were compared with sample's feature vectors.

a method using MENNs (Multi-Edit-Nearest-Neighbors). Initially, MENNs iteratively picked ideal training speech samples, resulting in samples with excellent separability. Subsequently, DNNEs (Decorrelated Neural Network Ensembles), learnt from acquired samples using ensembles. Finally, the taught ensembles were applied on test data for classifying PDs. The study compared their approach with currently used validation algorithms on latest public datasets.

DNNs as FSAs in an attempt to prove their efficacy by comparing performances of traditional DNNs with other integrated systems. The study developed EOFSCs (Ensembles of Optimal Features and Sample Dependant Base Classifiers) to capitalise on recent discoveries by studies. According to recent researches, distinct optimum models are developed for different forms of speech data that are sensitive to sample variations and subsets of attributes. Using the suggested integrated system, further consolidations of the findings were advised based on the development of EOFSCs. Basic classifiers show sensitiveness towards subsets of characteristics obtained from vowel phonations. This work's suggested EOFSCs use base classifier's for examining characteristics. The final forecasts of EOFSCs were evaluated using majority voting procedure. The results of their experiments suggested that combining FSAs with NNs improved performances of traditional methods. Moreover, integration of FSAs with h DNNs showed superior feature selection integrations with standard MLTs.

3. PROPOSED METHODOLOGY

This work suggest FSAs and classification scheme for approach for identifying PDs in patients. KPCAs (Kernel based Principal Component Analyses) reduce dimensionalities followed by OBEFSs executions which include FMBOAs, LFCSAs, and AFAs. EDL classifiers such as FCBi-LSTMs, CAEs , and SAEs are used in classifications. Finally all classifiers results are combined using Stacked Generalizations and the resultant outputs are evaluated with performance metrics like F-Measures, MCCs, accuracies, and errors. Figure 1 depicts the overall flow of this work's suggested scheme.

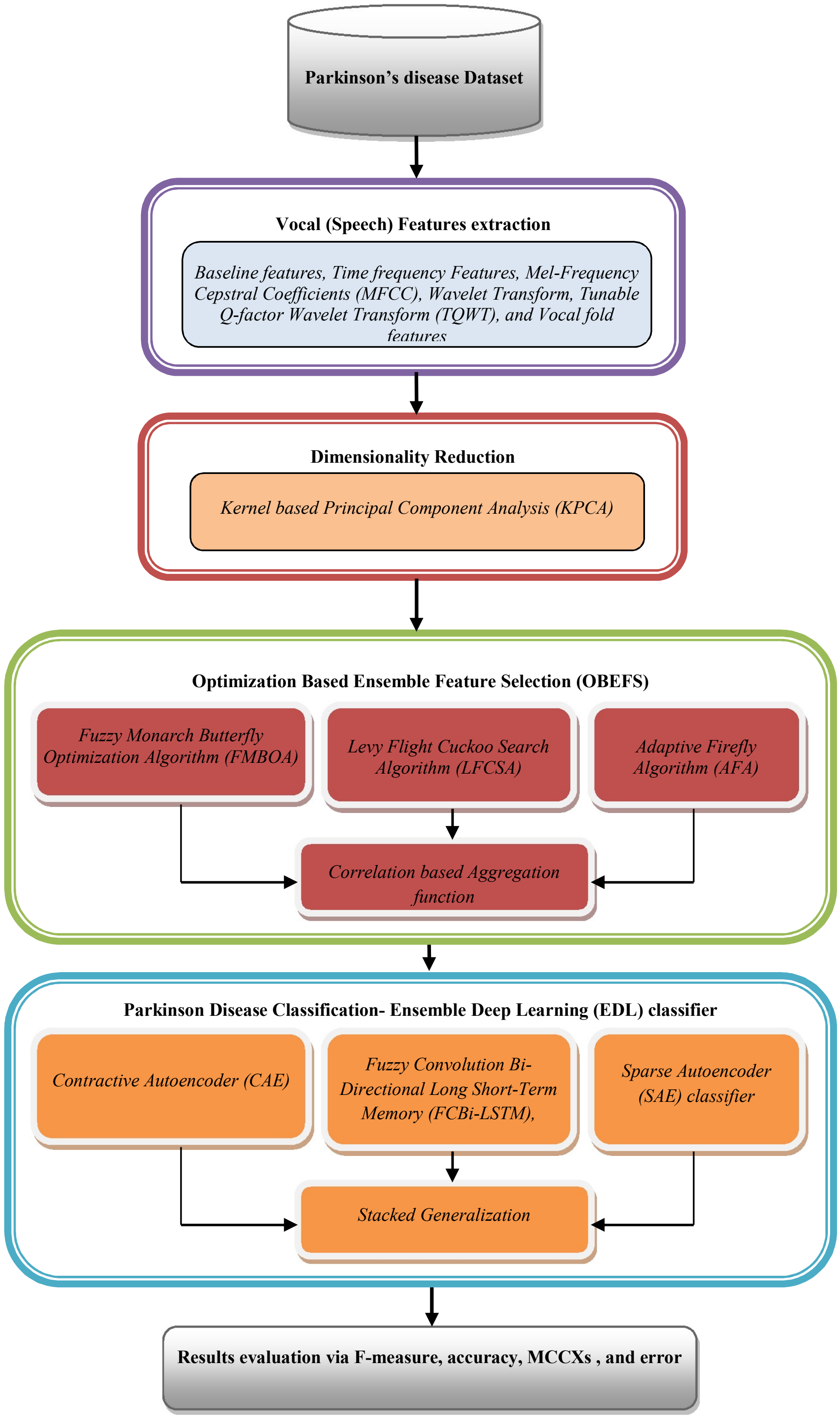

Fig. 1 - Overall Flow Of This Work's Suggested Scheme.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1. Parkinson's disease Dataset

This work used PDs dataset from machine learning repositories of UCI The dataset used in the study was obtained from Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University and includes 188 patients with PDs (107-males, 81-females) and 64 non-diseased individuals (23-Males and 41-Females). The patient's age ranges were The ages of healthy individuals ranged between 41 to 82 years. The voices were recorded in 44.1 KHz and following doctor's inspections, three duplicates of the vowel /a/ letter of persons were obtained.

3.2. Feature extraction

Features including Baseline features, Time frequency Features, MFCCs (Mel-Frequency Cepstral Coefficients), WTs (Wavelet Transforms), TQWTs (Tunable Q-factor WTs), and Vocal fold features were retrieved from the dataset.

i) Baseline characteristics: Patients affected with PDs experience impaired speech even during early stages of the disease's onset and hence characteristics of speech can be effectively utilised to evaluate PDs and track its progression for medical therapies. Examples of baseline features can be: Jitter and glow-based features, fundamental frequency parameters (#5), harmonicity parameters (#2), RPDEs (Recurrence Time Density Entropies) (#1), DFAs (Detrended Fluctuation Analyses) (#1), and PPEs (Pitch Period Entropies) (#1) are commonly used speech characteristics in Parkinson's disease studies

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

iii) MFCCs: MFCCs based extraction methods uses triangular overlapping filter banks to combine cepstral analysis with spectral domain partitions. The data contained 84 MFCCs related parameters generated using mean and standard deviations of the first 13 MFCCs, as well as the signal's log energy's 1st and 2nd [40], apart from the vocal folds, to identify PD effects in the vocal tract (#84).

iv) WTs: WTs are common approaches for making decisions about signals in general, especially ones with minor geographical variances. This approach produces 182 WT-based characteristics for both the approximation and detailed coefficients, including energy, Shannon's and log energy entropy, and Teager-Kaiser energy.

v) TQWTs): TQWTs use three adjustable parameters (Q (Q-factors), r (redundancies), and J (count of levels)) to improve the quality of signal conversion based on signal behaviour Using this dataset, many tests yielded 432 TQWT-related variables

vi) Vocal fold characteristics: Vibrations of voices in Features were examined for noises. The data was used to extract features like GQs (Glottis Quotients) (#3), GNEs (Glottal to Noise Excitations) (#6), VEFRs (Vocal Fold Excitation Ratios) (#7), and EMDs (Empirical Mode Decompositions) features .

vii) Concat characteristics: Concat features combine baseline, vocal fold and time frequency features.

3.3. Reducing Dimensionalities with KPCAs

KPCAs reduce dimensionalities i.e. they reduce linear dimensions of the high dimensional sound waves for examining variances in the case of PDs . Assume that PDs dataset is labelled i,i=1,...N, with each an I being a D-dimensional sound recording characteristics vector. The sound recordings' dimensionality reduction feature vector is found

using this vector. OBEFSs are utilised for feature selection using dimensionality reduced feature vectors.

3.4. Feature selection by OBEFSs

In this work, OBEFSs techniques by aggregating the feature rankings provided by the single feature selectors such as FMBOAs , LFCSAs , and AFAs into a final consensus ranking via correlation.

3.4.1. FMBOAs

FMBOAs are methods for choosing feature sets. FMBOAs select characteristics for samples based on the degree of existence effects. MBOs are based on the migratory behaviors. FMBOAs have been exploited for good results in classifications when employed without modifications. This implies that they manage to balance their searches both global and local. Butterflies relate and distribute information locally within the swarm and thus assisting in enhanced information for systems where the search operations and sharing of information are algorithmically executed using operators for migrations and butterfly adjustments.

3.4.2. LFCSAs

CSAs (Cuckoo Search Algorithms) are based on cuckoo species which use suitable hosts to optimise selection of attributes from datasets in order to nurture their brood. Some cuckoo species employ a suitable host to optimise the selection of features from a dataset in order to nurse their brood. CSAs are used to reduce the risk of egg losses (irrelevant features) to other species without avoiding parental commitments in raising their progeny. The final qualities are determined by putting eggs (features) in several nests. LFCSAs start with a population of N host nests .A cuckoo's egg (feature), say i, is picked at random at each iteration t, and new feature solutions f_i^{t+1} are generated. The Lévy flights are a sort of random walks where steps are characterized in terms of step-lengths and following certain probability distributions. The step's orientations must be isotropic and random. Probabilities prb_a are used to compute replacement rates in a stochastic manner which are tuned for better performances. The best solutions (features) achieved so far are saved as feature vectors f_{best}, and all current solutions (feature selection solutions) are ordered according to their fitnesses (accuracies) at iterations. The procedure is repeated until the specified stopping threshold is reached.

3.4.3. AFAs

The idealised behaviour of firefly flashing is the basis for FAs Fireflies steps need to placed far away from optimal answers as possible. For the optimal selection of attributes from datasets, fireflies are employed and global and local searches are balanced. As a result, firefly's strides also need to consider past and current location's data. Historical data of Fireflies which contain optimal values of past two iterations are taken into account in this work. The distances between current fitness values and population's best fitness values determine further steps of fireflies in iterations. The steps may vary with iterations, and steps of fireflies are also altered in iterations.

3.4.4. Correlation function

Correlation coefficient matrices from features are computed for selecting outputs of ensemble's feature selection and use Equation (1),` vv`

$$Correlation\ Coefficient = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}} \quad (1)$$

Where *x, y* represent examined attribute values while *N* stands for instance counts. Feature sets chosen based on correlations by ensembles become inputs for classifiers.

3.5. classifications of PDs via Ensemble Deep Learning (EDL) classifier

In this work, PDs are classified using an Ensemble Deep Learning (EDL) classifier. Ensemble learning can considerably boost the learning system's generalisation capabilities. EDL, a machine-learning approach that uses numerous classifiers such as FCBi-LSTMs, CAEs, and SAEs to construct an ensemble learner using Stacked Generalization to enable greater generalisation of learning systems, has shown remarkable success in classifications of PDs . Deep EDL combines the benefits of three deep learning models as well as ensemble learning to provide a model with improved generalisation performance. The EDL classifier learns numerous base classifiers and aggregates their results using specified criteria. The rule used to aggregate the outputs defines an ensemble's effective performance..

3.5.1. FCBi-LSTMs

The FCBi-LSTMs are used for classifications based on characteristics chosen from PD datasets . Convolution and pooling layers make up these networks. They carry out convolutions then pooling where outputs are provided as inputs for subsequent Convolution layers. Convolution layers using filters divide features into multiple matrices of sizes (Nm+1) for better representations. These resultant matrices are reduced in dimensions by CNN's pooling layers but manage to maintain links between features. These averaging layers which use preceding Convolved inputs, send their outputs as inputs to subsequent layers. Bi-LSTMs use final Convolution layer's outputs as intermediary variables and hence perform better than uni-directional LSTMs and retain more structural information. The final outputs of Bi-LSTMs are processed by two Convolution layers to obtain diagnosis of PDs. MFB (Multi-modal Factorized Bilinear Pooling) is a technique for combining information from CNNs with Bi-LSTM's obtained features.

3.5.2. SAEs

The SAEs classifier neurons labelled as (+1) are bias units introduced to the FFNNs (feed-forward NNs) through cost functions. This phase drives AEs more correctly to replicate inputs x without over fits .Figure 2 depicts an overview of the model. The sparse auto encoder bottlenecks are utilised as input vectors to DNNs .

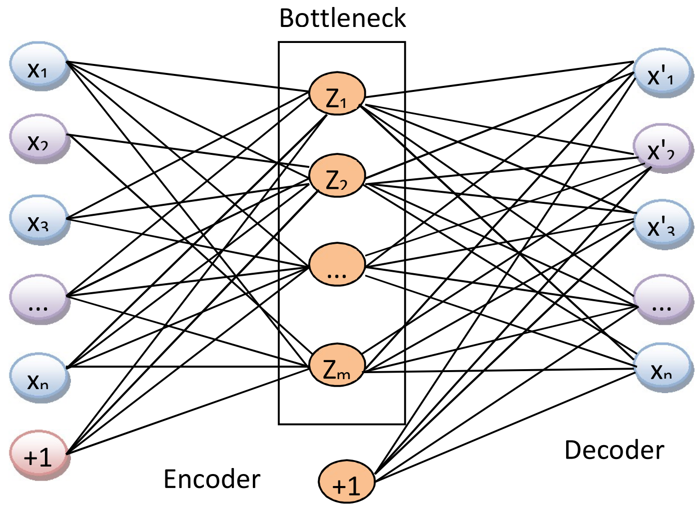

Figure 2. Sparse Model Of Saes

SAEs are proposed in which cost functions are made of 3 parts and are detailed below. Assuming dataset has N training instances $(x_1, x_2, \dots x_n)$, where x_i stands for ith input. SAEs learn to reconstruct inputs x_i using cost functions $h_{W,b}(x_i)$ nearing x_i where MSEs, weight decays, and sparsity are included in the function. The cost functions for N training samples MSEs and weight decays are defined in the following equation [53],

$$J_{sparse}(W,b) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\left\|h_{W,b}(x^i) - x^i\right\| + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(W_{ji}^l)^2 \quad (2)$$

Weight decays specified by the equation above avoids over fits of data, since small values of λ can lead to data over fits, while its larger values can result in data under fits. Sparsity, the third component of the cost functions, helps in activating hidden layers of AEs and prevent data over fits. It has the ability to minimise the number of hidden layer areas evaluated. The average active value of the hidden layer was calculated using the following equation, where a denotes the activation function, which is rectifier (ReLU),

$$\hat{p}_j = \frac{1}{N}\sum_{i=1}^{N}\left(a_j^2(x^i)\right) \quad (3)$$

Sparsity is computed for getting $\hat{p}_j$ nearer to p (sparsity parameter) helps in deviations from p and results in activating or de-activating hidden layer's neurons. It can be defined using Kullback-Leibler divergences as depicted in Equation (4)

$$\sum_{j=1}^{s_l} KL(p\|\hat{p}_j) = \sum_{j=1}^{s_l}\left[p\log\frac{p}{\hat{p}_j} + (1-p)\log\frac{1-p}{1-\hat{p}_j}\right] \quad (4)$$

SAEs cost functions add all three component results which is depicted as Equation (5),

$$J_{sparse}(W,b) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\|h_{W,b}(x^i) - x^i\| + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^l\right)^2 + \beta\sum_{j=1}^{s_l}KL(p\|\hat{p}_j) \quad (5)$$

where β represents sparse penalties. The deep NNs classifier is trained using the bottleneck of the SAEs as inputs, and the SAEs are trained to minimise their cost function as mentioned above. The SAEs and the classifier were both trained at the same time, resulting in improved feature extraction while the classifier's choice was optimised. The training procedure takes 30 iterations, using an 8 instances per batch. The sparsity parameter p was set to 0.05 while weight decay λ was set to 0.0001and sparse penalty term to 2. DNN fine-tuning on the final 10 iterations to change classifier parameters and reduce the softmax cost function while SAE parameters remain fixed. The parameters are updated using the Adam optimizer depending on the calculated gradients.

3.5.3. CAEs

Contractive AEs transform learned representations are robust towards small changes around the training examples It achieves that by using different penalty term imposed to the representation. The loss function (ℓ_2) is used for the reconstruction term. AEs use the encoder function f to map inputs x to internal representations (or codes) f(x). The decoder functions g are then used to map f(x) back to their input spaces. Reconstruction functions r made up of the functions f and g i.e. r(x) = g(f(x)), and reconstruction loss functions l penalises errors, with r(x) considered as forecasts of x. CAEs [54] are regularised AEs that learn to minimise the regularised reconstruction errors and depicted below.

$$\mathcal{L}_{CAE} = \mathbb{E}\left[\ell(\mathrm{x}, \mathrm{r}(\mathrm{x})) + \lambda\left\|\frac{\partial \mathrm{f}(\mathrm{x})}{\partial \mathrm{x}}\right\|_{\mathrm{F}}^2\right] \quad (6)$$

where $r(x) = g(f(x))$ and $\|A\|_F^2$ is the sum of the squares of the elements of A. Both the squared loss,

$$\ell(\mathrm{x}, \mathrm{r}) = \|\mathrm{x} - \mathrm{r}\|^2 \quad (7)$$

and the cross-entropy loss,

$$\ell(\mathrm{x}, \mathrm{r}) = -\mathrm{x}\log \mathrm{r} - (1 - \mathrm{x})\log(1 - \mathrm{r}) \quad (8)$$

Because of the simpler mathematical method it enables, concentrate your investigation on the squared loss. It's worth noting that minimising the CAEs criteria is very dependent on the parameterization of f and g, particularly the linked weights restriction imposed by the equation below (9),

$$\mathrm{f}(\mathrm{x}) = \mathrm{sigmoid}(\mathrm{Wx} + \mathrm{b}) \quad (9)$$

$$g(h) = sigmoid(W^T h + c) \quad (10)$$

Because of the linked weights, the foregoing regularising term compels f (as well as g) to be contractive, i.e., to have singular values < 1. Larger values of λ produce greater contractions (smaller singular values) where it has least impacts on reconstruction errors, i.e. in local directions with little or no variability of data.

3.5.4. Stacked Generalization

Stacked generalisation deduces generalized biases in relation to given learning sets. Cross validation data and least squares with non-negativity constraints were used to determine the best weights of combination in regression to generate a good linear combination of the base learners . Consider the linear combination of the base learners' predictions $f_1, f_2, \ldots, f_m$ given by equation (11),

$$f_{stacking}(x) = \sum_{j=1}^{m} w_j f_j(x) \quad (11)$$

where w is the optimal weight vector learned by the meta learner

4. EXPERIMENTAL RESULTS

Journal of Theoretical and Applied Information Technology
15th May 2023. Vol.101. No 9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This work's experimental findings of the suggested EDL classifier and its comparisons with other approaches including FCLSTM-CNNs (Fuzzy Convolution Long Short-Term Memories) based CNNs, FCBi-LSTMs, and CNNs are detailed in this section. MATLAB R2016a was used to conduct the studies on recognising PDs and classifying them. The following system requirements were followed during implementation: 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.

4.1. EVALUATION METRICS

To analyse prediction performances of classifiers, evaluation metrics are required. Confusion matrix in Table 1 represents the counts of properly and wrongly categorised occurrences for classes based on binary classifications. The symbols tp, fp, fn, and tn in the confusion matrix represent true positive (tp), false positive (fp), false negative (fn), and true negative (tn), respectively. Precisions, recalls, F-measures, accuracies, and error rates were computed using formulae given in Equations (12-16),

$$Precision = \frac{tp}{tp + fp} \quad (12)$$

$$recall = \frac{tp}{tp + fn} \quad (13)$$

$$F - measure = \frac{2 * precision * recall}{precision + recall} \quad (14)$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (15)$$

Error Rate=100-Accuracy (16)

Another statistic for evaluating the quality of binary classifications is the MCCs . MCCs consider tp, fp, fn, and tn counts and are recognised as balanced measures that can be applied even when class distributions are imbalanced. MCCs are correlation coefficients between actual and projected occurrences that range in the interval [-1, +1] where +1 represents flawless predictions and -1 shows differences between forecasts and actual in labels.

Table 1. Confusion Matrix For Two-Class Classification

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

4.2. RESULTS COMPARISON

Experimentations using 4 features and classifiers were evaluated using the metrics of accuracies, errors, F-measures, and MCCs. The combination of MFCC + Wavelet + Concat features with a CNNs classifier yields an accuracy rating of 94.1752 percent. Despite the fact that the accuracy percentage of the MFCC + Wavelet + Concat combination is 95.1557 percent for the FCLSTM-C FCBi-LSTM classifier gives the accuracy results of 96.6381%, 98.0244%, 97.3457% and 98.7720% respectively for TQWT+MFCC+Wavelet, TQWT+Wavelet+Concat, TQWT+MFCC+Concat, and MFCC+Wavelet+Concat. Proposed EDL classifier with MFCC + Wavelet + Concat combination achieves the highest performance with an accuracy rate of 99.903%, F-Measure rate of 99.633%, and 72.431% for MCCs (See Table 2).

Table 2. Results Of Classifiers With Triple Feature (Kpca+ Obefs)

CNNs CLASSIFIER (%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR RATE**	**MCCs**
F1-TQWT+MFCC+Wavelet	85.8697	87.5696	12.4304	57.2007
F2-TQWT+MFCC+Concat	90.2315	91.9315	8.0684	61.4600
F3-TQWT+ Wavelet + Concat	86.3695	88.0694	11.9306	63.3994
F4-MFCC + Wavelet + Concat	92.4752	94.1752	5.8248	64.5384
FCLSTM-CNNs CLASSIFIER (%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR RATE**	**MCCs**

F1-TQWT+MFCC+Wavelet	94.2258	93.0470	6.9530	67.6669
F2-TQWT+MFCC+Concat	91.5250	93.0854	6.9146	67.7060
F3-TQWT+ Wavelet + Concat	93.4200	93.1261	6.8739	65.4457
F4-MFCC + Wavelet + Concat	91.6921	95.1557	4.8443	67.2960
FCBi-LSTM CLASSIFIER(%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR RATE**	**MCCs**
F1-TQWT+MFCC+Wavelet	98.3100	96.6381	3.3619	74.300
F2-TQWT+MFCC+Concat	96.5900	98.0244	1.9756	72.300
F3-TQWT+ Wavelet + Concat	97.5200	97.3457	2.6543	70.300
F4-MFCC + Wavelet + Concat	98.5010	98.7720	1.2280	71.400
EDL CLASSIFIER (%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR RATE**	**MCCs**
F1-TQWT+MFCC+Wavelet	99.449	97.771	2.2299	75.432
F2-TQWT+MFCC+Concat	97.722	99.156	0.8436	73.412
F3-TQWT+ Wavelet + Concat	98.652	98.378	1.6222	71.013
F4-MFCC + Wavelet + Concat	99.633	99.903	0.0966	72.431

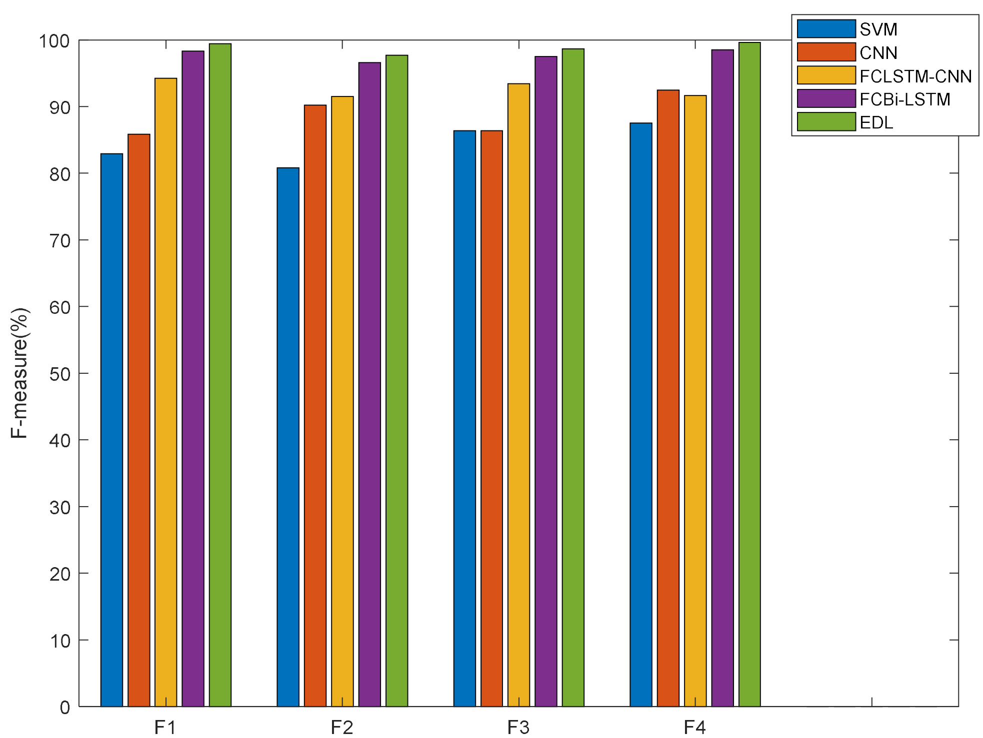

Figure 3. F-Measure Comparison Of Feature Level Combination Vs. Classifiers

Figures 3-6 shows F-measures, accuracies, MCCs , and errors using different feature sets. When compared to other feature combination sets, TQWT+MFCC+Wavelet combination of feature set yields greater results of 99.449 percent, 97.77 percent, 75.432 percent, and 2.2299 percent for f-measure, accuracy, MCCs , and error. The suggested classifier using the first feature level combination achieves a higher f-measure of 99.449 percent, whereas other approaches such as CNNs , FCLSTM-CNNs , and FCBi-LSTM achieve f-measures of 85.8697 percent, 94.2258 percent, and 98.3100 percent, respectively.

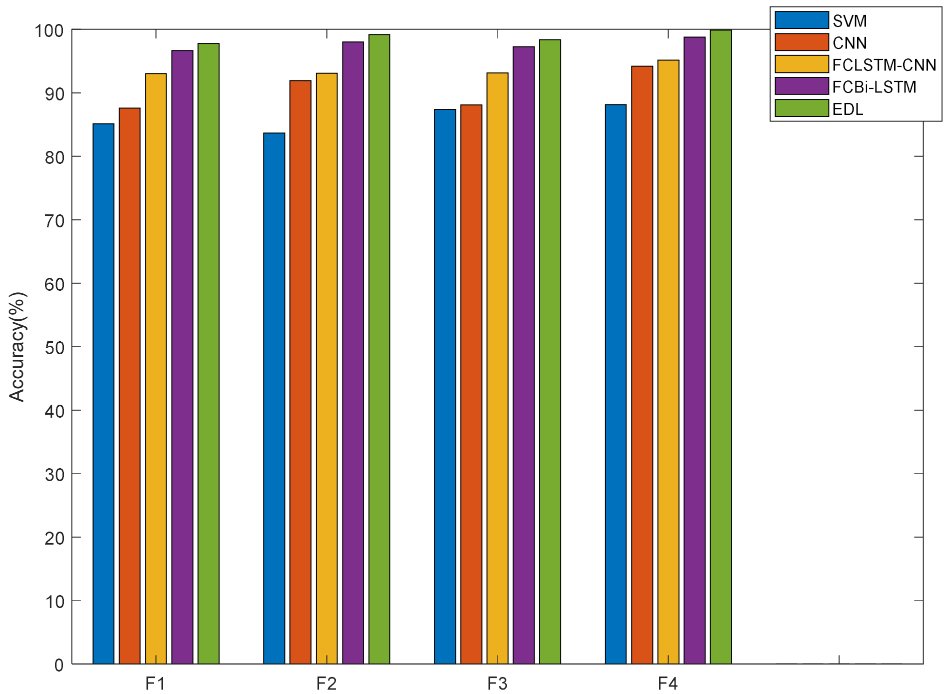

Figure 4. Accuracy Comparison Of Feature Level Combination Vs. Classifiers

Figure 4 shows that the x-axis results are measured using a feature-level combination of classifiers, and the y-axis results are shown as accuracy. The final feature level combination proposed for the classifier achieves 99.9030 percent accuracy, whereas other approaches such as CNNs , FCLSTM-CNNs , and FCBi-LSTM achieve 94.1752 percent, 95.1557 percent, and 98.7720 percent accuracy, respectively.

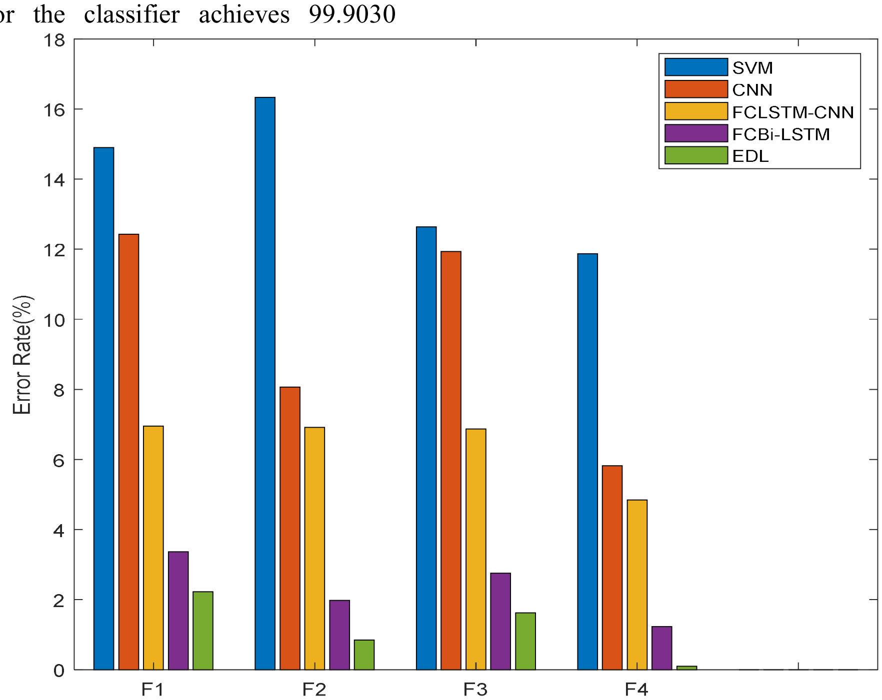

Figure 5. Error Rate Comparison Of Feature Level Combination Vs. Classifiers

Figure 5 compares the error rates of classifiers with four different feature level combinations. Figure 5 shows that the proposed classifier with the final feature level combination has a lower error rate of 0.0966 percent, whereas CNNs , FCLSTM-CNNs , and FCBi-LSTM approaches have larger error rates of 5.8248 percent, 4.8443 percent, and 1.2280 percent, respectively.

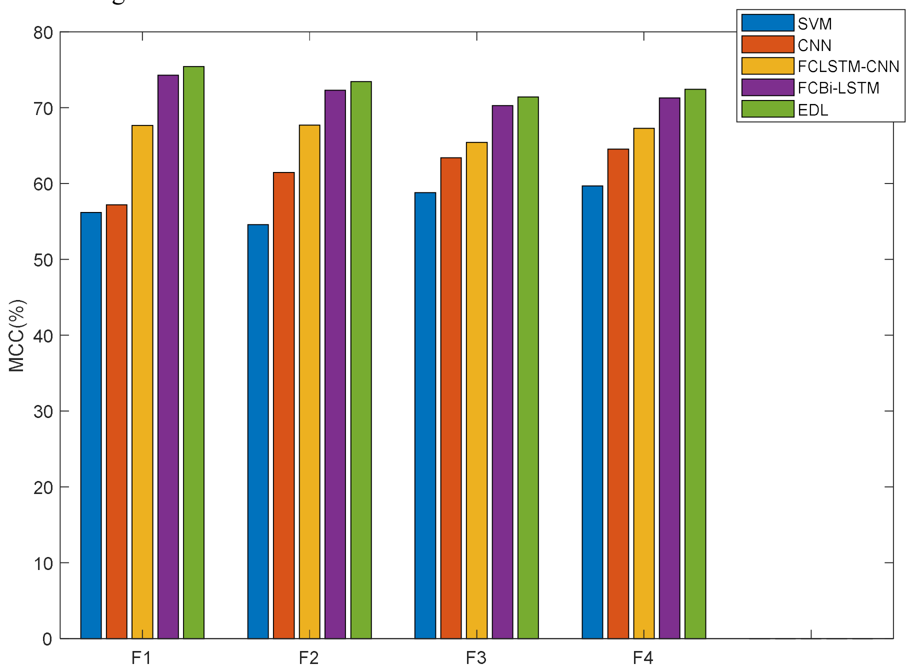

Figure 6. Mccs Comparison Of Feature Level Combination Vs. Classifiers

Figure 6 displays MCCs results comparison of first feature level combination produces greater results of 75.432 percent for proposed algorithm, 57.2007 percent , 67.6669 percent , and 74.300 percent for CNNs , FCLSTM-CNNs , and FCBi-LSTM approaches classifiers accordingly.

5. CONCLUSION AND FUTURE WORK

In this study, an ensemble learner known as an OBEFSs is proposed along with Ensemble Deep Learning (EDL) classifier that combines single approaches. This is done using algorithms such as FMBOAs, LFCSAs, and AFAs. To carry out OBEFS, the correlation function is used to choose the best features from the three feature subsets. After that, the EDL classifier is utilised to get a diagnostic of patients with PDs. The EDL classifier contained FCBi-LSTMs , CAEs , and SAEs . The stacked generalisation method was developed to aggregate the findings of classifiers that only choose the majority class (PDs) of the dataset. F-Measures, Accuracies, MCCs , and error rates were used to evaluate the categorization The classifier outputs are implemented using machine learning repositories of UCI. Experimentation is carried out with four different types of feature sets and classifiers. The proposed EDL classifier using MFCC+Wavelet+Concat combination obtains the greatest accuracy rate of 99.9030 percent, F-Measure rate of 99.633 percent, and MCCs rate of 72.431 percent. Most PDs research rely only on accuracy rates, which might be deceptive in cases of skewed class distribution. Instead, various statistical techniques like as cross-validation and the ANOVA test have been included to examine the classifier outcomes.

DECLARATION:

Ethics Approval and Consent to Participate:

No participation of humans takes place in this implementation process

Human and Animal Rights:

No violation of Human and Animal Rights is involved.

Funding:

No funding is involved in this work.

Data availability statement:

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study

Conflict of Interest:

Conflict of Interest is not applicable in this work.

Authorship contributions:

All authors are contributed equally to this work

Acknowledgement:

There is no acknowledgement involved in this work.

REFERENCES

[1]. H.Gao, X. Wei, J. Liao, R. Wang and J. Xu, "Lower bone mineral density in patients with Parkinson's disease: a cross-sectional study from Chinese Mainland," *Frontiers in aging neuroscience*, vol.7,2015, pp.1-9. https://doi.org/10.3389/fnagi.2015.00203 .

[2]. S.Halbgebauer, M. Nagl, H. Klafki, U.Haußmann, P. Steinacker, " Modified serpinA1 as risk marker for Parkinson's disease dementia: Analysis of baseline data,"*Scientific reports*, vol.6,2016, pp.1-8. DOI: 10.1038/srep26145

[3]. X.Wei, H. Gao, J. Zou, X. Liu, D. Chen , » Contra-directional coupling of Nur77 and Nurr1 in neurodegeneration: a novel mechanism for memantine-induced anti-inflammation and anti-mitochondrial impairment,"*Molecular neurobiology*, vol.53,2016, pp.5876-5892. DOI: 10.1007/s12035-015-9477-7

[4]. P.Ashok Babu, L. Kavisankar, J. Xavier, V. Senthilkumar, G. Kumar, T. Kavitha, A. Rajendran, G. Harikrishnan, A. Rajaram and A.G.Adigo, "Selfish Node Detection for Effective Data Transmission Using Modified Incentive Sorted Pathway Selection in Wireless Sensor Networks,"*Wireless Communications and Mobile Computing*, *2022*. https://doi.org/10.1155/2022/9359135.

[5]. J.F.Guo, K. Li, R.L. Yu, Q.Y. Sun, L. Wang, L.Y. Yao, " Polygenic determinants of Parkinson's disease in a Chinese population," *Neurobiology of aging*, vol.36,2015, pp.1765-e1. DOI:10.1016/j.neurobiolaging.2014.12.030

[6]. M.S.Bryant, D.H. Rintala, J.G. Hou and E.J.Protas, " Relationship of falls and fear of falling to activity limitations and physical inactivity in Parkinson's disease,"*Journal of aging and physical activity*,vol. 23,2015, pp.187-193. DOI:10.1123/japa.23.2.187

[7]. A.Rajaram and K.Vinothkumar, " Efficient Multipath Location Aware Routing Protocol for Mobile Ad Hoc Networks," *Journal of Theoretical and Applied Information Technology, Scopus Indexed*, Vol.*59,2019,* pp.130-38.

[8]. M.A.Nalls, N. Pankratz, C.M. Lill, C.B. Do, " Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease,"*Nature genetics*, vol.46,2014, pp.989-993. DOI:10.1038/ng.3043

[9]. H.Gürüler H., "A novel diagnosis system for parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method," *Neural Computing and Applications*, vol. 28, 2017, pp. 1657–1666. DOI:10.1007/s00521-015-2142-2

[10]. A.Rajaram, and A.Baskar, "Hybrid Optimization-Based Multi-Path Routing for Dynamic Cluster-Based MANET," *Cybernetics and Systems*.2023. https://doi.org/10.1080/01969722.2023.2166249

[11]. A.Rajaram and J.Sugesh, "Power aware routing for MANET using on-demand multipath routing protocol," *International Journal of Computer Science Issues (IJCSI)*, vol.*8, 2011,pp*.517. DOI:10.1109/ICDCSW.2004.1284112

[12]. M.Peker, " A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM," *Journal of medical systems*, 40,2016, pp.1-16. DOI:10.1007/s10916-016-0477-6

[13]. B.E.Sakar G. Serbes, and C. O. Sakar, "Analyzing the effectiveness of vocal features in early telediagnosis of parkinson's disease*," PloS one*, vol. 12, 2017, pp.1-18. DOI:10.1371/journal.pone.0182428.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[14]. K.J.Kubota , J. A. Chen and M. A. Little, "Machine learning for largescale wearable sensor data in parkinson's disease: Concepts, promises, pitfalls, and futures," *Movement disorders*, vol. 31, 2016, pp. 1314–1326. DOI:10.1002/mds.26693.

[15]. D.N.V.S.L.S. Indira, R.K., Ganiya, P. Ashok Babu, A. Xavier, L. Kavisankar, S. Hemalatha, V. Senthilkumar, T. Kavitha, A. Rajaram, K. Annam and A.Yeshitla, " Improved artificial neural network with state order dataset estimation for brain cancer cell diagnosis." *BioMed Research International*, *2022*. https://doi.org/10.1155/2022/7799812