IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11471

EFFICIENT RECOGNISE SYSTEM FOR PARKINSON’S DISEASE USING VOCAL RECORDINGS FEATURE SELECTION BASED ON L1-NORM SUPORT VECTOR MECHINE

Eedukondalu Dupati [1], Dr. G. N.R. Prasad [2]

MCA VI SEMESTER, Department of MCA, CBIT (A) – Hyderabad, Telangana, India-500075[1]

Sr. AssistantProfessor, Department of MCA, CBIT (A) – Hyderabad, Telangana, India-500075[2]

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.

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

II. LITERATURE SURVEY

Parkinson’s disease (PD) is a disabling disorder with progressive degeneration of the nigrostriatal pathway that classically impairs motor skills. In the last 15 years, the non-motor symptoms (NMS) of PD became the focus of clinical and scientific interest. Constipation is one of the most frequent and well-known NMS in PD patients [1]. A wide spectrum of

DOI: 10.17148/IJARCCE.2022.11471

prevalence of constipation in Parkinson's patients has been reported, ranging from 7% to 71% among different studies [2,3,4]. Constipation appears through all stages of Parkinson's disease and increases with advancing disease progression [1, 5]. Its impact on the quality of life is no less than motor symptoms. However, there was no suggested or recommended questionnaire/scale for PD constipation and the criteria or definitions of constipation in different studies were heterogeneous [1]. This might contribute to the variation of constipation prevalence, which is not helpful to characterize constipation.Moreover, constipation is supposed as an early, pre-motor manifestation of PD. Recently, one systematic review and meta-analysis proved that people with constipation have a higher risk of developing PD compared with those without. Constipation may precede the onset of Parkinson's cardinal motor symptoms by decade [6]. Several reports demonstrated constipation represents a risk factor for PD (determining a relative risk versus control ranging between 2 and 2.5) [2, 6, 7]. As a premotor symptom, the prevalence of constipation and the time interval between the occurrence of constipation and the motor symptom onset in Chinese Parkinson's patients has not been reported. Therefore, we conducted this cross-sectional investigation in Chinese Parkinson's patients in Shanghai to clarify the prevalence and clinical characteristics of subjective constipation and to evaluate the chronology of motor symptoms and constipation. Meanwhile, related literature would be reviewed to better understand the variation of constipation and its clinical characteristics between Western population and Asians.

III. METHODOLOGY

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

A. DATASET

Dataset used in the research was adopted from the repositoryof the University of Oxford (UO) with collaboration with national center for voice developed by little *et al.* [8] and isavailable at the UC Irvine repository of data mining [23]. Theoriginal research published that feature extraction methodsfor general voice disorders. The voice recordings of 31 people, including 23 people with Parkinson's disease contained16 males and 7 females) and 8 healthy controls (males = 3and females = 5) were deployed in the study. In the datasetable, each column for voice and each row are related to oneof 195 voice recording from an individual subject. Additionally, the people of age from 46 to 85 years with a mean valueof age is 65.8 and standard division 9.8. The main objectiveof this dataset was to classify people with Parkinson's disease from healthy people by fi 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 avowel was recorded for 36 second and total of 195 samples

Algorithm 1 Proposed System

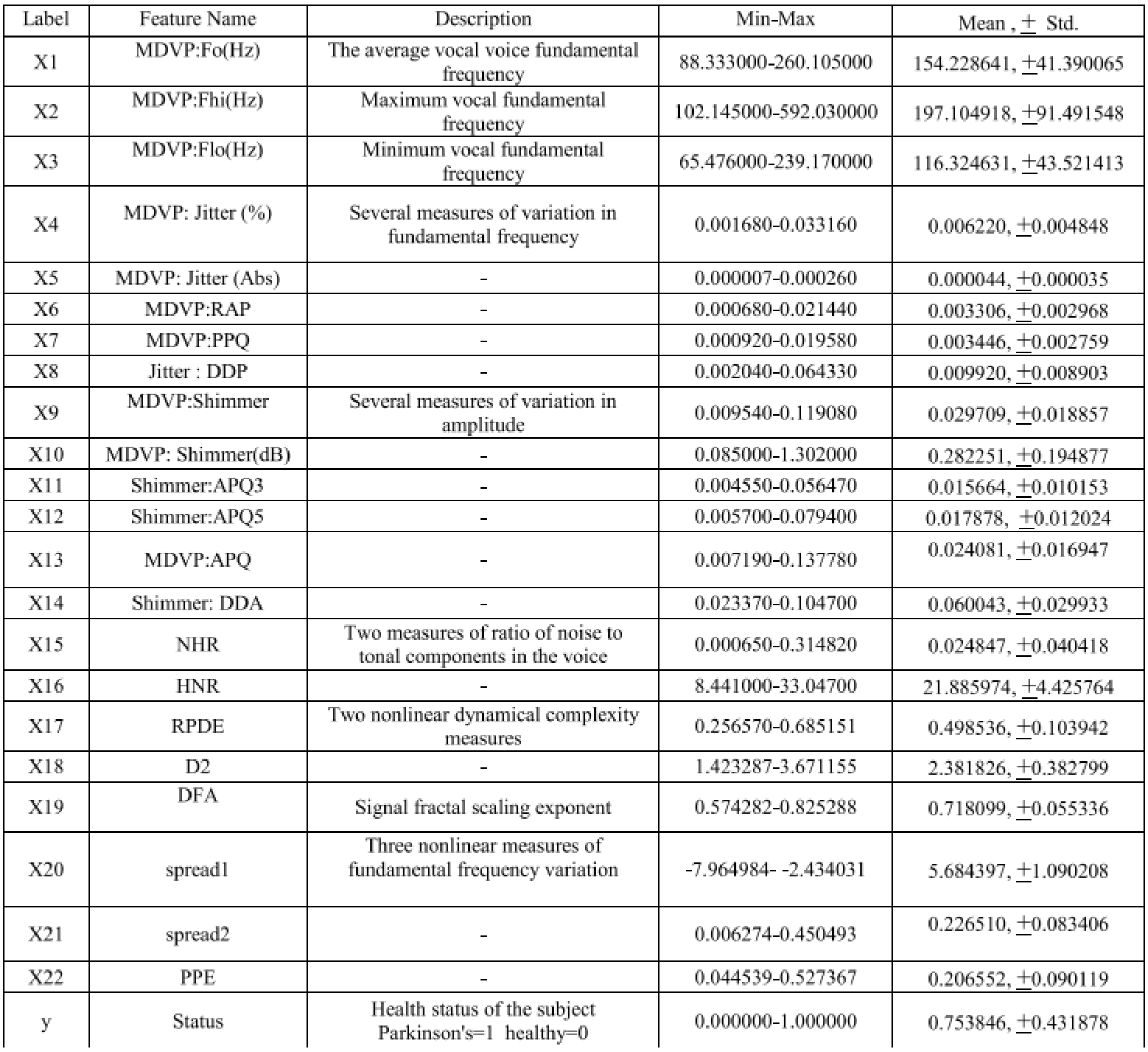
```
Begin
Step1: data preprocessing using standard scalar, and Min
Max scalar on PD dataset;
i.e. $V -$ =
$v-$
$min$
$max$
–
$min$ ($newmax - newmin$) + $newmin$ in Eq(1)
Step2: selected features by L1 –Norm SVM;
Step3: For j = 1: k, performance estimation applied k-fold
cross- validation, where k =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

DOI: 10.17148/IJARCCE.2022.11471

were recorded. The pho nations were recorded in industrial acoustic company sound-treated booth by the micro phone which at distance 8 cm from mouth and microphone was calibrated as presented in [24]. The voice speech signals werestored 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 wereextracted. Thus the extracted dataset size is 195∗23 matrixes.Table 2 shows the 23 features of voice signals of PD dataset.

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 classifiier effectively classifies PD and healthy subjects. Furthermore, the k-fold cross-validation echnique 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 fifive steps, preprocessing of the dataset, features selection, cross-validation, and machine learning classifier performance evaluation. The framework of the proposed classifification system as shown in Fig 1.



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

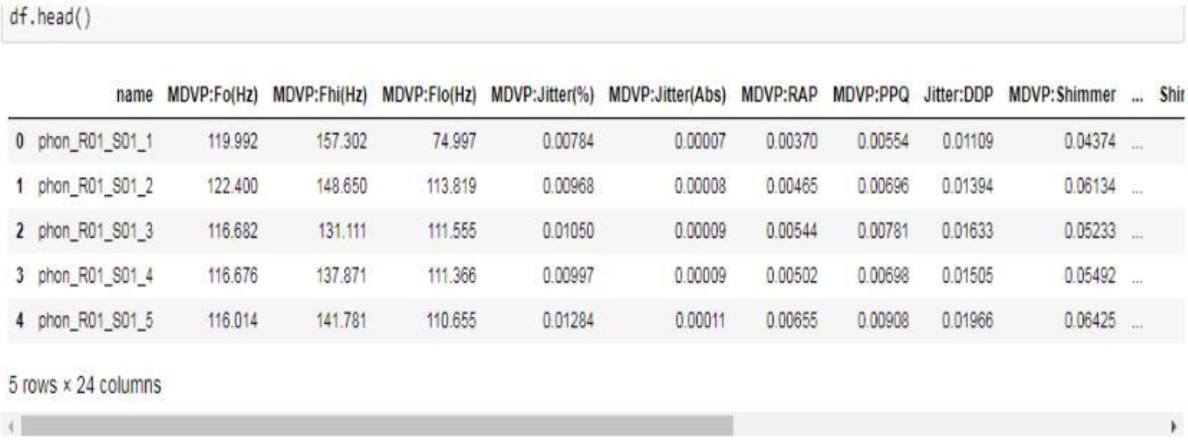
IV. RESULTS

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.

DOI: 10.17148/IJARCCE.2022.11471

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 fifig 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 signifificantly important for prediction of PD.

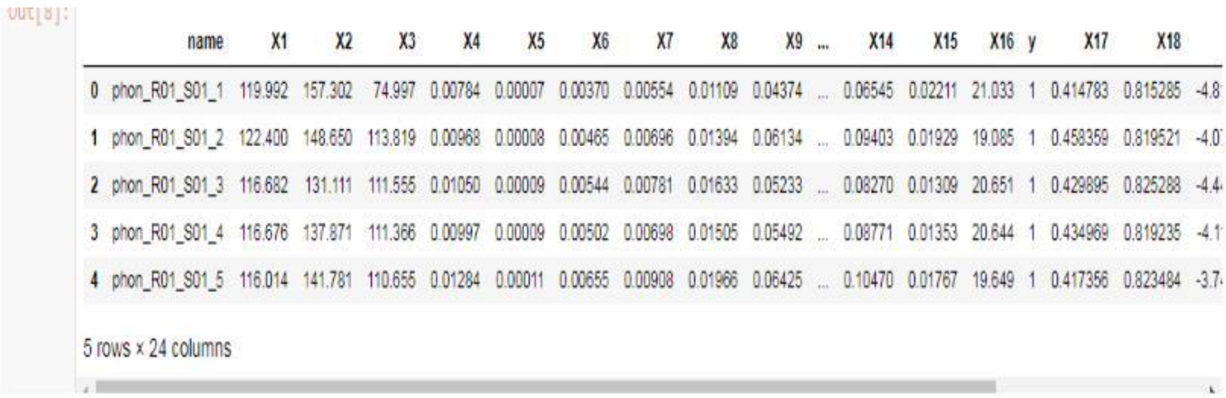


```
df.head()
```

	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	...	Shir
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.00007	0.00370	0.00554	0.01109	0.04374	...	
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	...	
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.00009	0.00544	0.00781	0.01633	0.05233	...	
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	...	
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	...	

5 rows × 24 columns

Fig : Parkinson’s disease Dataset

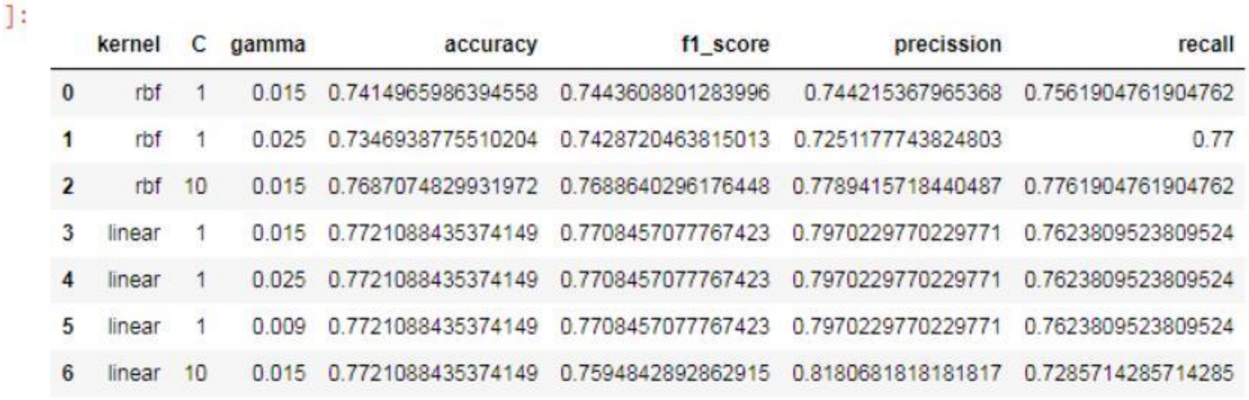


Out[8]:

	name	X1	X2	X3	X4	X5	X6	X7	X8	X9	...	X14	X15	X16	y	X17	X18	
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.00007	0.00370	0.00554	0.01109	0.04374	...	0.06545	0.02211	21.033	1	0.414783	0.815285	-4.8
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	...	0.09403	0.01929	19.085	1	0.458359	0.819521	-4.0
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.00009	0.00544	0.00781	0.01633	0.05233	...	0.08270	0.01309	20.651	1	0.429895	0.825288	-4.4
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	...	0.08771	0.01353	20.644	1	0.434969	0.819235	-4.1
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	...	0.10470	0.01767	19.649	1	0.417356	0.823484	-3.7

5 rows × 24 columns

Fig :After change of the column name

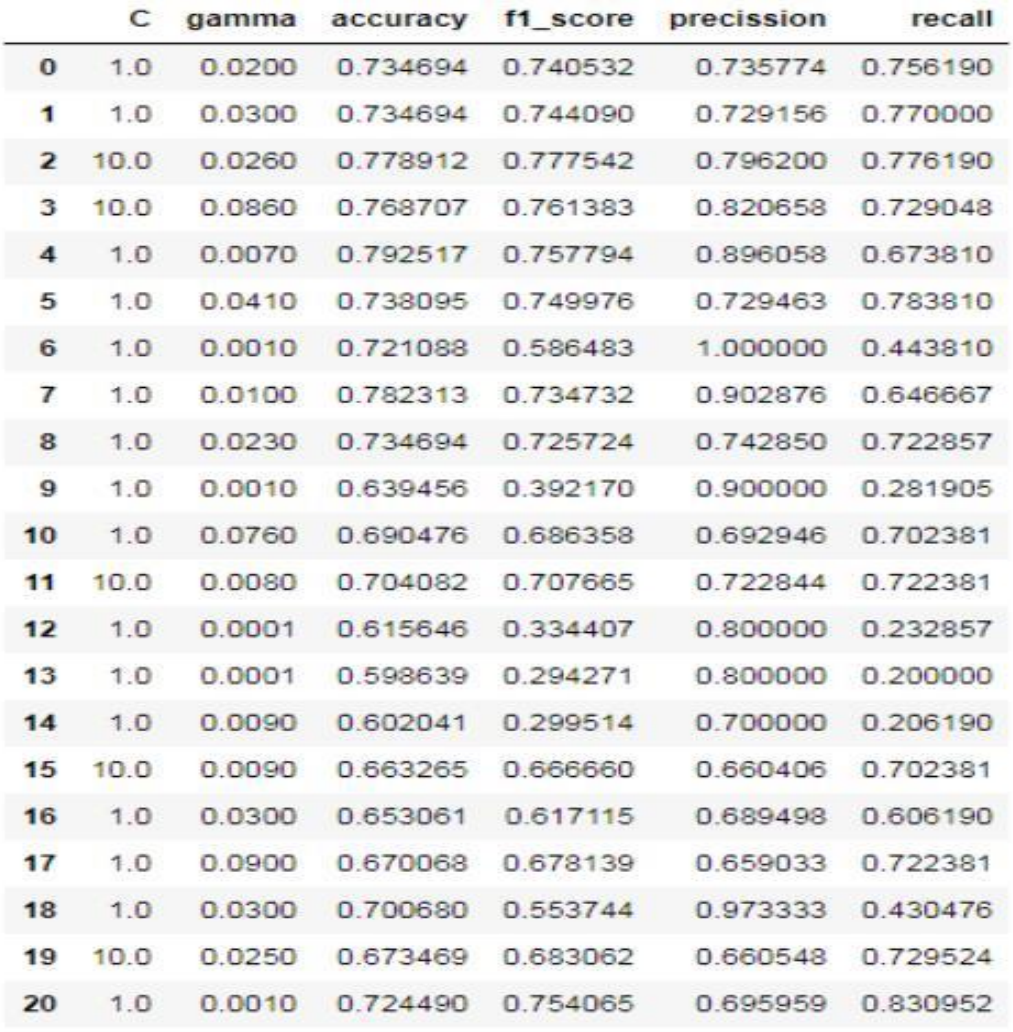


]:

	kernel	C	gamma	accuracy	f1_score	precission	recall
0	rbf	1	0.015	0.7414965986394558	0.7443608801283996	0.744215367965368	0.7561904761904762
1	rbf	1	0.025	0.7346938775510204	0.7428720463815013	0.7251177743824803	0.77
2	rbf	10	0.015	0.7687074829931972	0.7688640296176448	0.7789415718440487	0.7761904761904762
3	linear	1	0.015	0.7721088435374149	0.7708457077767423	0.7970229770229771	0.7623809523809524
4	linear	1	0.025	0.7721088435374149	0.7708457077767423	0.7970229770229771	0.7623809523809524
5	linear	1	0.009	0.7721088435374149	0.7708457077767423	0.7970229770229771	0.7623809523809524
6	linear	10	0.015	0.7721088435374149	0.7594842892862915	0.8180681818181817	0.7285714285714285

Fig :Performance measures with different parameter

DOI: 10.17148/IJARCCE.2022.11471



	C	gamma	accuracy	f1_score	precission	recall
0	1.0	0.0200	0.734694	0.740532	0.735774	0.756190
1	1.0	0.0300	0.734694	0.744090	0.729156	0.770000
2	10.0	0.0260	0.778912	0.777542	0.796200	0.776190
3	10.0	0.0860	0.768707	0.761383	0.820658	0.729048
4	1.0	0.0070	0.792517	0.757794	0.896058	0.673810
5	1.0	0.0410	0.738095	0.749976	0.729463	0.783810
6	1.0	0.0010	0.721088	0.586483	1.000000	0.443810
7	1.0	0.0100	0.782313	0.734732	0.902876	0.646667
8	1.0	0.0230	0.734694	0.725724	0.742850	0.722857
9	1.0	0.0010	0.639456	0.392170	0.900000	0.281905
10	1.0	0.0760	0.690476	0.686358	0.692946	0.702381
11	10.0	0.0080	0.704082	0.707665	0.722844	0.722381
12	1.0	0.0001	0.615646	0.334407	0.800000	0.232857
13	1.0	0.0001	0.598639	0.294271	0.800000	0.200000
14	1.0	0.0090	0.602041	0.299514	0.700000	0.206190
15	10.0	0.0090	0.663265	0.666660	0.660406	0.702381
16	1.0	0.0300	0.653061	0.617115	0.689498	0.606190
17	1.0	0.0900	0.670068	0.678139	0.659033	0.722381
18	1.0	0.0300	0.700680	0.553744	0.973333	0.430476
19	10.0	0.0250	0.673469	0.683062	0.660548	0.729524
20	1.0	0.0010	0.724490	0.754065	0.695959	0.830952

Fig:Performance with feature selection:

V. CONCLUSION

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. The performance of the proposed system is excellent and achieved 99% classification as compared to the classification performances of other proposed studies. 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

REFERENCES

[1] J. R. Duffy, *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. Amsterdam, The Netherlands: Elsevier, 2005.

[2] J.W.Langston,``Parkinson' sdisease:Current and future challenges," *Neuro Toxicology*, vol. 23, pp. 443_450, Oct. 2002.

DOI: 10.17148/IJARCCE.2022.11471

[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 Man- agement 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. 2011.36

[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] • . 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, ``Ef_cient 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 ef_cient 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.

© IJARCCE This work is licensed under a Creative Commons Attribution 4.0 International License