IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 19, 2019, accepted March 14, 2019, date of publication March 21, 2019, date of current version April 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2906350

Feature Selection Based on L1-Norm Support Vector Machine and Effective Recognition System for Parkinson's Disease Using Voice Recordings

AMIN UL HAQ[1], JIAN PING LI[1], MUHAMMAD HAMMAD MEMON[1], JALALUDDIN KHAN[1], ASAD MALIK[2], TANVIR AHMAD[3], AMJAD ALI[4], SHAH NAZIR[5], IJAZ AHAD[4], AND MOHAMMAD SHAHID[6]

[1]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]School of information science and Technology, Southwest Jiaotong University, Chengdu 611756, China
[3]School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China
[4]Department of Computer Science and Software Technology, University of Swat, Saidu Sharif Swat 19200, Pakistan
[5]Department of Computer Science, University of Swabi, Swabi 23430, Pakistan
[6]Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei 10607, Taiwan

Corresponding author: Amin Ul Haq (khan.amin50@yahoo.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61370073, in part by the National High Technology Research and Development Program of China under Grant 2007AA01Z423, and in part by the project of the Science and Technology Department of Sichuan Province.

ABSTRACT The patient of Parkinson's disease (PD) is facing a critical neurological disorder issue. Efficient and early prediction of people having PD is a key issue to improve patient's quality of life. The diagnosis of PD specifically in its initial stages is extremely complex and time-consuming. Thus, the accurate and efficient diagnosis of PD has been a significant challenge for medical experts and practitioners. In order to tackle this issue and to accurately diagnosis the patient of PD, we proposed a machine-learning-based prediction system. In the development of the proposed system, the support vector machine (SVM) was used as a predictive model for the prediction of PD. The L1-norm SVM of features selection was used for appropriate and highly related features selection for accurate target classification of PD and healthy people. The L1-norm SVM produced a new subset of features from the PD dataset based on a feature weight value. For the validation of the proposed system, the K-fold cross-validation method was used. In addition, the metrics of performance measures, such as accuracy, sensitivity, specificity, precision, F1 score, and execution time, were computed for model performance evaluation. The PD dataset was in this paper. The optimal accuracy achieved the best subset of the selected features that might be due to various contributions of the PD features. The experimental findings of this paper suggest that the proposed method can be used to accurately predict the PD and can be easily incorporated in healthcare for diagnosis purpose. Currently, the computer-based assisted predictive system is playing an important role to assist in PD recognition. In addition, the proposed approach fills in a gap on feature selection and classification using voice recordings data by properly matching the experimental design.

INDEX TERMS Classification, feature selection, L1-norm support vector machine, Parkinson's disease diagnosis, performance, voice recording.

I. INTRODUCTION

Parkinson's disease (PD) is considered a common neurological sickness around the globe. Parkinson disease is a progressive and long-term disorder the central nervous system that badly affects people whose age is usually above 60 years. The cells suffering from PD do not have a consistent flow of dopamine with the motor system. The vocal impairment is hypothesized initial signs of the disease [1]. Dr. James Parkinson in 1817, PD was discovered as ''shaking palsy''. He identified six causes of PD where three of them were examined by him [2]. Diagnosed that people with

The associate editor coordinating the review of this manuscript and approving it for publication was Xinyu Du.

Parkinsonism has vocal disorders problems that affect their speech volume level and face complexity in the pronunciation of syllables and so forth. Thus to use vocal measurements as an effective diagnostic tool for PD recognition [3]. Singh *et al.* [4] Parkinson disease is the critical disorder sickness second to Alzheimer's disease and the complete PD treatment has not discovered till now. The existing technique of therapies is good for tackle PD symptoms. However, researchers have made attempts to find out the effective treatment strategy that ensures recovery and treatment. In [5] the PD diagnosis is being typically based on conducted few invasive techniques and empirical tests and examinations. The invasive based techniques in order to diagnose the PD are very expensive, less efficient, as well as very complex equipment's needed to conducts and the accuracy is also not satisfactory.

New approaches are needed to diagnose PD. Therefore, less expensive, simplified and reliable methods should be adapted to diagnosis disease and ensure treatments. However, noninvasive diagnosis techniques of PD require being investigated. Machine learning techniques are used to classify PD and healthy people. It has been measured that vocal issues of disorders can be assessed for early PD detection [6]. PD diagnosing and controlling using speech signals is more reliable. However, the telemonitoring techniques that use speech signals permit far off monitoring of PD.

To classify PD and healthy people the usage of speech signals is an effective technique for diagnosing PD from speech impairments. In Literature, different machines learning based classification techniques have been proposed to classify PD and healthy people from speech signals, and are reported in the study.

Tsanas *et al.* [7] used a dataset consisting of 263 speech samples from 43 people and 76.7 % of dataset were PD, the leftover dataset was healthy. They utilized an updated version of the dataset that was utilized in [8]. Little *et al.* [8] present an assessment of measures for the identity of PD subjects from healthy by detecting dysphonia. They diagnosed 23 PD and 8 healthy people and their dataset recorded vowels and used a Support Vector Machine (SVM) for classification and achieved classification accuracy 91.4 %. In [9] 132 extracted features from speech signals applied dysphonia methods. The database only contained vowels and some features extraction algorithms such as Least Absolute Shrinkage Selection Operator (LASSO), Minimal Redundancy Maximal Relevance (MRMR), Relief and Local Learning Based Feature Selection (LLBFS), were used and 10 features selected from 132 were selected by FS algorithms. These 10 features were used as input parameters for classification with two machine learning algorithms (Random Forests and SVM). In another study Tsana *et al.* [10] process speech signals of PD to compute a relationship between severity of the PD and disorder of speech. In [11] Gök studied the dataset used in [8]. They applied an ensemble of k-nearest neighbor (k-NN) algorithms to increase the accuracy. Features selection was deployed to find suitable features for prediction of PD. Bayestehtashk *et al.* [12] proposed a diagnosis of the severity of PD using speech signals. They designed a system that used analysis of regression for prediction of the severity of PD through sustained phonations. In [13] Taha used a machine learning algorithm SVM for classification of speech signals in PD and utilized the N-fold Cross validation technique. The data set for the experiment contained 240 running voice samples recorded from 60 PD and 20 healthy people. Those samples of speech were clinically rated via unified Parkinson's, score scale motor exam of speech (UPDRS-S).

Sakar *et al.* [3] collected multiple voice recording from 40 people in which 20 PD and 20 healthy. The voice samples 26 inclusive of everyday sentences, numbers, words and contained vowels had been gathered for each concern and 26 features had been extracted from voice signals by Praat Acoustic Analysis Software [14]. They carried out Leave-One-Subject-Out (LOSO) and s-LOO validation methods to compute the performance of K-NN and LibSVM classifiers [15]. They compared the classifier performance using performance measuring metrics like accuracy, sensitivity, specificity and Matthews's correlation coefficient (MCC). İ. Cantürk and Karabibe [16] proposed approach was designed on a machine learning based system and use speech signals. Four FS algorithms (LASSO, relief, LLBS, and MRMR) were applied to filter out the most appropriate features from the dataset. Moreover; classifiers such as Ada boost, SVM, k-NN, multi-layer perceptron (MLP), and Naïve Bayes (NB) were applied for classification PD and healthy subjects. Moreover, two validation techniques i.e. k-Fold, and LOSO were utilized for correct classification of PD. The proposed system performances were measured by performance measuring metrics such as accuracy of classification, sensitivity and specificity, and MCC. The computation complexity of algorithm also computed and the system was evaluated on a PD dataset which contained multiple types of voice signals. Wen *et al.* [17] proposed an efficient feature selection and classification system for vehicle detection. They used Haar-like feature section technique and RBF-SVM for vehicle detection. The proposed method achieved better performance. Hong *et al.* [18] proposed a feature selection method to improve the effectiveness of the text mining analysis. A new genetic algorithm was designed for text mining to increase the search performance. Furthermore, FSGA improved the clustering and speed performance. Zhu *et al.* [19] propose a framework of using PU learning for SbME using latent topics identified by a topic model for feature dimension reduction. The LDA method has a significantly smaller dimension than the term based method it is more practical in a SbME setting, where computational efficiency is crucial in providing real time update of search results per the user's query documents. Daassi-Gnaba *et al.* [20] proposed a system for Wood Moisture Content Prediction Using Feature Selection Techniques and a Kernel Method. The proposed system obtained high performance. Cai *et al.* [21] proposed framework for prediction of PD. They used SVM classifier and relief

TABLE 1. Subjects details of sex, age, PD stage and number of years since detected.

Instance Name (no of recording)	Subject label	Sex	Age	Stages (H&Y)	Year since diagnosis
phon_R01_S01_1 _7)	S01(PD)	M	78	3.0	0
Phon_R01_S34_1_6	S34(PD)	F	79	2.5	$\frac{1}{4}$
Phon_R01_S44_1_6	S44(PD)	M	67	1.5	1
Phon_R01_S20_1_6	S20(PD)	M	70	3.0	1
Phon_R01_S24_1_7	S24(PD)	M	73	2.5	1
Phon_R01_S26_1_6	S26(PD)	F	53	2.0	$\frac{1}{2}$
Phon_R01_S08_1_6	S08(PD)	F	48	2.0	2
Phon_R01_S39_1_6	S39(PD)	M	64	2.0	2
Phon_R01_S33_1_7	S33(PD)	M	68	2.0	3
Phon_R01_S32_1_6	S32(PD)	M	50	1.0	4
Phon_R01_S02_1_6	S02(PD)	M	60	2.0	4
Phon_R01_S22_1_6	S22(PD)	M	60	1.5	$4\frac{1}{4}$
Phon_R01_S37_1_6	S37(PD)	M	76	1.0	5
Phon_R01_S21_1_6	S21(PD)	F	81	1.5	5
Phon_R01_S04_1_6	S04(PD)	M	70	2.5	$5\frac{1}{2}$
Phon_R01_S19_1_6	S19(PD)	M	73	1.0	7
Phon_R01_S35_1_7	S35(PD)	F	85	4.0	7
Phon_R01_S05_1_6	S05(PD)	F	72	3.0	8
Phon_R01_S18_1_6	S18(PD)	M	61	2.5	11
Phon_R01_S16_1_6	S16(PD)	M	62	2.5	14
Phon_R01_S27_1_7	S27(PD)	M	72	2.5	15
Phon_R01_S25_1_6	S25(PD)	M	74	2.5	23
Phon_R01_S06_1_6	S06(PD)	F	63	2.5	28
Phon_R01_S10_1_7	S10(healthy)	F	46	n/a	n/a
Phon_R01_S07_1_6	S07(healthy)	F	48	n/a	n/a
Phon_R01_S13_1_7	S13(healthy)	M	61	n/a	n/a
Phon_R01_S43_1_7	S43(healthy)	M	62	n/a	n/a
Phon_R01_S17_1_6	S17(healthy)	F	64	n/a	n/a
Phon_R01_S42_1_6	S42(healthy)	F	66	n/a	n/a
Phon_R01_S50_1_6	S50(healthy)	F	66	n/a	n/a
Phon_R01_S49_1_7	S49(healthy)	M	69	n/a	n/a
	The labeled "n/a" for healthy people for which Parkinson, stage and years since detected is not applicable. "H&Y" refers to the Hoeln and Yahr PD stage, were higher values show that high level of disability [21]				

feature selection algorithm with bacterial foraging optimization (BFO) and achieved best classification performance. Naranjo *et al.* [22] proposed a classification system. They used two-stage features and classification approach (TSFSA) for Parkinson's disease diagnosis by applied sound recording replication and achieved the best performance. In another study Bi *et al.* [46] proposed a methodology for conducting importance-performance analysis through online review by the combination of LDA, IOVO-SVM and ENNM. The proposed method obtained effective analysis results with low cost and with small time. Liu *et al.* [47] proposed a framework for multi-class sentiment classification. They used different feature selection/machine learning algorithms and achieved good results compared to other existing studies. In [48] Liu *et al.* proposed a method for multi-class sentiment classification based on an improved one-vs.-one (OVO) strategy and the support vector machine (SVM) algorithm. The experimental results demonstrated that proposed method achieved high performance as compared to existing studies.

The main contribution of this study is to propose a machine-learning based system to successfully diagnose people with PD and improve the patient's life. Machine learning predictive model SVM was used for PD and healthy people classification. The L1-Norm SVM was used for appropriate features selection that improves the classification

performance of the classifier. We adopted the L1-Norm SVM for appropriate feature selection in this study because the classification performance of L1-Norm SVM FS based method is good as compare to other methods of classification for PD and healthy people. These methods where used other feature selection algorithms such as LASSO, MRMR, LLBFS [9], Relief with BFO [21] and two-stage feature selection method [18]. Furthermore, all these studies used these FS algorithms for the same dataset [8], [23]. The K Fold cross-validation was used in to select the best hyper parameters for best model evaluation. Performance evaluation metrics such as classification accuracy, sensitivity, and specificity were utilized to check the proposed system performance. The proposed system has been tested on PD data-set multiple types of sound signals.

Following are the key contributions of the proposed research study:

I. The performance of classifier checked on selected features subsets which are selected by L1-Norm SVM algorithm along with K–folds cross-validation technique.
II. The performance also checked on full features set and compared with performance on selected features sets.
III. The system has been tested on PD dataset and achieved very high classification performance.
IV. We suggest that the proposed system can be effectively diagnosis PD and easily incorporated in the healthcare system.

The paper reaming sections are organized as follows. Section 2, explored the PD dataset, features selection technique and classification algorithm in detail. The validation technique and performance evaluation metrics also briefly discuss in this section. In section 3 the experimental results are analyzed and discussed in details. The paper conclusion and future work direction are given in the last section 4.

II. MATERIALS AND METHOD

The sub-sections below discuss the materials and method of the proposed research work.

A. DATASET

Dataset used in the research was adopted from the repository of the University of Oxford (UO) with collaboration with national center for voice developed by little *et al.* [8] and is available at the UC Irvine repository of data mining [23]. The original research published that feature extraction methods for general voice disorders. The voice recordings of 31 people, including 23 people with Parkinson's disease contained 16 males and 7 females) and 8 healthy controls (males = 3 and females = 5) were deployed in the study. In the dataset table, each column for voice and each row are related to one of 195 voice recording from an individual subject. Additionally, the people of age from 46 to 85 years with a mean value of age is 65.8 and standard division 9.8. The main objective of this dataset was to classify people with Parkinson's disease from healthy people by finding differences in vowel vocalization. The ''status'' attribute is set to 0 for healthy and 1 for PD people. For each subject, an average of 6 phonation of a vowel was recorded for 36 second and total of 195 samples were recorded. The phonations were recorded in industrial acoustic company sound-treated booth by the microphone which at distance 8 cm from mouth and microphone was calibrated as presented in [24]. The voice speech signals were stored in the computer using a computerized speech laboratory. Table 1 shows the details of the subject [8] of each recording based on different measurements like vocal perturbation and nonlinear measurements and thus 23 features were extracted. Thus the extracted dataset size is 195*23 matrixes. Table 2 shows the 23 features of voice signals of PD dataset.

Algorithm 1 Proposed System

Begin
Step1: data preprocessing using standard scalar, and Min-Max scalar on PD dataset;
i.e. $V^- = \frac{v-min}{max-min}(new_{max} - new_{min}) + new_{min}$ in Eq(1)
Step2: selected features by L1 –Norm SVM;
Step3: For j = 1: k, performance estimation applied k-fold cross- validation, where k = 10
Training set = k-1 sub-group of 195 instances;
Testing set k-9 sub-set of 195 instances;
Step4: train classifiers with k-1 sub-groups with initial hyper- parameters values(C, γ);
Step5: validate classifier on a test set of 10- folds and achieved the best combination of hyper-parameters;
Repeat step 3 and 4;
Step6: Compute average classification results of 10 fold processing
i.e. $E = \frac{1}{10}\sum_{i=1}^{10} Ei$; Eq(13)
Step 7: performance of the best predictive model on j testing set;
Step8: finish;

B. THE PROPOSED SYSTEM METHODOLOGY

The proposed system designed to classify PD and healthy people. In the development of the proposed system, the machine learning predictive model SVM was used. The L1-Norm SVM algorithm was used for appropriate features selection that classifier effectively classifies PD and healthy subjects. Furthermore, the k-fold cross-validation technique was applied for best hyper-parameters and for predictive model selection. Four performance evaluation metrics were used for predictive model evaluation. The PD dataset which online available at UC Irvine data mining repository was used for testing of the proposed system. The methodology of the proposed system is structured into five steps, preprocessing of the dataset, features selection, cross-validation, and machine learning classifier performance evaluation. The framework of the proposed classification system as shown in Fig 1.

The following is the pseudo code of the proposed system.

TABLE 2. 23 Features of voice samples of PD dataset [8].

Label	Feature Name	Description	Min-Max	Mean , ± Std.
X1	MDVP:Fo(Hz)	The average vocal voice fundamental frequency	88.333000-260.105000	154.228641, ±41.390065
X2	MDVP:Fhi(Hz)	Maximum vocal fundamental frequency	102.145000-592.030000	197.104918, ±91.491548
X3	MDVP:Flo(Hz)	Minimum vocal fundamental frequency	65.476000-239.170000	116.324631, ±43.521413
X4	MDVP: Jitter (%)	Several measures of variation in fundamental frequency	0.001680-0.033160	0.006220, ±0.004848
X5	MDVP: Jitter (Abs)	-	0.000007-0.000260	0.000044, ±0.000035
X6	MDVP:RAP	-	0.000680-0.021440	0.003306, ±0.002968
X7	MDVP:PPQ	-	0.000920-0.019580	0.003446, ±0.002759
X8	Jitter : DDP	-	0.002040-0.064330	0.009920, ±0.008903
X9	MDVP:Shimmer	Several measures of variation in amplitude	0.009540-0.119080	0.029709, ±0.018857
X10	MDVP: Shimmer(dB)	-	0.085000-1.302000	0.282251, ±0.194877
X11	Shimmer:APQ3	-	0.004550-0.056470	0.015664, ±0.010153
X12	Shimmer:APQ5	-	0.005700-0.079400	0.017878, ±0.012024
X13	MDVP:APQ	-	0.007190-0.137780	0.024081, ±0.016947
X14	Shimmer: DDA	-	0.023370-0.104700	0.060043, ±0.029933
X15	NHR	Two measures of ratio of noise to tonal components in the voice	0.000650-0.314820	0.024847, ±0.040418
X16	HNR	-	8.441000-33.04700	21.885974, ±4.425764
X17	RPDE	Two nonlinear dynamical complexity measures	0.256570-0.685151	0.498536, ±0.103942
X18	D2	-	1.423287-3.671155	2.381826, ±0.382799
X19	DFA	Signal fractal scaling exponent	0.574282-0.825288	0.718099, ±0.055336
X20	spread1	Three nonlinear measures of fundamental frequency variation	-7.964984- -2.434031	5.684397, ±1.090208
X21	spread2	-	0.006274-0.450493	0.226510, ±0.083406
X22	PPE	-	0.044539-0.527367	0.206552, ±0.090119
y	Status	Health status of the subject Parkinson's=1 healthy=0	0.000000-1.000000	0.753846, ±0.431878

1) PREPROCESSING OF DATA

For a good representation of data preprocessing is a very important step and machine-learning classifier should be trained and tested effectively. Techniques of preprocessing include removing of missing values, standard scalar, Min-Max Scalar have been applied to the dataset. In standard Scalar ensures that every feature has the mean 0 and variance 1. Similarly, in Min-Max Scalar arrange the data such that all features are between 0 and 1, [26]. The features having missing values that feature row are deleted from the dataset. Mathematical form Min-Max normalization is expressed in equations (1).

Min-max normalization:

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

where V is the old feature value and V^{-} is the new one.

2) FEATURES SELECTION (FS) ALGORITHM

Features selection algorithms are necessary to remove irrelevant features from feature space. The reduced features will improve the accuracy of classification and deduced execution time of classifier. In this study, we use L1–Norm SVM algorithm for features selection. The formulation of the L1-Norm support vector machine is given below:

Consider a data set S with n instances. As expressed in equation (2):

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

where $\mathbf{x_i}$ is the $\mathbf{i^{th}}$ instance that has n features and a class label $\mathbf{y_i}$ where $\mathbf{x_i}$ expressed in equation (3):

$$x_i = \{x_{i1}, x_{i2} \ldots\ldots, x_{in}\} \qquad (3)$$

where $\mathbf{x_{ij}}$ is the values of the $\mathbf{j^{th}}$ features in instance $\mathbf{x_i}$.

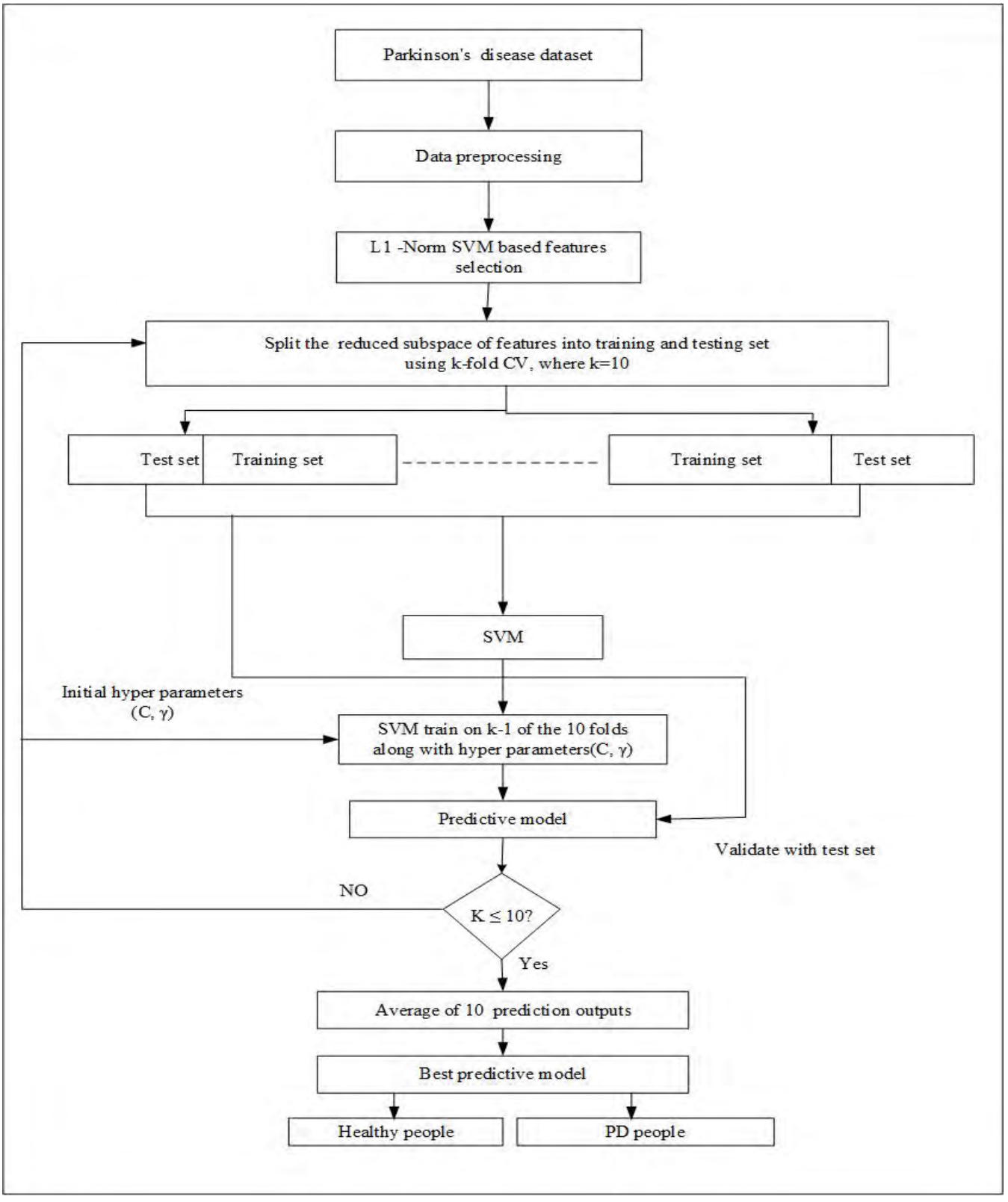

FIGURE 1. A framework of the proposed classification system.

The two class categorization problem, support vector machine SVM learns the separating hyper-plane $w \times x = b$ that maximize the margin distance $\frac{2}{\|\alpha\|_2^2}$, where α is the weight vector and b is the bias term. The primal form of SVM is given

in [27] and expressed mathematically in equation (4):

$$\min_{\alpha\ b} \frac{1}{2}\|\alpha\|_2^2 \tag{4}$$

Subject to the following conditions:

$$\left\{y_i\ (\alpha x_i - b) \geq 1, i = \overline{1, k}\right\}$$

In Cortes and Vapnik [28] suggested the SVM modified version and named it soft Margin SVM and mathematically written in equation (5):

$$\min_{\alpha,b,\S} \frac{1}{2}\|\alpha\|_2^2 + C\sum_{i=1}^{k} \S i \tag{5}$$

Subject to conditions below:

$$\begin{cases} y_i\ (\propto x_i - b) \geq 1 - \xi_i \\ \xi i \geq 0, \quad i = \overline{1, k} \end{cases}$$

where §(*i*) is slack variable, which computes the degree of misclassification of instance $\mathbf{x_i}$, $C > 0$ is error penalty parameter.

In [29], Bradley and Mangasarian proposed L1-Norm SVM algorithm for features selection as a consequence of the resulting sparse solutions and expressed in equation (6):

$$\min_{\alpha,b,\S} \|\alpha\|^1 + C\sum_{i=1}^{k} \S i \tag{6}$$

Subject to following conditions:

$$\left\{\begin{array}{c} y_i\ (\propto x_i - b) \geq 1 - \xi_i \\ \xi i \geq 0, \quad i = \overline{1, k} \end{array}\right\}$$

The optimization problem of L1-Norm SVM could be described in the below equation (7) [30].

$$\min_{\alpha,b} \|\alpha\|^1 + C\sum_{i=1}^{k} \max\left(0, 1 - yi\left(\alpha^T xi + b\right)\right)^2 \tag{7}$$

The following is the pseudo code of L1-Norm SVM algorithm for features selection:

Algorithm 2 Feature Selection

Begin
Step1: create n instances, $x_i = \{x_{i1}, x_{I2} \ldots x_{in}\}$ of the dataset; equation (3)
Step2: applied CV test on each instance for adjusting regularized hyper -parameter C and γ;
Step3: using L1-Norm SVM on each instance and calculated weight vector α for each feature;
Step4: remove the features who coefficient $\alpha = 0$ in each instance;
Step5: Compute the average CV score for each instance;
Step6: repeat step 2 to 5 up to that no features of $\alpha = 0$ for all each instance of dataset;
Step7: choose the subset of features for each instance of the dataset who cross-validation score values are high;
Step 8: combined all remaining features into a new reduced set of features;
Step9: produced x to reduced features set x^- that includes features in k;
Step10: finish.

3) MACHINE LEARNING CLASSIFIER

In this study, the following classifier was used for PD and healthy people classification. Here is the brief theoretical and mathematical background of the classifier is presented.

The support vector machine is classifier and for classification problem used mostly, [21], [31]–[35], [38]. Due to the good performance of classification SVM are used in various applications widely, [34], [35]. In the classification of the binary problem, the instances are separated with a hyper plane $w^T x + b = 0$, where w is a d-dimensional coefficient vector, which is normal to the hyper plane of the surface. The offset value from the origin is, and x is the data set values. The SVM get results of w and b. The W can solve by introducing Lagrangian multipliers in the linear case. On borders, the data set point are called support vectors. The solution of "w" can be expressed in equation (8):

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i \tag{8}$$

where support vectors number is n, y_i are output labels to x. The w and b values are computed, the linear function of discriminant can be written as in equation (9):

$$f(x) = \text{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i x_i^T x + b\right) \tag{9}$$

The non-linear case, the kernel trick, and decision function can be expressed in equation (10):

$$f(x) = \text{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right) \tag{10}$$

The positive semi definite functions that obey Mercer's condition as kernel functions [33]: Such as the polynomial kernel as expressed in equation (11):

$$\left(K(x, x_i) = \left(\left(x^T x_i\right) + 1\right)^d\right) \tag{11}$$

The Gaussian kernel as expressed in equation (12):

$$\left(K(x, x_i) = \exp\left(-\gamma\ \|x - x_i\|^2\right)\right) \tag{12}$$

There are two parameters that should be determined in the SVM model: C and γ.

4) VALIDATION METHOD

To check the proposed system performance K-folds Cross-validation (CV) [45], method and three evaluation metrics were used. In this study we used K-fold cross-validation and according to k-fold the data set was split into k identical components. The k-1 groups were applied for training and leftover was used for testing purposes in each step. The k times the process is iterated. Then the average of k results is computed to get the performance of the classifier. The cross-validation different Value of k was selected and we used the value of $k = 10$ in our experiments. In 10 fold CV process, 90% of the data used for training and 10% data were used for

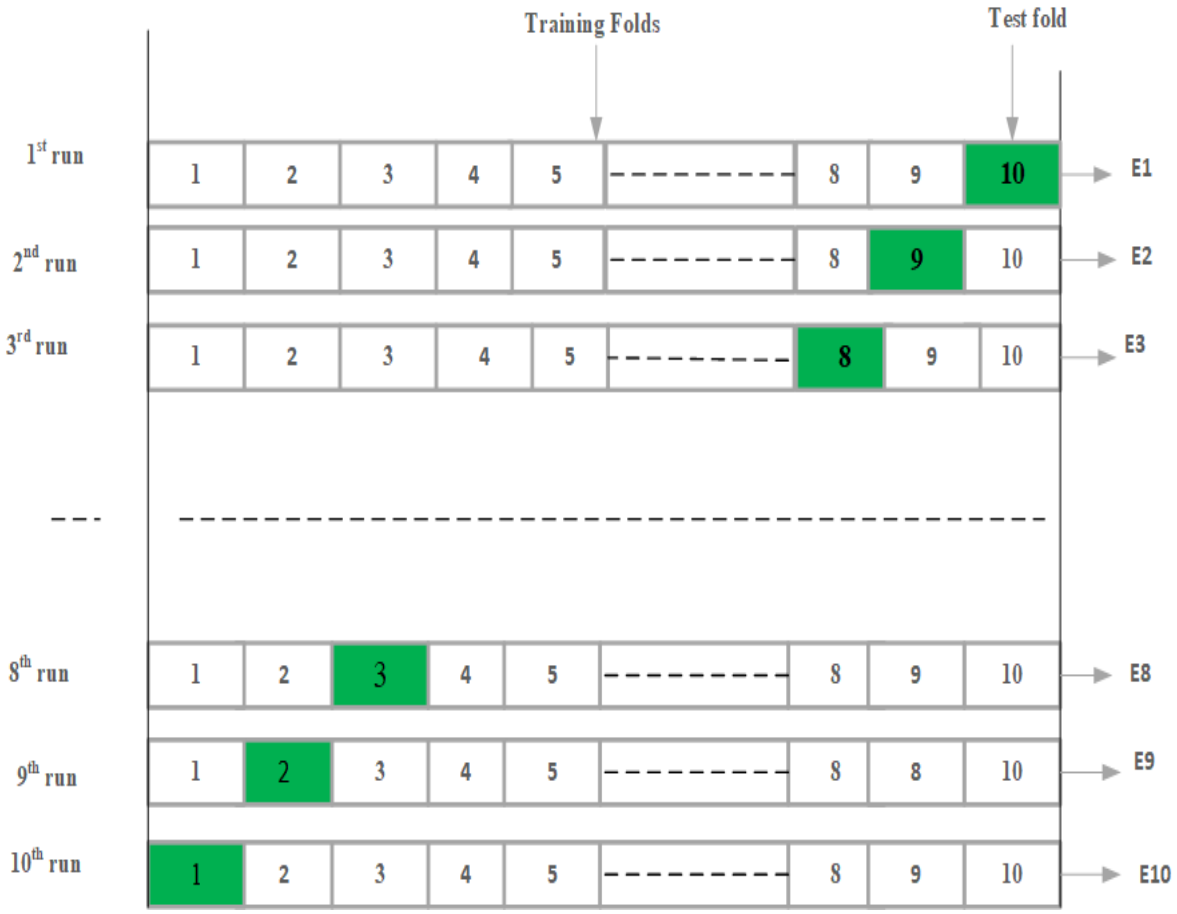

FIGURE 2. The 10-fold cross-validation process.

testing. The 10-time repeated the validation process. In the process of each fold, all samples are randomly distributed in the training and test groups over the entire dataset prior to selection of new training and test sets for the new cycle. Finally, at the end of 10 folds Processes, an average of all performance metrics are computed. As shown in fig 2 estimated performances Ei for each fold were computed and then used to calculate the estimated average performance E of the model. The mathematical equation for estimated average performance written in equation (13):

$$E = \frac{I}{10} \sum_{i=1}^{10} E_i \tag{13}$$

The k-fold cross-validation process, were k = 10 as shown in fig 2.

5) PERFORMANCES EVALUATION METRICS

Evaluation metrics used to evaluate the performance of classifier [3], [39]–[44]. In this study, three performance evaluation metrics were used. Table 3 shows the confusion matrix of the binary classification problem.

TABLE 3. Confusion matrix [16]–[37].

	Predicted PD subject	Predicted healthy subject
Actual PD subject	TP	FN
Actual healthy subject	FP	TN

According to Table 3 we compute the following metrics and mathematically expressed in equations (14), (15), (16), (17) and (18) respectively.

TP(True Positive) if the subject is classified as PD.

TN (True Negative) if a healthy subject is classified as healthy.

FP (False Positive) if a healthy subject is classified as PD.

FN (False Negative) if a PD is classified as healthy.

Classification Accuracy: Accuracy shows the overall performance of the classification system. Accuracy is the diagnostic test probability that correctly performed.

$$Ac = \frac{TN + TP}{TP + TN + FP + FN} * 100\% \tag{14}$$

TABLE 4. Weights and rank of 22 features in descending order.

S.no	Label order by weight	Feature name	Weight
1	X2	MDVP:Fhi(Hz)	197.1049
2	X1	MDVP:Fo(Hz)	154.2286
3	X3	MDVP:Flo(Hz)	116.3246
4	X16	HNR	21.88597
5	X18	D2	2.3181826
6	X19	DFA	0.71809
7	X17	RPDE	0.498536
8	X10	MDVP: Shimmer(dB)	0.282251
9	X21	spread2	0.22651
10	X22	PPE	0.206552
11	X14	Shimmer: DDA	0.060043
12	X9	MDVP: Shimmer	0.029709
13	X15	NHR	0.024847
14	X13	MDVP:APQ	0.024081
15	X12	Shimmer:APQ5	0.017878
16	X11	Shimmer:APQ3	0.015664
17	X8	Jitter: DDP	0.00992
18	X4	MDVP: Jitter	0.00622
19	X7	MDVP:PPQ	0.003446
20	X6	MDVP:RAP	0.003306
21	X5	MDVP: Jitter(Abs)	0.00044
22	X20	spread1	-5.6844

Sensitivity/Recall: The ratio of correctly classified heart patient subjects to all number of heart patient subjects. The sensitivity of the classifier the o detecting positive instances, it is known also as ''true positive rate''. Sensitivity (true positive fraction) that a diagnostic test is positive and subject has the disease Sensitivity (Sn) /Recall/True positive rate

$$Sn = \frac{TP}{Tp + FN} * 100\% \tag{15}$$

Specificity: Specificity shows that a diagnostic test is negative and the person is healthy.

$$\text{Specificity} = Sp = \frac{TN}{TN + FP} * 100\% \tag{16}$$

$$Precision = \frac{Tp}{(TP + FP)} * 100\% \tag{17}$$

F1- score: The traditional F-measure or balanced F-score (F1 score) is the harmonic mean of precision and recall:

$$F1 - score = 2 \cdot \frac{precision.recall}{precision + recall} \tag{18}$$

III. EXPERIMENTAL RESULTS ANALYSIS AND DISCUSSION

A. EXPERIMENTAL SETUP

In order to test the proposed diagnosis system performance, various experiments were conducted in this study. The first, experiment is concerned about features selection by L1-Norm SVM algorithm. In the second experiment, we checked the performance of support vector machine on PD dataset on full features with k-fold cross validation where k = 10. In the remaining experiments, classifiers performance were checked on 22 different features subsets that produced by L1 Norm SVM FS algorithm with 10 folds cross-validation. To check the performance of classifier performance measuring metrics were computed. The Experimental results of the proposed study compared with some of the states of the art methods. Furthermore, all experimental results are tabulated in tables, graphically demonstrated for better understanding.

Based on experimental results analysis and discussion various conclusions were derived and reported in conclusion sections. Additionally, the proposed approach fills in a gap on feature selection and classification using voice recordings data by properly matching the experimental design.

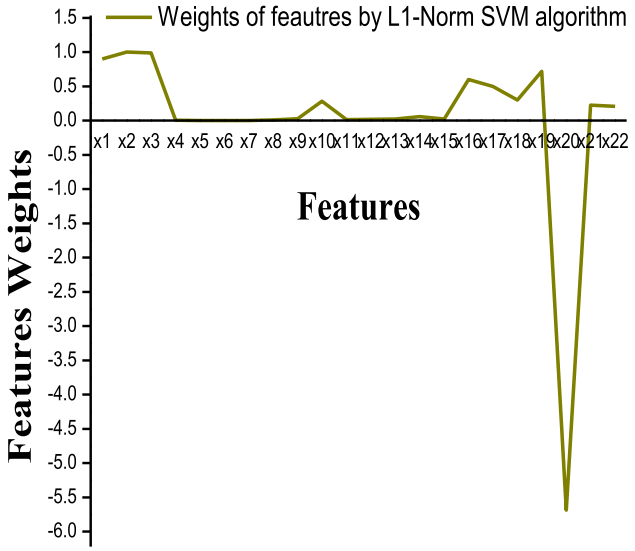

FIGURE 3. L1-Norm SVM algorithm based features weight and ranking.

The python on an Intel(R) Core^{TM} i5 -2400CPU @3.10 GHz CPU, 4 GB RAM, and window 10 experimental setups was used for computation of all experimental results.

B. EXPERIMENTAL RESULTS

1) EXPERIMENT 1. RESULTS OF THE SELECTED 22 DIFFERENT SUBSETS OF FEATURES BY L1-NORM SUPPORT VECTOR MACHINE (FS) ALGORITHM

To recognize the prediction of PD with reducing features subspace, L1-Norm SVM was used for creating reduce different subsets of features from the PD dataset. L1-Norm SVM features selection process based on feature weight. Thus 22 different subsets of features were constructed by eliminating feature step by step from feature set based on feature weight from lower to higher rank. The 22 features weight and ranking as shown in fig 3. The 22 features subsets were constructed in a detrimental way. The features such as X1 = MDVP: Flo (Hz), X2 = MDVP: Fhi (Hz), X3 = MDVP: Flo (Hz), X16 = HNR, X10 = DVP: Shimmer (dB), X17 = RPDE, X18 = D2 and X19 = DFA have very high weight value and these features includes in most subsets of features. Furthermore, all these features are critically necessary for PD prediction. The feature X 20 = spread1 have negative value among all the features and less significantly important for prediction of PD.

Table 5 shows the 22 different feature subsets created by L1-Norm SVM algorithm.

The fig 3 show the ranking of features.

2) EXPERIMENT 2. CLASSIFICATION PERFORMANCE ON FULL FEATURES SET WITH 10 FOLDS CV

In this experiment SVM kernels, RBF and linear were used for classification. The 10 folds cross-validation average value

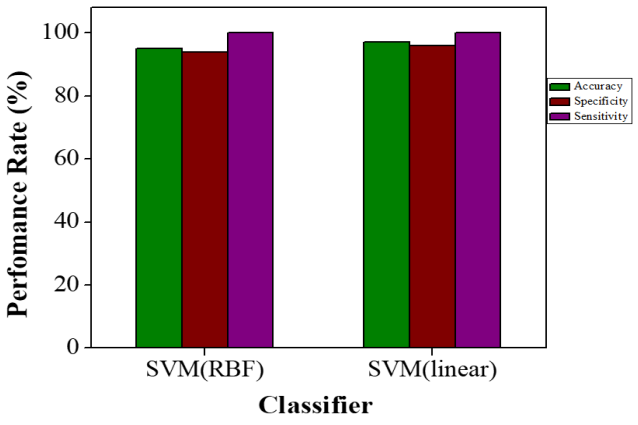

FIGURE 4. Classification performance on full features.

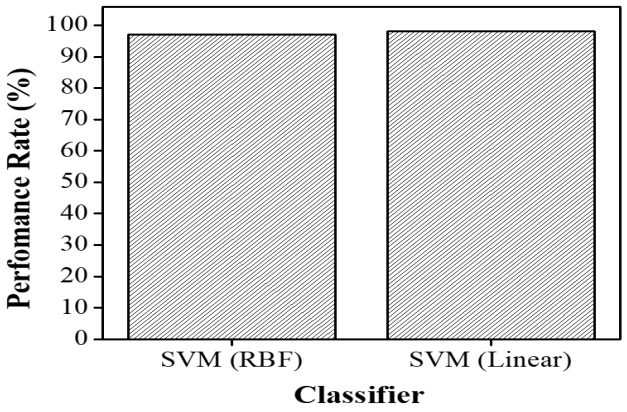

FIGURE 5. The F1- score evaluation of the classifier.

was computed on full features set with different values of hyper-parameters C and γ. These results were tabulated into Table 6. The classification performance of SVM kernel RBF with 10 folds CV on full features set where hyper-parameters values were C = 1 and γ = 0.025 and obtained good classification results as compared to other values of hypermeters. The average value of 10 folds the classification performance obtained 95% accuracy, 94 % specificity, and 100% sensitivity. The classification performance of SVM kernel linear on full features with hyper-parameters C = 1 and γ = 0.025 was good as compared SVM(RBF) and average classification performance of 10 folds achieved 97% accuracy, 96% specificity, and 100% sensitivity. The 96% specificity shows that the correct classification of healthy people, similarly 100 % sensitivity value shows that classifier accurately classification Parkinson disease people. The F1 –score performance metric also computed. In fig 4 the good classification performances of SVM (kernel = RBF) and SVM (kernel = linear) was shown graphically for better understanding. Fig 4 shows the classification performance of full features.

Fig 5 show the F1-score of the classifier evaluation.

TABLE 5. Feature subsets produced by L1-norm SVM (FS) algorithm.

Number of features in the subset	Features in subset
22	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14, X9, X15, X13, X12, X11, X8, X4, X7, X6, X5, X20 }
21	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14, X9, X15, X13, X12, X11, X8, X4, X7, X6, X5 }
20	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14, X9, X15, X13, X12, X11, X8, X4, X7, X6 }
19	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14, X9, X15, X13, X12, X11, X8, X4, X7 }
18	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14, X9, X15, X13, X12, X11, X8, X4 }
17	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14, X9, X15, X13, X12, X11, X8 }
16	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14, X9, X15, X13, X12, X11 }
15	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14, X9, X15, X13, X12 }
14	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14, X9, X15, X13 }
13	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14, X9, X15}
12	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14, X9}
11	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22, X14}
10	{X2, X1, X3, X16, X18, X19, X17, X10, X21, X22}
9	{X2, X1, X3, X16, X18, X19, X17, X10, X21}
8	{X2, X1, X3, X16, X18, X19, X17, X10}
7	{X2, X1, X3, X16, X18, X19, X17}
6	{X2, X1, X3, X16, X18, X19}
5	{X2, X1, X3, X16, X18}
4	{X2, X1, X3, X16}
3	{X2, X1, X3}
2	{X2, X1}
1	{X2}

C. RESULTS OF THE PROPOSED CLASSIFICATION SYSTEM

1) EXPERIMENT 3. CLASSIFICATION PERFORMANCE ON SELECTED SUBSETS OF FEATURES WITH 10 FOLDS CV

In this section, 22 experiments were performed on 22 subsets of the features. Each set of features with 10 folds cross-validation method was used and computes the average value of 10 folds. Thus 22 different records for 22 sets of features were constructed as reported in table 7. The SVM classifier was used as a classifier for the classification of PD and healthy people. Hyper-parameters C and γ different values were passed to the classifier. According to Table 7, the classification performance of SVM classifier on 22 features set was 97%, 96% and 100% in terms of accuracy, specificity, and sensitivity respectively. The performance on 21, 20, 19, 18 features sets as the same to 22 features set approximately in term of accuracy. The Classification performance of SVM on 17, 16, 15, 14, 13 and 12 features set was good as compared to 22 to 18 features sets. Classification performance on features sets 11 and 10 was very good as compared to other subsets of features. At 10 features set the classifier obtained classification accuracy of 99%, specificity 99%, and sensitivity 100%. Thus we reduced features subset 10 is suitable features subset for PD and healthy subject classification.

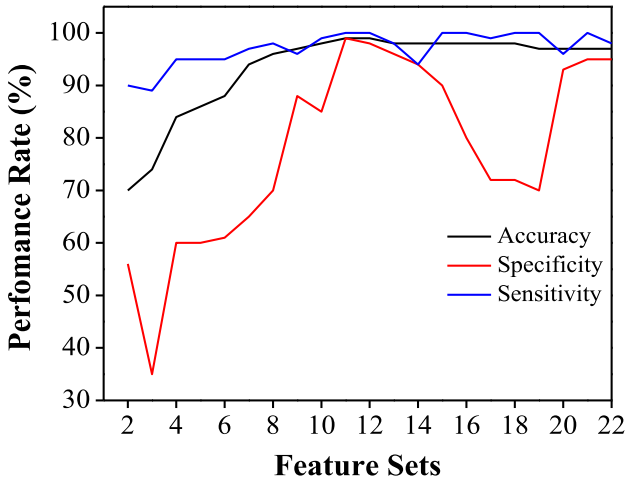

FIGURE 6. Classification performance on selected features subset.

The classification accuracy 99% shows the overall correct proposed system performance. The specificity 99% at 10 features set show that these features effectively predicted health people. Similarly, 100 % sensitivity show that the classifier accurately detected PD people on 10 features set.

TABLE 6. Classification on full features with 10 folds crosses validation.

Classifier	Hype-parameter		Performance evaluation metrics average value of 10 folds CV				
	C	γ	Accuracy (%)	Specificity (%)	Sensitivity/ Recall (%)	Precision (%)	F1-score
SVM (RBF)	1	0.015	93	92	90	91	91
	1	0.025	95	94	100	93	97
	10	0.015	85	86	75	87	81
SVM (Linear)	1	0.015	95	95	100	94	97
	1	0.025	97	97	100	96	98
	1	0.009	89	96	81	95	88
	10	0.015	86	90	66	89	76

The classification performance gradually reduced as the size of the dataset decreased. On reducing the feature set size 1 the classifier obtained 70% accuracy, 56% specificity, and 90% sensitivity, precision 99%. In fig 6 the classification performances on a different number of features subset graphically shown for better understanding of the results. The F1-score performance evaluation metric also computed for all experiments. In fig 7 the F1- score performance of classier on different selected feature sets shown. The execution of classifier was 1.120 seconds on full future set while on selected feature sets the execution time gradually reduced to 0.002 seconds. Hence feature selection reduced the execution time of classifier. In fig 8 the execution time of classifier on has been shown on full and on selected features set.

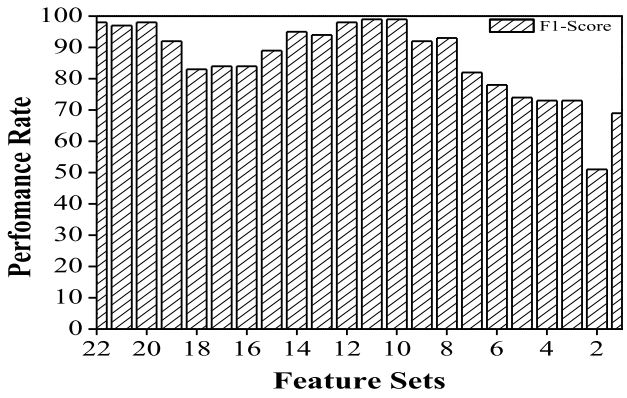

FIGURE 7. F1-score evaluation of the proposed classification system.

D. PERFORMANCE COMPARISION ON FEATURES FULL SET AND ON BEST SELECTED FEATURES SET

Fig 9 shows the classification performance of the classifier on features full set and on best-selected features set. On selected features subset of size 10 i.e. {X2, X1, X3, X16, X18, X19, X17, X10, X21, X22}, the classifier obtained 99% accuracy, 99% specificity, and 100% sensitivity and on full features set the obtained 97% accuracy of classification, 96 % speci-

TABLE 7. Classification performances on subsets of selected feature by L1-norm SVM algorithm with 10 folds cross-validation.

Classifier	Size of subset	Hyper-parameters		Performances evaluation metrics average value of 10 folds CV					
	n	C	γ	Accuracy (%)	Specificity (%)	Sensitivity / Recall (%)	Precision (%)	F1-score	Execution Time(s)
SVM	22	1	0.02	97	97	100	96	98	1.120
	21	1	0.03	97	96	98	95	97	1.111
	20	10	0.026	97	93	100	95	98	0.912
	19	10	0.086	97	93	96	93	92	0.800
	18	1	0.007	97	75	100	70	83	0.277
	17	10	0.041	98	74	100	72	84	0.264
	16	1	0.001	98	72	99	72	84	0.110
	15	1	0.01	98	86	100	80	89	0.080
	14	1	0.023	98	91	100	90	95	0.060
	13	1	0.001	98	94	94	94	94	0.045
	12	1	0.076	98	96	98	96	98	0.038
	11	10	0.008	98	92	100	98	99	0.026
	10	1	0.0001	99	99	100	99	99	0.022
	9	1	0.001	98	83	99	85	92	0.022
	8	1	0.009	97	88	96	88	93	0.005
	7	10	0.009	96	70	98	70	82	0.005
	6	1	0.03	94	66	97	65	78	0.005
	5	1	0.09	88	65	95	61	74	0.004
	4	1	0.003	86	61	95	60	73	0.003
	3	10	0.025	84	61	95	60	73	0.003
	2	1	0.001	74	35	89	35	51	0.002
	1	1	0.032	70	52	90	56	69	0.002

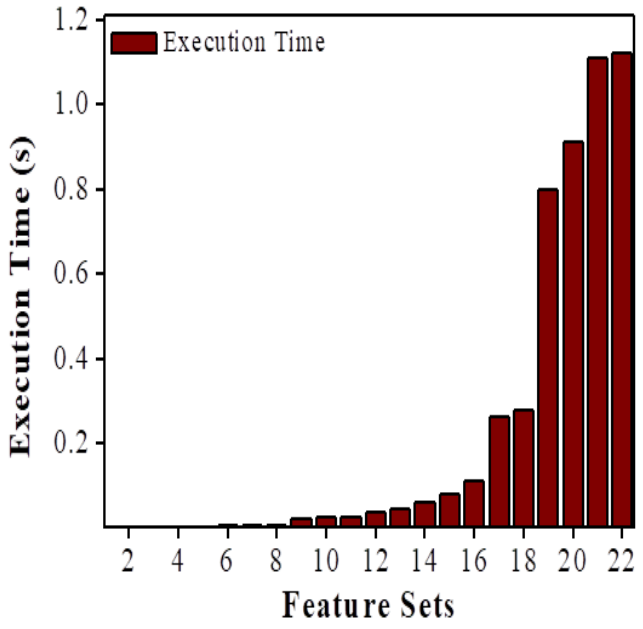

FIGURE 8. Execution time of classifier on different selected feature sets.

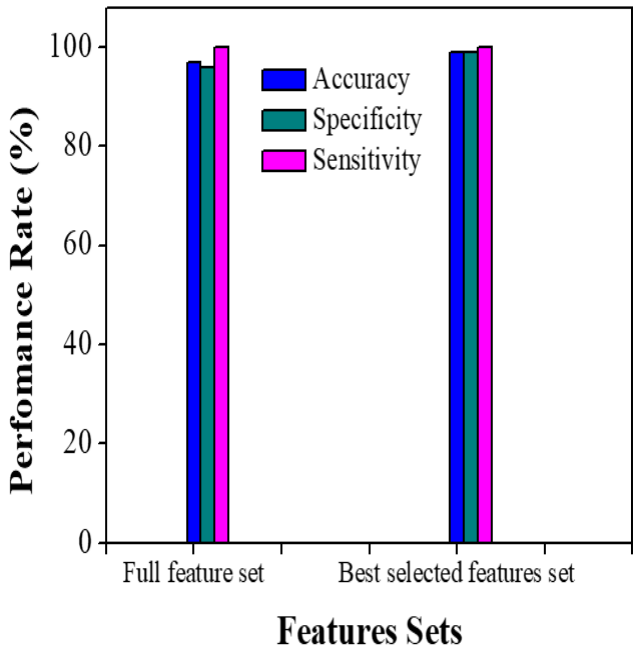

FIGURE 9. Classification performances on the full and on best-selected feature set.

ficity, and 100 % sensitivity. Thus from above experiments 2 and 3, we analyzed that relevant features increased the classification performance of the classifier. Fig 9 shows the classification performance on the full feature set and on the best-selected feature set. The execution time of classifier on full feature set was 1.120 seconds and on best feature sub was.022 seconds. Fig 10 shows the execution of classifier on the full feature set and on best-selected feature set shown graphically.

The fig 11 show the classification accuracy of the proposed method and other proposed previously methods.

TABLE 8. Proposed study classification performance and results of other previously proposed methods.

S/No	Reference	Method	year	Accuracy (%)
1	[18]	Two-stage variable selection and classification approach (TSFSA)	2017	86.2
2	[32]	Deep belief network (DBN)	2016	94
3	[17]	RF-BFO-SVM	2017	97.42
4	The proposed method	Feature selection based on L1-Norm support vector machine and classification method for PD recognition	2018	99

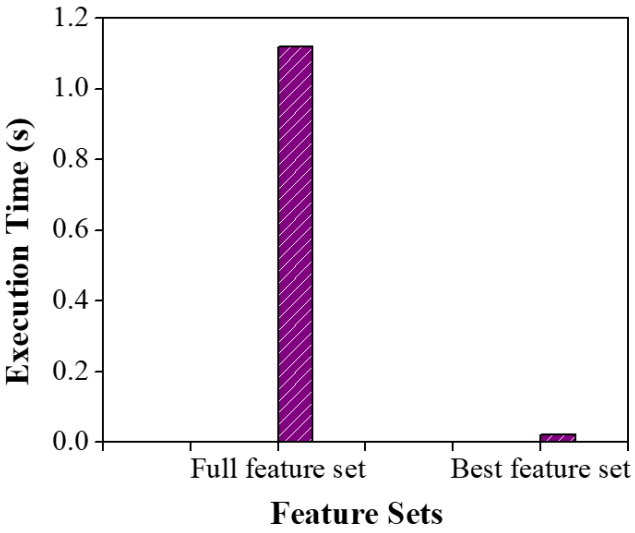

FIGURE 10. Classifier execution time on full feature set and on best selected feature set.

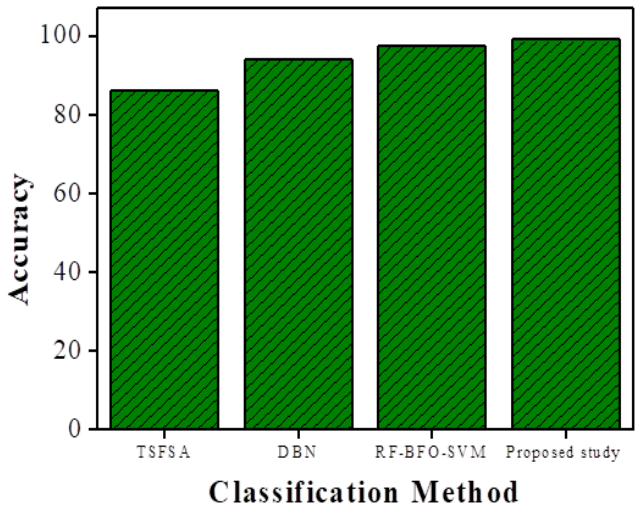

FIGURE 11. Classification accuracy of the proposed method and other methods previously proposed.

According to table 8 and Fig 11, the classification accuracy of the proposed study is good as compare to other previous studies for Parkinson disease recognition. The proposed study achieved 99 % classification accuracy, 99% specificity show that the proposed technique is good for detection of healthy people. The 100% sensitivity correctly detected the Parkinson disease people. Similarly, the F1 score performance of the proposed approach is good. The Proposed study concluded that the predictive performance of the machine learning classifier, improved when using certain important features instead of all features in the dataset. Due to appropriate features selection classifier predictive performance increased and reduced the computational complexity. Therefore effective feature selection algorithms definitely take critical roles to select the best feature from feature space for the optimal performance of machine learning classifier. The L1-Norm SVM based selected features effectively discriminated the people of Parkinson disease and healthy people. Due to these reasons, the propped approach preferences are excellent from previous studies of PD recognition. Currently, computer-based assisted predictive system playing an important role to assist in PD recognition. Additionally, the proposed approach fills in a gap on feature selection and classification using voice recordings data by properly matching the experimental design.

IV. CONCLUSIONS

To the best of the author's knowledge in this study, an efficient diagnosis system was developed for the prediction of PD. In the development of the system, machine learning classifier SVM was used for PD and healthy people classification. L1-Norm support vector machine of feature selection was used for appropriate and highly related features for accurate target classification of PD and healthy people. L1-Norm SVM algorithm produced new subsets of features from PD dataset based on feature coefficient weight value. The k-folds cross-validation, where $k = 10$ was applied to select the optimal values of tuning parameters for the best classification model. Additionally, the techniques of performance measuring metrics such as accuracy, sensitivity, and specificity, precision, recall, and F1-score and execution time(seconds)

were used for model performance evaluation. The PD dataset of 23 attributes and 195 instances available on UC Irvine data mining repository was used for testing of the proposed system. Machine learning libraries in python were used for the implementation and development of the proposed system. The experimental results analysis shows that the proposed system classify the PD and healthy people effectively. The improvement of PD prediction might be due to various contributions the PD features. These findings suggest that the proposed diagnosis system could be used to accurately predict PD and furthermore could be easily incorporated in healthcare. Currently, computer-based assisted predictive system is playing an important role to assist in PD recognition. Additionally, the proposed approach fills in a gap on feature selection and classification using voice recordings data by properly matching the experimental design.

The reduced space of features by L1-Norm SVM FS algorithm show that these are highly important features that diagnosis PD accurately as compared to original features space. The classification performance of SVM classifier on reduced features subset 10 was excellent as compared full features set and on other reduced features subsets. The average 10 folds cross-validation of classifier obtained 99% accuracy, 99% specificity, and 100% sensitivity. The 99% specificity value shows that it is good for the detection of healthy people. Similarly, 100% sensitivity show that classifier effectively detected PD people. The 10 folds cross-validation is an effective method of validation. According to L1-Norm support vector machine feature selection algorithm, select the most important features are MDVP: Fhi (Hz), MDVP: Fo (Hz), MDVP: Flo (Hz), HNR, D2, DFA, RPDE, MDVP: Shimmer (dB), spread2, and PPE. These features have great impacts on the classification of PD and healthy people.

The novelty of this study is developing a system of diagnosis to classify PD and healthy People. The system used the FS algorithm L1-Norm support vector machine, classifier, cross-validation technique, and performance measuring metrics for PD diagnosis. As we think that decision support system development through machine learning approach it will be better for prediction of PD. Furthermore, we know that irrelevant features also degrade the performance of the diagnosis system and computation time increase. Hence, another innovative part of proposed study to used features selection algorithm to select a relevant subset of features that improve the classification performance diagnosis system. According to Table 8 and figure 11 the performance of the proposed system is excellent and achieved 99% classification as compared to the classification performances of other proposed studies. The execution time of classifier on full feature set was 1.120 seconds and on best feature sub was 0.022 seconds. Hence feature selection reduced the execution time of classifier.

In the future other features selection algorithms, optimization and deep neural network classification methods will be utilized to further increase the performance of the diagnosis system for PD diagnosis.

AVAILABILITY OF DATA AND MATERIALS

The dataset used in this research work is available on UC Irvine machine learning repository.

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES

[1] J. R. Duffy, *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. Amsterdam, The Netherlands: Elsevier, 2005.
[2] J. W. Langston, "Parkinson's disease: Current and future challenges," *NeuroToxicology*, vol. 23, pp. 443–450, Oct. 2002.
[3] B. E. Sakar *et al.*, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 4, pp. 828–834, Jul. 2013.
[4] N. Singh, V. Pillay, and Y. E. Choonara, "Advances in the treatment of Parkinson's disease," *Prog. Neurobiol.*, vol. 81, pp. 29–44, Jan. 2007.
[5] *Parkinson's Disease: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care*, Nat. Collaborating Centre Chronic Conditions, London, U.K., 2006.
[6] B. Harel, M. Cannizzaro, and P. J. Snyder, "Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study," *Brain Cogn.*, vol. 56, no. 1, pp. 24–29, Jun. 2004.
[7] A. Tsanas, 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 Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, May 2012.
[8] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009.
[9] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," *J. Roy. Soc. Interfaces*, vol. 8, pp. 842–855, Jun. 2011.
[10] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 884–893, Apr. 2010.
[11] M. Gök, "An ensemble of k-nearest neighbours algorithm for detection of Parkinson's disease," *Int. J. Syst. Sci.*, vol. 46, pp. 1108–1112, Apr. 2015.
[12] A. Bayestehtashk, M. Asgari, I. Shafran, and J. McNames, "Fully automated assessment of the severity of Parkinson's disease from speech," *Comput. Speech Lang.*, vol. 29, no. 1, pp. 172–185, Jan. 2016.
[13] K. Taha, J. Westin, and M. Dougherty, "Classification of speech intelligibility in Parkinson's disease," *Biocybern. Biomed. Eng.*, vol. 34, pp. 35–45, Jul. 2015.
[14] J. Howell, "When technology is too hot, too cold or just right," *Emerg. Learn. Des. J.*, vol. 5, no. 1, pp. 9–18, May 2017.
[15] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.
[16] İ. Cantürk and F. Karabiber, "A machine learning system for the diagnosis of Parkinson's disease from speech signals and its application to multiple speech signal types," *Arabian J. Sci. Eng.*, vol. 41, pp. 5049–5059, Dec. 2016.
[17] X. Wen, L. Shao, W. Fang, and Y. Xue, "Efficient feature selection and classification for vehicle detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 508–517, Mar. 2015.
[18] S.-S. Hong, W. Lee, and M.-M. Han, "The feature selection method based on genetic algorithm for efficient of text clustering and text classification," *Int. J. Adv. Soft Comput. Appl.*, vol. 7, pp. 2074–8523, Mar. 2015.
[19] M. Zhu, C. Xu, and Y.-F. B. Wu, "Positive unlabeled learning to discover relevant documents using topic models for feature selection," in *Proc. Int. Conf. Data Mining Streeing Committee World Congr. Comput. Sci., Comput. Eng. Appl. Comput.*, 2014, pp. 1–7.
[20] H. Daassi-Gnaba, Y. Oussar, M. Merlan, T. Ditchi, E. Géron, and S. Holé, "Wood moisture content prediction using feature selection techniques and a kernel method," *Neurocomputing*, vol. 237, pp. 79–91, May 2017.
[21] Z. Cai, J. Gu, and H.-L. Chen, "A new hybrid intelligent framework for predicting Parkinson's disease," *IEEE Access*, vol. 5, pp. 17188–17200, Sep. 2017.

[22] L. Naranjo, C. J. Pérez, J. Martín, and Y. Campos-Roca, "A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications," *Comput. Methods Programs Biomed.*, vol. 142, pp. 147–156, Feb. 2017.

[23] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomed. Eng. Online*, vol. 6, no. 23, pp. 1–19, Jun. 2007.

[24] J. G. Švec, P. S. Popolo, and I. R. Titze, "Measurement of vocal doses in speech: Experimental procedure and signal processing," *Logopedics Phoniatrics Vocology*, vol. 28, no. 4, pp. 181–192, 2003.

[25] M. M. Hoehn and M. D. Yahr, "Parkinsonism: Onset, progression, and mortality," *Neurology*, vol. 50, no. 2, p. 318, 1998.

[26] S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, "Data preprocessing for supervised leaning," *Int. J. Comput. Sci.*, vol. 1, no. 2, pp. 111–117, 2006.

[27] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Science & Business Media, Jun. 2013.

[28] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[29] P. S. Bradley and O. L. Mangasarian, "Feature selection via concave minimization and support vector machines," in *Proc. 15th Int. Conf. (ICML)*, 1998, pp. 82–90.

[30] M. Moon and K. Nakai, "Stable feature selection based on the ensemble L1-norm support vector machine for biomarker discovery," *BMC Genomics*, vol. 17, p. 1026, Dec. 2016.

[31] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*. Cambridge, U.K.: Cambridge Univ. Press, 2000, p. 23.

[32] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[33] H.-L. Chen, B. Yang, J. Liu, and D.-Y. Liu, "A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis," *Expert Syst. Appl.*, vol. 38, pp. 9014–9022, Jul. 2011.

[34] J. Mourao-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data," *NeuroImage*, vol. 28, pp. 980–995, Nov. 2005.

[35] V. D. A. Sánchez, "Advanced support vector machines and kernel methods," *Neurocomputing*, vol. 55, pp. 5–20, Sep. 2003.

[36] A. H. Al-Fatlawi, M. H. Jabardi, and S. H. Ling, "Efficient diagnosis system for Parkinson's disease using deep belief network," in *Proc. IEEE Congr. Evol. Comput.*, Jul. 2016, pp. 1324–1330.

[37] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," *Mobile Inf. Syst.*, vol. 2018, Dec. 2018, Art. no. 3860146.

[38] A. U. Haq *et al.*, "Comparative analysis of the classification performance of machine learning classifiers and deep neural network classifier for prediction of parkinson disease," in *Proc. IEEE Int. Comput. Conf. Wavelet Active Media Technol. Inf. Process.*, Dec. 2018, pp. 101–106.

[39] A. Ul Haq, J. P. Li, M. H. Memon, M. H. Memon, J. Khan, and S. M. Marium, "Heart disease diagnosis through machine learning predictive model using sequential backward selection feature algorithm," in *Proc. IEEE 5th Int. Conf. Converg. Technol. (I2CT)*, 2019, pp. 1–5.

[40] M. Durairaj and N. Ramasamy, "A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate," *Int. J. Control Theory Appl.*, vol. 9, no. 27, pp. 255–260, 2016.

[41] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3240–3247, 2009.

[42] M. M. Islam, H. Iqbal, M. R. Haqueet, and M. K. Hasan "Prediction of breast cancer using support vector machine and K-nearest neighbors," in *Proc. IEEE Region Hum. Technol. Conf. (R10-HTC)*, Dhaka, Bangladesh, Dec. 2017, pp. 226–229.

[43] D. Zhang, L. Zou, X. Zhou, and F. He, "Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer," *IEEE Access*, vol. 6, pp. 28936–28944, Jun. 2018.

[44] N. Liu, J. Shen, M. Xu, D. Gan, E.-S. Qi, and B. Gao, "Improved cost-sensitive support vector machine classifier for breast cancer diagnosis," *Math. Problems Eng.*, vol. 2018, Oct. 2018, Art. no. 3875082.

[45] S. Nazir, S. Shahzad, and L. S. Riza, "Birthmark-based software classification using rough sets," *Arabian J. Sci. Eng.*, vol. 42, pp. 859–871, Feb. 2017. doi: 10.1007/s13369-016-2371-4.

[46] J.-W. Bi, Y. Liu, Z.-P. Fan, and J. Zhang, "Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews," *Tourism Manage.*, vol. 70, pp. 460–478, Feb. 2019. doi: 10.1016/j.tourman.2018.09.010.

[47] Y. Liu, J.-W. Bi, and Z.-P. Fan, "Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms," *Expert Syst. Appl.*, vol. 80, pp. 323–339, Sep. 2017. doi: 10.1016/j.eswa.2017.03.042.

[48] Y. Liu, B. Jian-Wu, and Z.-P. Fan, "A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm," *Inf. Sci.*, vol. 394, pp. 38–52, Jul. 2017. doi: 10.1016/j.ins.2017.02.016.

AMIN UL HAQ received the M.S. degree in computer science. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, China. He has a vast academic, technical, and professional experience in Pakistan. He is a Lecturer with Agricultural University Peshawar, Pakistan. He is also associated with the Wavelets Active Media Technology and Big Data Laboratory as an International Student. His research interests include machine learning, medical big data, the IoT, e-health and telemedicine, and concerned technologies and algorithms.

JIAN PING LI is currently the Chairman of the Computer Science and Engineering College and the Model Software College, University of Electronic Science and Technology of China. He is also the Director of the International Centre for Wavelet Analysis and Its Applications. He is the Chief Editor of the International Progress on Wavelet Active Media Technology and Information Processing. He also serves as a Technical Adviser and holds a dozen academic and social positions at the National Science and Technology Award Evaluation Committee, the National Natural Science Foundation Committee of China, and the Ministry of Public Security of China. He is also an Associate Editor of the *International Journal of Wavelet Multimedia and Information Processing*.

MUHAMMAD HAMMAD MEMON was born in Pakistan. He received the M.Eng. degree in computer science and engineering and the Ph.D. degree from the University of Electronic Science and Technology of China, in 2014 and 2017, respectively, where he is currently a Postdoctoral Researcher with the School of Computer Science and Engineering. In his Ph.D. degree, he has studied the problem related to the field of image processing and wavelet transformation. He has professional certifications in computer science, networking, and business management from USA, United Arab Emirates, and Pakistan. He has published over 23 research papers in recent years. His research interests include artificial intelligence, network security, cloud computing, image processing, and CBIR techniques. During his study, he received several excellent prestigious awards, such as the Academic Achievement Award 2015–2016 and the Excellent Performance Award 2014–2015 and 2015–2016 from UESTC. He is an Associate Editor of the IEEE Access and for various reputed international journals.

JALALUDDIN KHAN received the M.S. degree in computer science from Aligarh Muslim University, Aligarh, India. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China. He has an impressive academic, research, and professional experience in Saudi Arabia. He was a Lecturer in the Deanship of Skills Development as well as a Researcher in the Center of Excellence in Information Assurance (COEIA), King Saud University, Riyadh, Saudi Arabia, for seven years. He is accompanying with the Wavelets Active Media Technology and Big Data Laboratory under the supervision of Prof. J. P. Li and collaborated work with researcher Mr. A. U. Haq. He has authored some research papers. His research interests include the IOT, security and privacy, e-health and telemedicine, machine learning, and medical big data concerned technologies. He is currently focusing on the IOT security with medical data.

ASAD MALIK received the B.Sc. degree (Hons.) in computer application from Aligarh Muslim University, in 2012, and the master's degree in computer application from Jamia Millia Islamia University, India, in 2015. He is currently pursuing the Ph.D. degree with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China. His research interests include information security, image processing, reversible data hiding in encrypted images, and machine learning.

TANVIR AHMAD received the master's degree in software engineering from the University of Electronic Science and Technology of China, in 2017. He is currently pursuing the Ph.D. degree with North China Electric Power University, Beijing, China, where he has been with the Center of Deep Knowledge Engineering, School of Control and Computer Engineering, since 2017. His research interests include machine learning, object detection recognition, pedestrian detection, computer vision, pattern recognition, medical image segmentation, and image and video processing.

AMJAD ALI received the Ph.D. degree in real-time systems from Gyeongsang National University, South Korea. He is currently an Assistant Professor and the Chairman of the Department of Computer Science and Software Technology, University of Swat. He published several research papers in international journals and conferences.

SHAH NAZIR received the Ph.D. degree in computer science with a specialization in software engineering. He was with the University of Peshawar. He is currently an Assistant Professor and the Head of the Department of Computer Science, University of Swabi. He has several research publications in well-reputed international journals and conference proceedings. His research interests include component-based software engineering, software birthmark, systematic literature review, and decision-making. He is a Reviewer for several journals and conferences.

IJAZ AHAD was born in Lal Qilla Lower Dir, KPK, Pakistan, in 1990. He received the B.S. degree from the University of Malakand, in 2010, the M.S. degree in computer science from Hazara University, Mansehra, in 2012, the B.Ed. degree from Al Khair University, in 2014, and the M.E. degree in software engineering from the School of Information and Software Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2018. He secured 80% marks in M.E. software engineering and received the Best Academic Achievement Award from UESTC. He is currently a Lecturer with the Department of Computer and Software Technology, University of Swat. He has published three articles in various international journals and conference proceedings. His research interests include mobile computing, wireless communication and networking, cloud computing, the Internet engineering, information security, system engineering, and software engineering.

MOHAMMAD SHAHID received the M.S. degree in computer science from Aligarh Muslim University, Aligarh, India. He is currently pursuing the Ph.D. degree with Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology. He has an extraordinary academic, research, and professional experience in computer science. His research interests include data mining, the IOT, machine learning, and medical big data concerned technologies.