[1]P Pradeep

[2]Dr Kamalakannan J

Design of a Predictor Model for Feature Selection using Machine Learning Approaches

Journal of Electrical Systems

Abstract: - Parkinson's disease is a neural degenerative disease where patients' faces various critical neurological disorders. Thus, the earlier prediction of PD helps to enhance the patients' life. The prediction of PD in earlier stage is complex and it consumes huge time. Therefore, effectual and appropriate prediction of PD is measured to a challenging factor for the health care experts and practitioners. To deal with this issue and to accurately predict the PD in earlier stage, this work concentrates on machine learning approaches for designing a predictor system. For developing the anticipated model, L1-norm based Genetic algorithm (L1-GA) is applied for predicting PD in the earlier stage. This L1-GA is utilized for selecting the influencing features for accurate prediction. This L1-GA produces newer feature subset from UCI Machine Learning (ML) dataset for PD for measuring feature weights. For validation , this work considers k-fold cross validation (CV) is used. Also, metrics like accuracy, error rate and execution time are evaluated. The inputs are taken from The PD dataset which is available online for preceding the feature selection process. The optimal accuracy attained with these newly selected sub-sets are considered for further computation. The simulation is performed in Python environment and the experimental findings determine that this study recommends that L1-GA provides better contribution towards PD feature selection and to predict PD in earlier stage. In recent times, Clinical Decision Support System (CDSS) plays an essential role for assisting PD recognition. As well, the anticipated model lays a bridge to fill the gap encountered in feature selection using the available data. The anticipated model gives better trade-off in contrast to prevailing approaches.

Keywords: Parkinson's disease, machine learning, L1-norm based Genetic algorithm, cross validation, feature selection

1. Introduction

Parkinson's disease (PD) is a neuro-degenerative, long-term, and age dependent disease which affects central nervous system and influences motor system of the individuals. Specifically, the inference and frequency of PD over 1000 people is estimated with the 1-2 persons [1]. The occurrence of PD over the individual is assumed to be increases with age. Approximately, 1% of human population is suffering with PD for more than 60 years old [2]. There are certain genetic disorder factors that are identified with 5-10% of disease affected individuals with some environmental affected cause. Some common symptoms of PD include walking, shaking movement, rigidity, and some issues over cognitive behavior and thinking [3]. With the constant disease progression, it turns to be another condition known as dementia. When an individual is influenced by PD possess risk of two-six times towards dementia when compared to other population [4].

As well, depression and anxiety are some common issues related to the people under this condition [5]. Similarly, gender factor is also another factor to be considered, as the chances of PD over female is lesser than male, i.e., with the ratio of 3:2. Although there is no complete cure towards this disease, certain treatment process provides an improvement with this condition [6]. Anti-Parkinson Medication Levodopa merged with Dopamine Agonists (L-DOPA) proves that there are some positive responses towards the treatment of PD patients [7]. While investigating with the drug resistance over numerous cases, an effectual surgical technique termed as brain stimulation is carried out to diminish motor symptoms [8]. Also, PD causes neuropsychiatric disturbances ranges from gentle to severe [9]. It is classified as behavior, cognition, mood, and thought. It also impairs drowsiness, sleep disorders, behavior disorder, REM and so on. The prediction of PD is made easier in numerous ways. Examination of medical history and neurological analysis are carried out by physicians for preliminary evaluation of condition. Imaging technologies like Computer Tomography (CT) and Magnetic Resonance Imaging (MRI) are generally adopted for predicting PD [10]. MRI is comparatively efficient than CT for appropriate disease

[1] Research Scholar, VIT University, Vellore.India

pradeep.p@vit.ac.in

[2]Sr Associate Professor SCORE, VIT University, Vellore.India

jkamalakannan@vit.ac.in

prediction. A diffusion technique of MRI is thoughtful for differentiating PD over the added symptoms/syndrome of PD.

With the arrival for more sophisticated computational diagnosis system, various diseases are analyzed and adopt the intelligent predictor model for enhancing the efficiency [11]. Domestic and industrial applications collect huge data over the days for effectual analysis. Some other applications like earth simulator, protein folding models, flight simulators, weather forecasting, and so are adopts the intelligent predictor model [12]. The cognitive ability over the learning model is acquired via various computer science fields, specifically from Artificial Intelligence (AI), mathematics, Machine Learning (ML), computer vision and cognitive psychology respectively. In recent times, mobile phone applications are modeled to predict the causes of Parkinson's disease [13]. It attempts to identify PD and the progression of the disease constantly. Scoring criteria is also followed for predicting the motor scores over certain period.

Also, PD detection approaches attempt the essential symptoms that recognize the use of diverse medical devices and equipment. The notable signs of the PD occurrence are the vocal pattern of the affected individuals. This defect is identified in earlier disease progression. Therefore, vocal disorder significance possess stronger link towards PD and it is extremely essential for modeling various computational approaches for diagnosing the condition effectually. In various studies, features vectors are hauled out with assistance of diverse speech signal technique. The extracted features are fed as an input to the intelligent disease predictor model to identify the hidden patterns from the provided data [14]. These predictor models pretend to determine discriminative factors to appropriately categorize samples of diverse groups. ML based prediction approach acquires enormous fame these days, as prediction with this model is reliable and accurate. Learning process complexity is extremely high when compared to previous model; however it is best suited for diverse complex applications [15]. Generally, most ML approaches outperforms when training samples are enormously higher. In some cases, certain unstructured data is extremely suited for ML approaches.

The significance of the automated diagnosis system for predicting PD relies over the severity. In case, if the disease is diagnosed, it turns to be a life-threatening factor. Similarly, the prior diagnosis also drastically enhances the affected condition of every individual rapidly. This investigation attempts to determine the discriminative patterns among the PD control and affected cases from vocal features over dataset with operative learning models. Here, an effectual processing channel is modeled for identifying the most influencing features for better classification. Initially, PD based datasets are attained from the online available UCI repositories. It is pre-processed for eliminating missing values and noise. Subsequently, the candidate/influencing features are predicted with L1-norm based Genetic Algorithm (L1-GA). The L1-norm is merged with the efficient GA for enhancing the prediction performance. It hauls out latent representation of influencing features and given to the algorithms. K-fold CV is applied for evaluation and training. The model performance is computed with various validation metrics. The outcomes rely over optimal significance of PD from the anticipated framework. Moreover, it enhances the significance of diverse computational model for healthcare sectors for superior disease prediction.

The work is structurally expressed as: Section 2 depicts the brief discussion regarding the prevailing methodologies. Section 3 includes the elaboration of dataset, L1 norm, and GA for extracting the latent features to build a better predictor model. The outcomes are analyzed and discussed in section 4. Section 5 summarizes anticipated framework and highlights significance of model with future research direction.

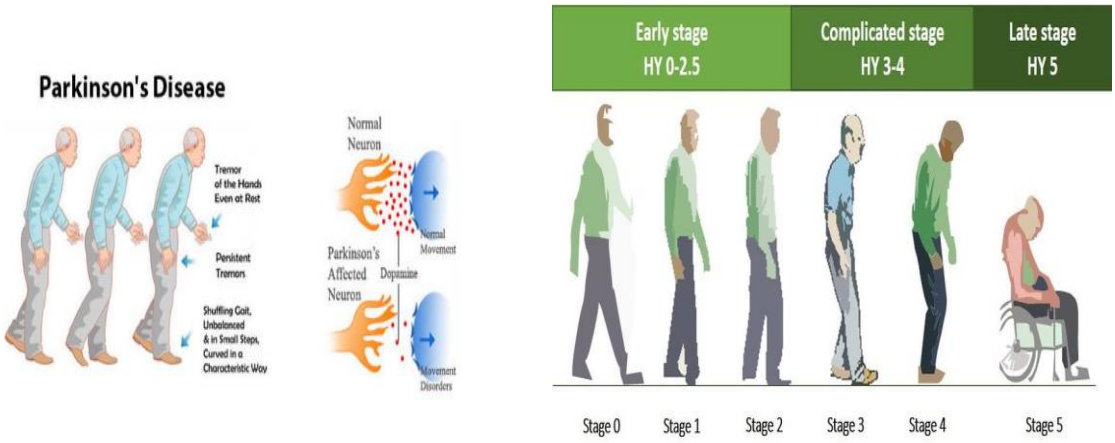

Fig 1: Generic view of PK and its stages

2. Related works

This section demonstrates the prevailing approaches over Parkinson's disease prediction with various available dataset. It is explained by various investigators and these researchers pretend to attain better outcomes for assisting clinical healthcare systems.

In the past few decades, ML-based disease categorization is extensively applied in diverse medical applications and attained notable outcomes. W. Dickson et al., in [16], considered Gaussian based density model from five diverse PD corpora. The author utilized phonetic text-based features that needs vocal tract feature of diverse speech signals. The researchers evaluated vowels, sentences, words, and monologues in corpora of male patients. The attained results are superior while comparing it with other datasets and it provides 82% accuracy. Similarly, I. Obeso et al., in [17], applied wrapper based selection approach for

words and vowels by recordings and attained an accuracy of 71%. E. Adeli et al., in [18] used traditional features like BoW and frequencies of monologues. He used various recordings and states language pronunciation comprises of appropriate information for categorizing and attaining better accuracy. A. Abos et al., in [19] evaluates decomposition and inherent based features from vowels from two diverse dataset. The author reports 96% accuracy with Spanish dataset by incorporating SVM and RF.

Liu et al., in [20], demonstrates the adoption of biomarkers for feature articulation from vowels and sentences of Spanish speech spectrograms which is effectual for predicting Parkinson's disease. The author also considered intelligibility and articulation features from words such as Spanish Gita recordings. He also reports 89% accuracy while training the classifier with articulation and intelligibility based features.

Adeli et al., in [21] anticipated novel method by determining on/off state of vocal folds (candidate initiates speaking and off while stop speaking). It comprises of vowels, monologues, words, and sentences from various recordings. The author reports 95% accuracy during speech signals for classifying PD. The author evaluated the suitability of vectors during PD classification with phonation, prosody, and articulation of vowels, words and sentences from available dataset. The author evaluated cosine variations from disease detection by contributing 79% accuracy with articulation features. Lie et al., in [22], modeled an approach that considers voice signal during phone calls and considers syllable. The investigator measures the onset and severity of PD in various works.

Gillies et al., in [23] evaluated the aging parameters in his work. The author considers prosody, articulation features along with age factors from speech vowels. Also, the author accounts for gender parameters and evaluated the age factor which plays a crucial role in PD classification. The investigators measured that some speaker signals are contributing PD classification procedure than old age speaker signals. The investigator modeled multi-class and binary SVM and evaluated the outcomes with NN by showing 96% accuracy. Chen et al., in [24] provided a novel method for categorizing speech signal towards PD/health patients. The author anticipated a phonological feature based model with monologues, speech signal of words, and text. The investigator evaluates this model with essential features of PD patients over clinical data. The author utilized conventional ML approaches while

considering phonological and articulation-based features for PDs detection. The author evaluated kinetic features for sustained and two-read sentences from available dataset from PD speech signals. The author carried out classification tasks with vectors and GMM and reports 89% accuracy.

Liu et al., in [25] anticipated open source software for PD. The author utilized articulation, phonation, prosody, and intelligibility speech signal dimensions from various dataset for vowels with traditional ML approaches to predict PD. The author modeled a system that is easily used by diverse physicians to evaluate various voice based diseases. The author modeled a framework for evaluation of PD with individual speech signal analysis. The author evaluated articulation, phonation, and prosody to design constant speech and read text recording from various channels (online class, phone and skype calls). Here, the author observed speech signals from skype were effectual in distant observation of PD patients. The author carried out the evaluation with GMM and vectors by attaining 77% correlation respectively.

O.B. Tysnes et al., in [26] evaluated an enhanced version of FDA for predicting PD. He considers prosody, articulation, phonation, and intelligibility features with Spanish vowels, words and sentences. The author considers vowel recordings with online available datasets. He integrates statistical pooling for improving the features and applies ReliefF based feature selection and attained 92% accuracy with SVM. Kim et al., in [27] applies gait physionet dataset for diagnosing PD. The author evaluated PD with gait using deep NN. He achieved 98% accuracy with DNN and 85% for predicting the PD severity. Brooks et al., in [28], anticipated an octopus based multiple pooling approaches (eight different pooling method) for extracting features. He evaluates vowels dataset and acquired superior accuracy with SVM. Arora et al., in [29] offered a technique for appropriate PD prediction with various dataset comprising of vowel pronunciation in various languages (UCI data). Polat et al., assessed phonetic and acoustic speech characteristics [30]. The author attained 99% accuracy with RF.

3. Methodology

This section discusses about the anticipated method used for extracting features from PD dataset for enhancing the classification accuracy. Here, L1-norm is used for performing Latent representation of features and the effectual feature extraction analysis is done with Genetic Algorithm (GA). The extracted features are fed as input to the classifier model. The block diagram of anticipated model is given in Fig 2. Similarly, flow of the anticipated model is expressed in Fig 3.

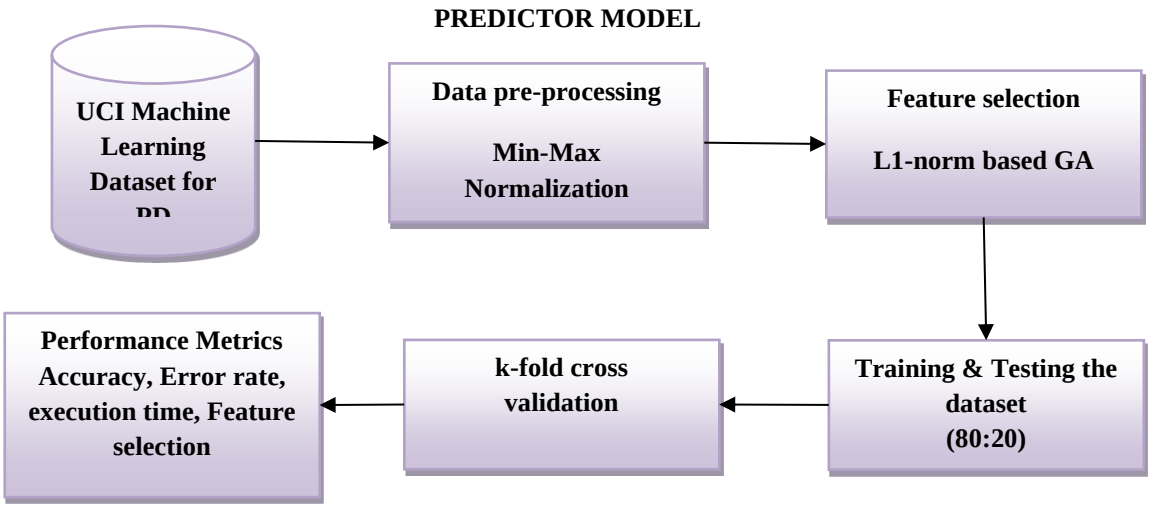

Fig 2: Flow diagram of proposed model

a. Dataset description

The data which is used in this investigation is gathered from 188 patients (81 women and 107 men) with PD. The age of the patients ranges from 33 to 87 from the department of Neurology in Istanbul University. Similarly, the control groups comprises of 64 healthy individuals (41 women and 23 men). During the process of data collection,

the microphone value is set as 44.1 KHz. It is based on physician's examination with phonation of vowels collected from every individual with three repetitions. Various speech signal processing algorithms like Wavelet Transform, Time Frequency Features, Mel Frequency Cepstral Coefficients (MFCCs), TWQT and vocal fold features are used for speech recordings of PD patients to extract essential information clinically from PD evaluation. Table I depicts the description of the PD dataset.

Table I: dataset descriptions

Characteristics	Multi-variate
Attributes characteristics	Real, Integer
Associated tasks	Classification
Total instances	756
Total attributes	754
Missing values	N/A

The ultimate objective of using this dataset is to categorize individuals with PD from healthy from the affected individuals by measuring the differences among the vowel vocalization. The dataset size is 195*23 matrixes. Here, 0 is set for healthy persons and 1 for PD patients. For an individual, average vowel phonation recorded is 6 over 36 seconds with a total of 195 samples.

b. Pre-processing

The appropriate representation of data is extremely an essential step before performing classification with ML approaches. The pre-processing technique includes standard scalar, removing of missing values, Min-Max scalar are applied to dataset. It has to fulfill the standard scalar feature with a mean of 0 and variance 1. Similarly, the data ranges from 0 to 1. Here, Min-Max normalization is provided as in Eq. (1):

$$V^{-} = \frac{v - min}{max - min}\,(new_{max} - new_{min}) + new_{min} \qquad (1)$$

Here, $'V'$ is old feature and V^{-} is newer feature.

Feature selection with L1-norm

This step is essential for eliminating irrelevant features from feature space. The redundant features can enhance accuracy and diminishes execution time during selection process. L1-norm with GA is used for feature selection. L1-norm is applied for latent representation during feature selection. Assume the given dataset with $'n'$ samples which is expressed as in Eq. (2):

$$S = \left\{(x_i, y_i)\middle| X_i\ \varepsilon\ R^n, y_i\ \varepsilon\ \{-1,1\}\right\}_{i=1}^{k} \qquad (2)$$

Here, x_{ij} is j^{th}feature value with $'i'^{th}$ sample and class label y_i where x_i expressed as in Eq. (3):

$$X_i = \{X_{i1}, X_{i2}, \dots, X_{in}\} \qquad (3)$$

Here, x_{ij} is j^{th} feature values with x_i samples. When $'X'$ specifies $n * p$ matrix where rows are expressed as $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$; $1 \le i \le n$, where $'Y'$ specifies dependent variables $(y_1, y_2, \dots, y_n)^T$. For control parameters λ $(\lambda > 0)$, which is a common regularization form expressed as in Eq. (4):

$$L\ (\lambda, , \beta) = \arg \min_{\beta} \{R(\beta) + \lambda\ P(\beta)\} \qquad (4)$$

Here, $\beta\ \epsilon\ R^P$ as estimated coefficients, $R(\beta)$ is loss function, $P(\beta)$ is regularization term. It is a most commonly used regularization approach with selection operators and absolute shrinkage, i.e., $P(\beta) = \sum_{j=1}^{p} |\beta_j|^1$. it carries out constant shrinkage and gene selection process at same time. There are some other L1 norm based regularization method is also anticipated.

The existing feature selection approaches show inconsistency during feature evaluation, feature interactions and large scale instances and features. Here, L1-norm based regularization is integrated with GA to perform effectual selection process with latent representation. This work anticipates a new chromosome specification comprising penalized control parameters and feature coefficients in learning model. In the initial process, GA population is randomly initialized with every chromosome encoded with feature coefficients and penalized control parameters to reach local optimal solutions or to enhance individual's fitness in search population. The operators like mutations, and crossovers are performed with penalized control parameters and selection operator's produces successive population. This process is repeated until the stopping condition is fulfilled. This process is explained as below. The representation is provided by two penalized parameters λ, α and coefficient of feature subset $(\beta_1, \beta_2, .., \beta_p)$ which is encoded as penalized control parameters and feature coefficients, $(\alpha, \lambda, \beta_1, \beta_2, \dots, \beta_p)$. The chromosome length is specified as $p + 2$ where $'p'$ is total number of features. Chromosome is a string (real value) and penalized control parameters are optimized globally with the operators. Even though the search space is multi-modal and non-convex, GA possesses global optimal ability as the dimension is extremely lower. Subsequently, the feature coefficient is optimized using regularization process for synchronous feature selection and learning model construction. In feature coefficient part, the non-zero β_i value with corresponding features are eliminated and the appropriate coefficients are equal to zero. The maximal available number of non-zero feature of available chromosome is specified as $'T'$. With number of optimally available features, the limit of $'T'$ is pre-defined values; else it is equal to $'p'$.

c. Objective function

The objective function is expressed as in Eq. (5):

$$Fitness\ (chromosome) = accuracy\ (\lambda, \alpha, \beta_1, \beta_2, \dots, \beta_p) \qquad (5)$$

Here, β_i non-zero specifies appropriate feature subset encoded with learning feature coefficient of chromosome. The objective function computes the significance of feature subset. Here, objective function fitness is evaluated as accuracy of regularization model with chromosome $\lambda, \alpha, \beta_1, \beta_2, .., \beta_p$ using L1-norm penalty method. When two chromosome are considered to possess same fitness, i.e., difference among the fitness is lower than smaller $'e'$ value (e^{-5}); the feature with smaller value/number is provided as higher chances to next generation.

d. Genetic operators with regularization process

In evolution process with regularization process, some standard GA operators like crossover, fitness selection (proportionate), and uniform mutation operators are used. However, with prior knowledge on optimal amount of features, the number of β_i based non-zero values of chromosome as constrained to maximal $'T'$ over evolution process.

Parameter	Value
Population size (P)	200
Crossover probability	0.90
Mutation probability	0.1
Stopping criteria	2000

e. Crossover

Initially, choose the parents randomly (ma, pa) from the available population for further breeding process. Subsequently, crossover operation is applied to cross over population probability ($p_c = 0.85$ respectively for producing offspring with inherent characteristics of the parents. Single crossover on feature coefficient of ma and pa chromosomes is produced among penalized control parameters $'\alpha'$ and $'\lambda'$; then these penalized control parameters are swapped among the parents to generate control parameters of offspring $'c_1'$ and $'c_2'$ respectively. The feature coefficients $'\beta'$ of these offspring chromosome are measured with local optimization.

Algorithm 1
1. Begin
2. Pre-processing data with Min-Max normalization as in Eq. (1)
3. Chose the most influencing feature with the L1-norm based on GA
4. Initialize population size, cross over probability, mutation probability, and stopping criteria
4. Perform k-fold CV with available training and testing set
5. Train classifier with (P, G, pc, pm)
6. validate selection with test sets and achieve best feature selection parameters
7. Evaluate objective function using Eq. (4)
8. Use genetic operators for regularization
9. Analyze mutation probability with $p_m = 0.1$
10. Generate next selection process with available offspring and parents using Eq. (5)
11. Measure the performance of prediction model and testing set with candidate process
12. end the process

f. Mutation

This operator facilitates population diversity and larger exploration of search space. In this step, the randomly chosen penalized control parameters $'\lambda'$ and $'\alpha'$ with mutation probability ($p_m = 0.1$) are mutating to chosen chromosome. The fitness value and feature coefficients of newer chromosome produced by mutation operation and measured by local optimization.

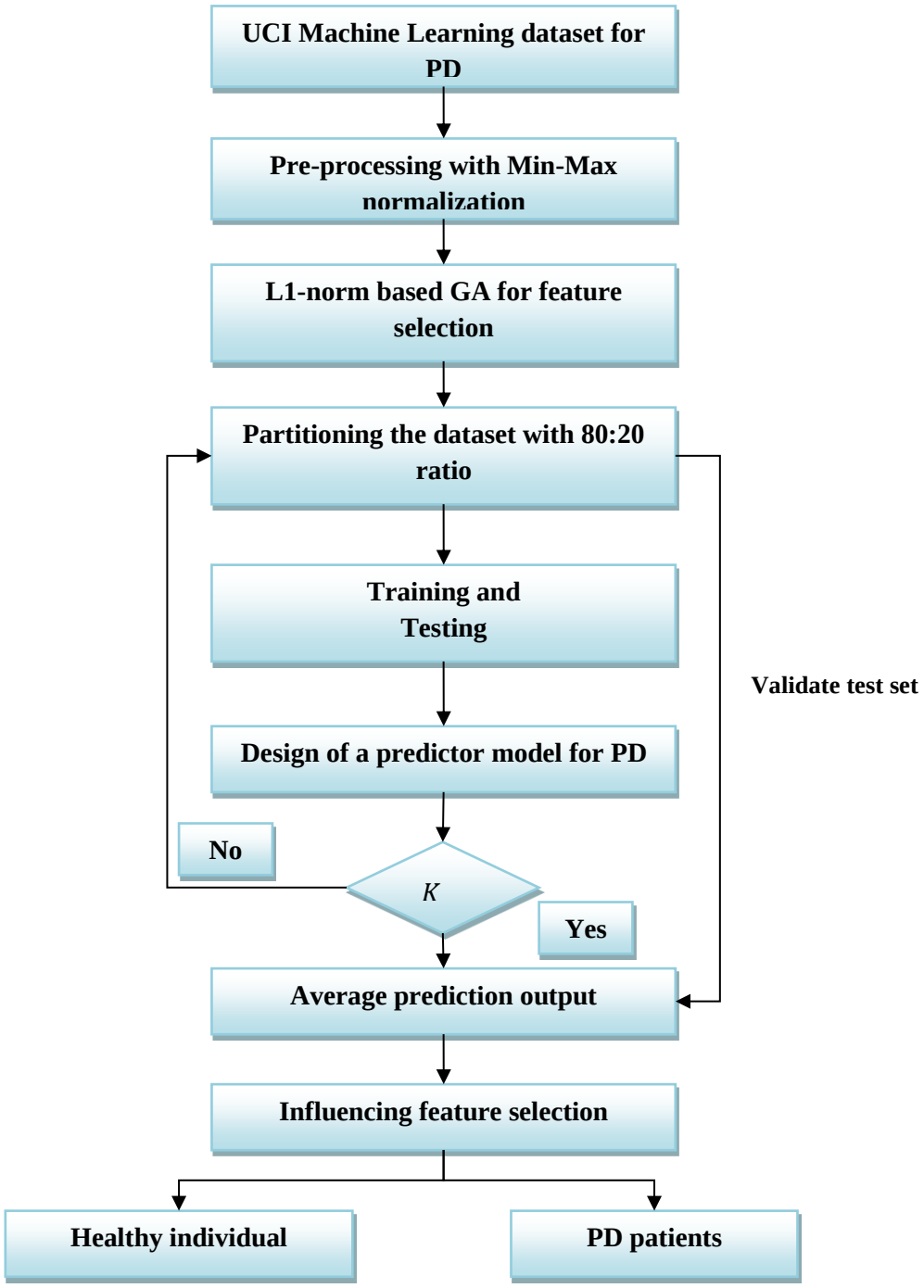

Fig 3: Flow chart of L1-norm based GA model

g. Selection process

Roulette-wheel selection process is utilized to produce successive or next generation from offspring and parent populations. Selection probability of chromosome $prob_c$ is directly proportional to fitness which is expressed as in Eq. (6):

$$prob_c = \frac{f(c)}{\sum f(parent) + \sum f(offspring)} \quad (6)$$

During genetic selection process, candidate chromosome produces better accuracy which is less likely to be removed and still have possibility of selection.

Algorithm 2
1. Initialize the process
2. Generate samples $x_i = \{x_{i1}, x_{i2}, \dots, x_i\}$ of dataset
3. perform CV on available samples for adjusting L1-norm regularization
4. Use L1-norm GA on every samples
5. Eliminate feature coefficient $\alpha = 0$ for all instances
6. Repeat the process with $\alpha = 0$ for features of all dataset samples
7. select feature subset for all instances
8. merge all the available features into newer feature set
9. generate $'x'$ to reduce feature set that includes $'k'$ features
10. end the process

4. Numerical results

Here, the simulation is performed using Python environment using Intel ® core i5 processor- 2400 CPU @ 3.10 Ghz,4 GB RAM and windows 10. The performance of the anticipated model is performed to measure the accuracy of feature selection. The values are mathematically expressed as in Eq. (7):

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} * 100\% \quad (7)$$

Here, TP is True Positive in which the individuals are classified correctly as PD; TN is True Negative in which the patient is categorized as healthy individual; FP is False Positive where healthy individual is classified as PD; FN is False Negative where the PD patient is classified as healthy. The model accuracy relies over the overall performance of the L1-norm based GA. To validate the performance of the diagnostic system various experimentations has been performed. This experimentation is related based on feature selection using L1-norm GA algorithm. The performance of the GA over UCI ML repository for PD over the available features performed using k-fold CV where $k = 2$. The experimental outcomes of the anticipated model are compared with various approaches. With the experimental results, various solutions are attained and provided a conclusion towards the research. This model pretends to fill the gap to promote feature selection and classification by an appropriate predictor model.

Here, the values attained during all iterations are given. The testing accuracy of the L1-norm based GA model is 98.67% with individual values [0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0]. There are 11 features over the feature subset is given as ['MDVP:Flo(Hz)', 'MDVP:PPQ', 'MDVP:Shimmer', 'Shimmer:APQ3', 'Shimmer:APQ5', 'Shimmer:DDA', 'NHR', 'HNR', 'RPDE', 'DFA', 'spread1']. The error value attained with this model is 0.013223140. Similarly, the final features are chosen based on random feature selection. The numbers of input features are 24 respectively ['name' 'MDVP: Fo (Hz)' 'MDVP: Fhi (Hz)' 'MDVP: Flo(Hz)' 'MDVP:

Jitter(%)', 'MDVP: Jitter(Abs)' 'MDVP:RAP' 'MDVP:PPQ' 'Jitter: DDP' 'MDVP: Shimmer', 'MDVP: Shimmer(dB)' 'Shimmer:APQ3' 'Shimmer:APQ5' 'MDVP:APQ' 'Shimmer: DDA', 'NHR' 'HNR' 'status' 'RPDE' 'DFA' 'spread1' 'spread2' 'D2' 'PPE']. Therefore, the optimally chosen features are 11 features out of 24 features. ['MDVP: Flo(Hz)', 'MDVP:PPQ', 'MDVP: Shimmer', 'Shimmer:APQ3', 'Shimmer:APQ5', 'Shimmer: DDA', 'NHR', 'HNR', 'RPDE', 'DFA', 'spread1'].

To perform PD prediction with reducing features sub-space, L1-norm based GA model is employed for generating various subset features from PD dataset. The feature selection process is based on weighted features. Therefore, 11 diverse feature subsets are modeled by available 24 features. They are: ['MDVP: Flo(Hz)', 'MDVP:PPQ', 'MDVP: Shimmer', 'Shimmer:APQ3', 'Shimmer:APQ5', 'Shimmer: DDA', 'NHR', 'HNR', 'RPDE', 'DFA', 'spread1']. The weighted features are based on feature subset. Some features possess negative values among available features and less significant for predicting PD as in Table II and III. Fig 4 depicts the error rate computation, Fig 5 shows fitness based accuracy measure, Fig 6 shows weighted feature selection, Fig 7shows accuracy measure and Fig 8 shows execution time of the process.

Table II: Features with weighted values

S.No	Weight	Feature name	Weight
1	F_2	MDVP:Fhi(Hz)	197.1
2	F_1	MDVP:Fo(Hz)	154.22
3	F_3	MDVP:Flo(Hz)	116.32
4	F_{16}	HNR	21.88
5	F_{18}	D2	2.31
6	F_{19}	DFA	0.71
7	F_{17}	RPDE	0.49
8	F_{10}	MDVP: Shimmer (dB)	0.28
9	F_{21}	Spread2	0.22
10	F_{22}	PPE	0.20
11	F_{14}	Shimmer: DDA	0.06
12	F_9	MDVP: Shimmer	0.029
13	F_{15}	NHR	0.024
14	F_{13}	MDVP: APQ	0.017
15	F_{12}	Shimmer: APQ5	0.015
16	F_{11}	Shimmer: APQ3	0.009
17	F_8	Jitter: DDP	0.006
18	F_4	MDVP: Jitter	0.003
19	F_7	MDVP: PPQ	0.003
20	F_6	MDVP: RAP	0.0004
21	F_5	MDVP: Jitter (Abs)	0.003
22	F_{20}	Spread 1	-5.6

Table III: Feature subset

Number of features	Features subset
1	F_2
2	F_1, F_2
3	F_1, F_2, F_3
4	F_2, F_1, F_3, F_{16}
5	$F_1, F_2, F_3, F_{16}, F_{18}$

6	$F_1, F_2, F_3, F_{16}, F_{18}, F_{19}$
7	$F_1, F_2, F_3, F_{16}, F_{18}, F_{17}, F_{19}$
8	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}$
9	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}$
10	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}$
11	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{14}$
12	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{9,}F_{14}$
13	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{9,}F_{14}, F_{15}$
14	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{9,}F_{14}, F_{15}, F_{13}$
15	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{9,}F_{14}, F_{15}, F_{13}, F_{12}$
16	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{9,}F_{14}, F_{15}, F_{13}, F_{12,}F_{11}$
17	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{9,}F_{14}, F_{15}, F_{13}, F_{12,}F_{11}, F_8$
18	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{9,}F_{14}, F_{15}, F_{13}, F_{12,}F_{11}, F_8, F_4$
19	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{9,}F_{14}, F_{15}, F_{13}, F_{12,}F_{11}, F_8, F_4, F_7$
20	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{9,}F_{14}, F_{15}, F_{13}, F_{12,}F_{11}, F_8, F_4, F_7, F_6$
21	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{9,}F_{14}, F_{15}, F_{13}, F_{12,}F_{11}, F_8, F_4, F_7, F_6, F_5$
22	$F_2, F_1, F_3, F_{16}, F_{18}, F_{19}, F_{17}, F_{10}, F_{21}, F_{22}, F_{9,}F_{14}, F_{15}, F_{13}, F_{12,}F_{11}, F_8, F_4, F_7, F_6, F_5, F_{20}$

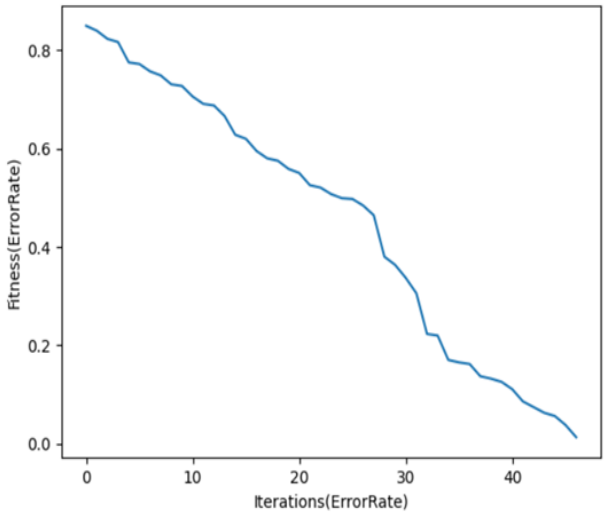

Fig 4: Error rate computation

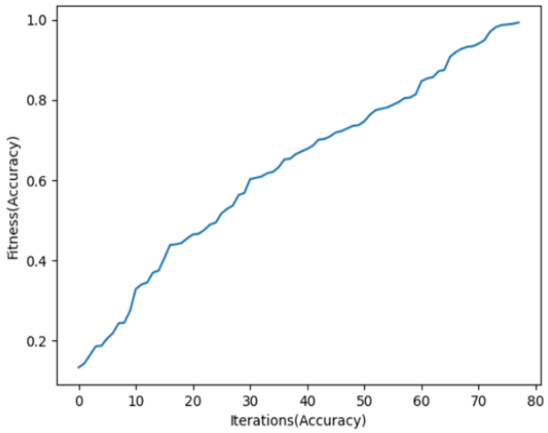

Fig 5: Fitness (accuracy) computation

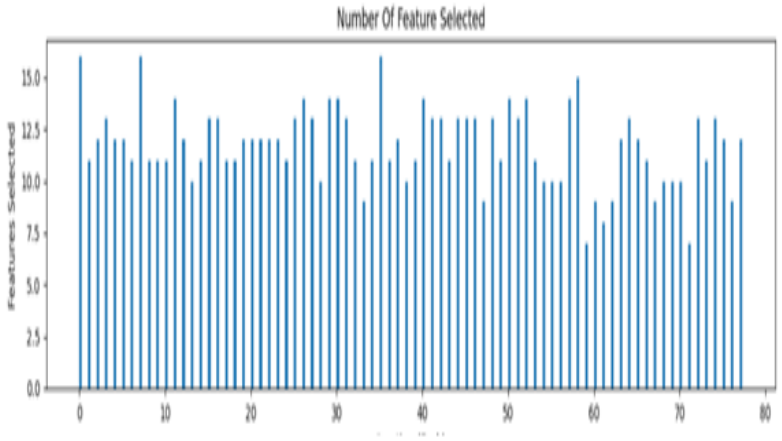

Fig 6: Feature selection computation

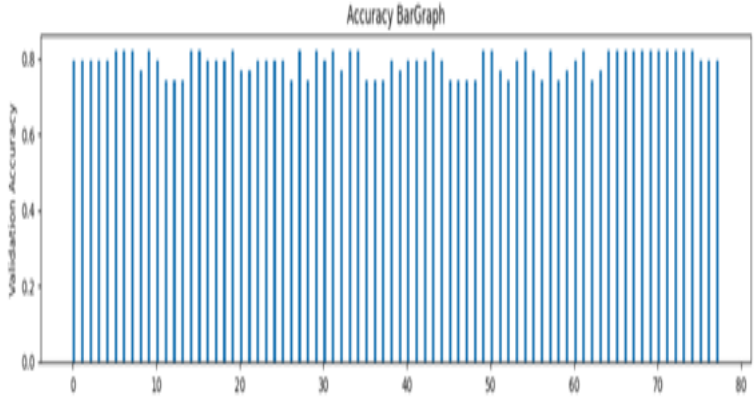

Fig 7: Accuracy computation

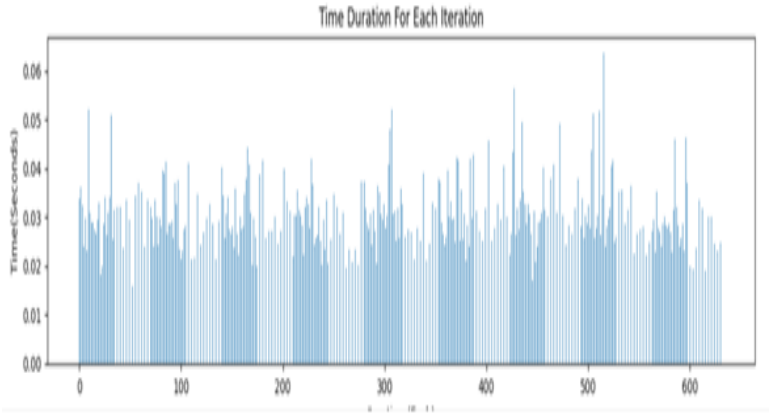

Fig 8: Execution time

Table IV: Accuracy comparison

S. No	Methods	Accuracy
1	Two-stage variable selection	86%
2	Deep Belief network	94%
3	RF-SVM	97%
4	Feature selection with SVM	99%
5	L1-norm GA	98.67%

The anticipated model is compared with prevailing approaches like two-stage variable selection, DBN, RF-SVM, and SVM model respectively where the anticipated model gives better accuracy than other models as in Table IV. However, it shows 0.33% accuracy lesser than SVM; 12.67%, 4.67%, and 1.67% respectively. Fig 9 depicts the accuracy computation of the anticipated model where L1-norm based GA gives 98.67% accuracy is selecting the most influencing features than the other models.

5. Conclusion

From this investigation, an efficient diagnosis system is modeled for predicting PD. In this predictor model, a ML approaches is used for predicting PD using the feature subset selection. Here, L1-norm based GA is used for feature selection with suitable and appropriate feature for accurate prediction of PD. This method generates newer feature subset with reduced complexity. Hence, an effectual feature selection process is used for handling critical factors and to choose better feature from the feature space for attaining optimal classifier performance. Feature selection approach discriminate the performance of anticipated model by differentiating the healthy persons from PD. Based on these factors, the anticipated provides excellence over the prevailing approaches for PD prediction. At present, CDSS is playing an effectual role in assisting the prediction of PD. As well, the anticipated model fills the research gap among the feature selection with voice recordings by appropriate experimental model. Reduced feature subset is attained with L1- norm based GA which shows highly influencing features that predicts PD more accurately when compared to unique feature spaces. GA performance is measured for reducing the feature subset compared to other models. Based on L1-norm based GA for feature selection, the most appropriate features are chosen from 11 out of 24 features. The randomly chosen features are: ['MDVP: Flo(Hz)', 'MDVP:PPQ', 'MDVP: Shimmer', 'Shimmer:APQ3', 'Shimmer:APQ5', 'Shimmer: DDA', 'NHR', 'HNR', 'RPDE', 'DFA', 'spread1']. These chosen features provide superior impact during classification process, i.e., healthy Vs PD. The novelty of this investigation is developed based on predictor model to select effectual features. The performance metrics like error rate, accuracy, fitness values are used for selecting the features effectually. This CDSS is modeled with ML approaches to show better prediction results. Moreover, the irrelevant features are eliminated as it degrades the system performance and increases the execution time. Therefore, an appropriate model is designed for predicting the features of PD and to treat it in the initial stage. This feature selection process is more appropriate to enhance the classification performance of diagnosis system. In future, feature selection can be performed in a hybridized manner for attaining better classification model and it is used for further enhancement in performance of diagnosis system for PD diagnosis.

REFERENCES

[1] D. K. Simon, C. M. Tanner, and P. Brundin, "Parkinson Disease Epidemiology, Pathology, Genetics, and Pathophysiology," *Clinics in Geriatric Medicine,* 2019.

[2] D. Georgiev, M. Domellof, K. Hamberg, L. Forsgren, and G. M. Hariz, "Sex differences, quality of life and non-motor symptoms in Parkinson's disease," 2019.

[3] J. C. Vásquez-Correa, N. Garcia-Ospina, J. R. Orozco-Arroyave, and E. Nöth, "Automatic Intelligibility Assessment of Parkinson’s Disease with Diadochokinetic Exercises," in *Workshop on Engineering Applications*, 2018, pp. 223-230.

[4] Vásquez-Correa, M. Stede, and E. Nöth, "Automated Cross-language Intelligibility Analysis of Parkinson’s Disease Patients Using Speech Recognition Technologies,"

in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, 2019, pp. 74-80.

[5] T. Arias-Vergara, J. C. Vásquez-Correa, and J. R. Orozco-Arroyave, "Parkinson’s disease and aging: analysis of their effect in phonation and articulation of speech," *Cognitive Computation,* vol. 9, pp. 731- 748, 2017.

[6] R. Castrillon, A. Acien, J. R. Orozco-Arroyave, A. Morales, J. Vargas, R. Vera-Rodrıguez*, et al.*, "Characterization of the Handwriting Skills as a Biomarker for Parkinson Disease," *arXiv preprint arXiv:1903.08226*, 2019.

[7] Miikkulainen, J. Liang, E. Meyerson, A. Rawal, D. Fink, O. Francon, *et al.*, "Evolving deep neural networks," in *Artificial Intelligence in the Age of Neural Networks and Brain Computing*, ed: Elsevier, 2019, pp. 293-312.

[8] Karan, Biswajit, SitanshuSekharSahu, and KartikMahto. "Parkinson disease prediction using intrinsic mode function based features from speech signal." Biocybernetics and Biomedical Engineering (2019).

[9] Espay, Alberto J., Joaquin A. Vizcarra, "Revisiting protein aggregation as pathogenic in sporadic Parkinsonand Alzheimer diseases." Neurology 92, no. 7 (2019): 329-337.

[10] Fan, Cheng, Fu Xiao, Yang Zhao, and Jiayuan Wang. "Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data." *Applied energy* 211 (2018): 1123- 1135

[11] Tan, Chun Chet, and ChikkannanEswaran. "Performance comparison of three types of autoencoder neural networks." In *2018 Second Asia International Conference on Modelling & Simulation (AMS)*, pp. 213-218. IEEE, 2018

[12] Kubota, Ken J., Jason A. Chen, and Max A. Little. "Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures." Movement disorders 31, no. 9 (2016): 1314-1326.

[13] Amin Ul Haq et.al "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Mobile Information Systems,vol:2018,pages 21, 2,December 2018

[14] Amin et al.,"Comparative analysis of the classification performance of machine learning classifiers and deep neural network classifier for Parkinson disease prediction," IEEE Int. computer conf. on wavelet active media technology and information processing, 2018.

[15] R. B. Postuma, D. Berg, M. Stern, W. Poewe, C. W. Olanow, W. Oertel, *et al.*, "MDS clinical diagnostic criteria for Parkinson's disease," *Mov Disord*, vol. 30, pp. 1591-601, Oct 2015

[16] D. W. Dickson, "Neuropathology of Parkinson disease," *Parkinsonism & Related Disorders*, vol. 46, pp. S30-S33, Jan 2018.

[17] I. Obeso, E. Casabona, R. Rodriguez-Rojas, M. L. Bringas, R. Macias, N. Pavon, *et al.*, "Unilateral subthalamotomy in Parkinson's disease: Cognitive, psychiatric and neuroimaging changes," *Cortex*, vol. 94, pp. 39-48, Sep 2017.

[18] E. Adeli, F. Shi, L. An, C.-Y. Wee, G. Wu, T. Wang, *et al.*, "Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data," *NeuroImage*, vol. 141, pp. 206-219, Nov 2016.

[19] A. Abós, H. C. Baggio, B. Segura, A. I. García-Díaz, Y. Compta, M. J. Martí, *et al.*, "Discriminating cognitive status in Parkinson's disease through functional connectomics and machine learning," *Scientific Reports*, vol. 7, p. 45347, Mar 2017.

[20] Liu, Q. Wang, E. Adeli, L. Zhang, H. Zhang, and D. Shen, "Exploring diagnosis and imaging biomarkers of Parkinson's disease via iterative canonical correlation analysis based feature selection," *Computerized Medical Imaging and Graphics*, vol. 67, pp. 21-29, Jul 2018

[21] Adeli, K. H. Thung, L. An, G. Wu, F. Shi, T. Wang, *et al.*, "Semi-Supervised Discriminative Classification Robust to Sample-Outliers and Feature-Noises," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1-1, 2018.

[22] Lei, Z. Huang, J. Zhang, Z. Yang, E.-L. Tan, F. Zhou, *et al.*, "Joint detection and clinical score prediction in Parkinson's disease via multi-modal sparse learning," *Expert Systems with Applications*, vol. 80, pp. 284-296, Sep 2017.

[23] G. E. Gillies, I. S. Pienaar, S. Vohra, and Z. Qamhawi, "Sex differences in Parkinson's disease," *Frontiers in Neuroendocrinology*, vol. 35, pp. 370-384, 2014

[24] H.-L. Chen, G. Wang, C. Ma, Z.-N. Cai, W.-B. Liu, and S.-J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinsonʻs disease," *Neurocomputing*, vol. 184, pp. 131-144, Apr 2016.

[25] M. Liu, J. Zhang, D. Nie, P. T. Yap, and D. Shen, "Anatomical Landmark based Deep Feature Representation for MR Images in Brain Disease Diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1-1, 2018.

[26] O. B. Tysnes and A. Storstein, ``Epidemiology of Parkinson's disease," J. Neural Transmiss., vol. 124, no. 8, pp. 901_905, 2017.

[27] Y. E. Kim and B. S. Jeon, ``Clinical implication of rem sleep behavior disorder in Parkinson's disease," J. Parkinson's Disease, vol. 4, no. 2, pp. 237_244, 2014.

[28] D. J. Brooks, ``Imaging approaches to Parkinson disease," J. Nucl. Med.,vol. 51, no. 4, pp. 596_609, Apr. 2010

[29] Arora, V. Venkataraman, A. Zhan, S. Donohue, K. Biglan, E. Dorsey, and M. Little, ``Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study," Parkinsonism Rel. Disorders,vol. 21, no. 6, pp. 650_653, Jun. 2015.

[30] Polat, ``A hybrid approach to Parkinson disease classi_cation using speech signal: The combination of SMOTE and random forests," in Proc. Sci. Meeting Elect.-Electron. Biomed. Eng. Comput. Sci. (EBBT),Apr. 2019, pp. 1_3.