Genetic Analysis with Feature Reduction to Predict the Onset of Parkinson's disease

Ranjith N, Lincy Mathews

Abstract: ***Parkinson is a disease which directly affects the brain cells and certain movement, voice and other disabilities. Hence curable medication is not available in market. The best solution is the early diagnosis to relieve the symptoms of Parkinson's disease affected people. One major concern effecting public is Parkinson's disease (PD). This paper studies the bias of various traditional algorithms on the voice-based data that has various parameters recorded from Parkinson patients and healthy patients. A brief survey of techniques are mentioned for the prediction of Parkinson's diseases is presented. To accomplish this task, identifying the best feature reduction approach was the primary focus. This paper further applies feature reduction techniques using a genetic algorithm for efficient prediction of Parkinson's disease along with machine learning-based approaches. The proposed method also presents higher accuracy in prediction by using this optimal feature reduction technique.***

Keywords : Parkinson's disease, Genetic algorithm, Feature reduction, Support Vector Machine, K-Nearest Neighbor, Naïve Bayes, Speech data.

I. INTRODUCTION

Parkinson's disease is a Deep-Rooted deteriorative epidemic of Central Nervous System [1]. This affects the Motor System that mainly includes gesture, speech and other cerebral parts of the body. Across the globe above 12 million people suffer from Parkinson's disease [2]. Most evident effects from people suffering PD will experience fluctuation in speech. When patients affected by PD speak at a fast rate, he cannot be understood by people around him. The speech deterioration of PD patients is called dysphonia [3]. Age has been seen as one of the most important factor in analyzing the onset of PD [4]. Traditional diagnosis that included physical study and observation of patient did not contribute effectively for the early detection of PD. Machine learning and Artificial Intelligence have effectively contributed by assessing these observations for a better and accurate diagnosis [5].

More than 90% of Parkinson affected population as speech data [6]. Hence, the analysis of speech data can generate meaningful information to confirm the presence of Parkinson's in people [7]. For the extraction of information from the feature datasets data mining techniques [8] plays an important role, which are the most commonly used by the professionals. The paper [9] gives a comprehensive analysis of various data mining techniques for diagnosis of Parkinson and a typical Parkinson disease in its genesis phase. The researchers of [10] had implemented the genetic algorithm and feature reduction using principal component analysis for feature reduction techniques. The traditional classifiers were used in comparison with the performance metrics for different feature sets. To differentiate people with PD and without PD, the voice recordings of the dataset were used for analysis. Both authors, Nazmeh, and Najmeh Samadiani [5] proposed an optimal set of 6 features using genetic algorithm for reduction of Parkinson's disease Ada-Boost & Bagging algorithms were applied for the automation process. The investigations on speech signals of PD patients and using various disparate images has been assessed with gait analysis [11] for the early detection. Gait Pattern was analyzed by employing the technique called silhouette [12]. In this paper [13] a predictive model was used which comprised of SVM. On generating latest set of characteristics with value of proposed method was utilized to get optimal predictions of PD using the voice recording data. To detect few irregularities not identified by medical personal, Meta-analysis [14] and Acoustical analysis method has been utilized by various researchers. In recent times in order to identify people with & without PD, Models have been developed to extract and identify discriminating features specific to a PD patient. For this purpose, voice recordings had been collected from various PD patients [15-17]. The research scholars [18] had discovered the tool to identify people with Parkinson and healthy by using recorded voice signals. From paper [4], author proposed hybrid intelligent system which contains preprocessing of feature dataset by using model based clustering and variants of feature reduction methods such as Head Component Analysis, Linear Discriminant Analysis. Three managed classifiers were used for classification like regression General Neural Network (GRNN), Probabilistic Neural Network (PNN) and Least-Square Support Vector Machine (LS-SVM). To reduce the space occupied by feature vector which had been used for the classification purpose, some studies recently set goal to present latest characteristics while other people had make use of various determination strategies to dispose of the reproduced highlights. Evaluating performance of various evolution methods in selecting optimal features and comparative study of different inspired algorithms was represented in this paper. The authors of this paper [19] employed linear classifiers, probabilistic classifiers and nonlinear classifiers as classification models to identify the Parkinson's disease symptom from fit people by extracting the speech recordings with selected features. By implementing recursive feature elimination algorithm, unique feature selection was implemented.

Revised Manuscript Received on July 10, 2020.

***** Correspondence Author

Ranjith N, Department of Information Science and Engineering, Ramaiah Institute of Technology, Bangalore, India. E-mail: ranjithn06@gmail.com

Lincy Mathews, Assistant Professor, Department of Information Science and Engineering, Ramaiah Institute of Technology, Bangalore, India. E-mail: lincymm99@msrit.edu

The creators thought about the presentation measurements, for example, particularity, positive prescient worth and negative prescient incentive for the component dataset. A set of analysed predictions was helpful for medical people to classify people with Parkinson from fit people from the reference of speech readings.

II. METHODOLOGY

The following section discusses the proposed system for effective identification of Parkinson's affected patients.

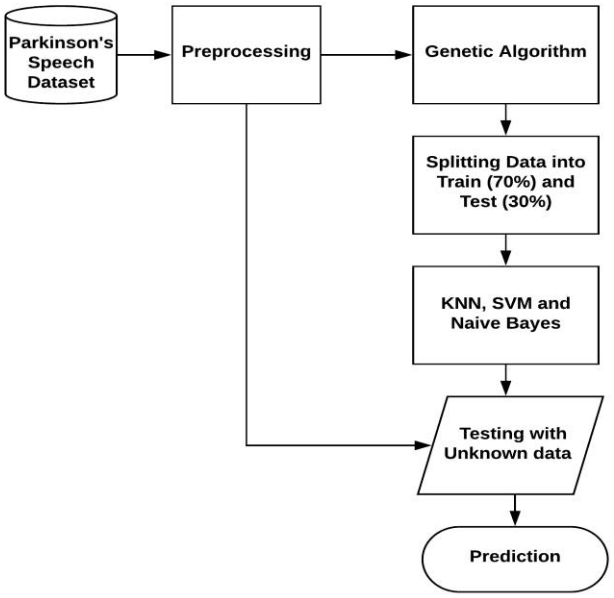

Figure 1. Proposed Model

Figure 1 represents the flow chart of the proposed system. Parkinson's speech data set are based upon mostly numerical features. Initially, the dataset is pre-processed to eliminate the null or negative values. Genetic algorithm is applied further to select the unique features (Feature Reduction). The base classifiers- Probabilistic model: Naïve Bayes , Lazy learner: K-Nearest Neighbor and Support Vector Machine are applied to the original dataset and the reduced dataset.Taking reference of [8], the specialists proposed Spirograph highlights to evaluate the engine work in PD. Different AI and order techniques were applied in examining the nearness of PD [20].Non-Motor features of sleep behavior disorder (RBD) [21] [22] and Cerebrospinal fluid and features of sleep behavior (non-motor) measurements were studied to detect early Parkinson's disease were used by authors of [23]. The models used for evaluation were Random Forest Classifiers, Support Vector Machine (SVM), Naïve Bayes and Boosted trees. In paper [24], the author proposed a technique by analyzing speech, gait and hand writing with start and stop positions. Based on this data, author trained the Convolutional Neural Network to distinguish the individuals with Parkinson's and healthy. The patient's Neurological condition was assessed from different phases like initial, intermediate and advanced. The proposed model is assessed based on the robustness by analyzing speech in different languages Spanish, Czech and German. Based on the final report, the authors classified patients with higher accuracy [25] in the above mentioned languages. The paper begins with the related works concerning the analysis of dysphonia for the study and characterization of PD patients. The methodology and analysis of our works is explained the following sections. The paper will be concluded with the observation and results.

III. FEATURE REDUCTION USING GENETIC ALGORITHM

Genetic algorithm effectively has been considered for its advanced feature selection algorithm. The genetic model is built on natural genetics and biological evolution [26].
One of the components to be analysed is the fitness function. This function decides how best every solution fits for the considered problem. For the given speech data set, let feature set, F={c0,c1,c2, … , c755}. Fitness function in this work is defined as a minimum bias towards the Parkinson's data by the base classifiers. Single point crossover has been implemented among the parent object (Original training data). Random feature is selected for the single point crossover. This genetic operator generates two off-springs (synthetic data) which are further added the training data. All data beyond that point in the feature set is swapped between the two parent data objects. Objects are classified by positional bias. Mutation is further applied to maintain genetic diversity from one generation of data instances to a better-defined data set. For slowing the convergence to the global optimum, mutation operators are used in an attempt to avoid local minima by incurring small changes to the existing population.

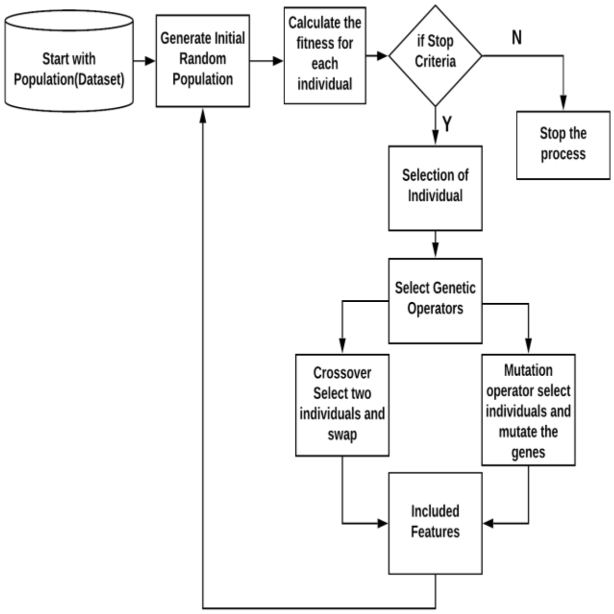

Figure 2. Process of Genetic Model

Figure 2 represents the flow chart of the Genetic Algorithm. Population is selected from the dataset and generated initial population value, fitness of each individual is calculated. If the fitness value is greater than or equal to the initial population then the individual is selected as genetic operator [27]. The crossover and mutation operators are further used to select the unique features from the Parkinson's speech dataset. This enhanced dataset will be used for analysis by various supervised machine learning algorithms.

Base Classifiers:
K-Nearest Neighbour is a non-parametric method which makes a non-existent limit to order the information. At the point when new information focuses comes in, the calculation will attempt to foresee that to the closest of the limit line.
Naïve Bayes algorithm [28] is a supervised and a highly scalable AI calculation which utilizes Bayes hypothesis [29]. This model falls under the probabilistic classifier family. The model expect that highlights are measurably autonomous of one another. Naïve Bayes classifier generates optimum results with the assumption model.
Support Vector Machine [30] is one among the most mainstream regulated AI calculations, which is utilized for arrangement just as Regression issues [31]. The objective is to convert n-dimensional spaces into classes. The new data point should be assigned the right class based on the vectors identified in constructing the hyperplane. These extraordinary cases are called as help vector and henceforth calculation is named as help vector machine. As mentioned above, K-Nearest Neighbour classifier, Support Vector Machine and Naïve Bayes are used as base classifier in this paper.

IV. EXPERIMENTAL RESULTS

The Parkinson's disease speech dataset has been Retrieved from the UC Irvine repository [8]. This dataset has been used for testing the proposed system.

Table 1. Unique features of Speech Dataset

Sl no	Feature Name	Description
1	MDVP:Fo(HZ)	Normal vocal voice key recurrence
2	MDVP-Fhi (HZ)	Most extreme vocal voice basic recurrence
3	MDVP-Flo (HZ)	Least vocal voice principal recurrence
4	MDVP:Jitter(%)	A few proportion of variety in major recurrence
5	MDVP:Shimmer	A few proportion of variety in central abundancy
6	NHR	Duo measure of ratio of nose to tonal components in the voice
7	RPDE	Two nonlinear dynamical intricacy measures
8	DFA	Signal fractal scaling type
9	Spread1	Three nonlinear proportions of essential recurrence variety
10	Status	Wellbeing status of the subject Parkinson's=1 healthy=0

Table 1 represents the unique features of speech samples from the dataset along with description. All the attributes are treated as numerical attributes before experimentation [23].

V.OBSERVATIONS

Table 2. Results for Parkinson's Speech Dataset before implementing Genetic Algorithm

SL NO	Classification Algorithm	Accuracy	Recall	Precision
1	K-Nearest Neighbor	69%	81%	78%
2	Support Vector Machine	72%	93%	75%
3	Naïve Bayes	70%	78%	80%

Table 2 represents the results of Naïve Bayes order K-NN and SVM algorithms with accuracy, precision and recall values before feature reduction. The K-Nearest Neighbour achieved an accuracy of 69%, recall of 81% and precision of 78%. An accuracy of 72% with a similar precision value was attained by Support Vector Machine. Naïve Bayes achieved an accuracy of 70%, recall of 78% and precision of 80%. The highest accuracy was obtained with Support Vector Machine model.

Table 3. Results for Parkinson's Speech Dataset with Feature Reduction

Sl No	Classification Algorithm	Accuracy	Recall	Precision
1	K-Nearest Neighbor	77.97%	98%	79%
2	Support Vector Machine	78.85%	100%	79%
3	Naïve Bayes	78.85%	100%	79%

The performance values with feature reduction are as shown in Table 3. K-Nearest Neighbor achieved an accuracy of 77.97%, recall of 98% and precision of 79%. Support Vector Machine showed an accuracy of 78.85%, recall of 100% and precision of 79%. Naïve Bayes scored an accuracy of 78.85%, recall of 100% and precision of 79%. Support Vector Machine and Naïve Bayes classification algorithms predicted with the highest accuracy on Parkinson's speech dataset. The above readings showed the effect of applying genetic approach [32-34] in selecting the unique features for better prediction of Parkinson's effected candidate.

VI. CONCLUSION

This paper studied the effects of feature reduction technique using genetic algorithm along with supervised AI calculations for better expectation of Parkinson's. From the observations, genetic algorithm with machine learning models were found to be effective in predicting the presence of Parkinson's symptoms. The effects of non-conversion of nominal attributes with the proposed system needs to be evaluated. Heuristic Search algorithm needs to be analyzed to understand its efficiency for early detection of PD.

REFERENCES

1. Jifeng Guo,Hongxu Pan,,Li Shu,,Xinxiang Yan, ,Qiying Sun,,Beisha Tang,, Dongxiao Liang,Qian Xu,,, "Genetic Impact on Clinical Features in Parkinson's Disease: A Study on SNCA-rs11931074,," pp. 2-4, 2018.
2. K.R, Chaudhuri,,C.Marrasand,,, "Nonmotor features of Parkinson's disease subtypes,," vol. 31, pp. 1095-1102, 2016.
3. Y.Wang et al,, Z.Liu,, J.Guo,, "Lack of association between IL-10 and IL-18 gene promoter polymorphisms and Parkinson's disease with cognitive impairement in a Chinese population,,," p. 6, 2016.
4. R.Sindhu,, M.Hariharan,, K.Polat,,, "A new hybrid intelligent system for accurate detection of Parkinson's deisease,,," vol. 113, pp. 904-913, 2014.
5. Najmeh-Samadiani,, Fayyazifar, Najmeh,, "Parkinson's disease detection using Ensemble techniques and Genetic Algorithm,,," Artificial Intelligence and Signal Processing Conference,, pp. 162-165, 2017.
6. Jinse Park,Ki-Won Choi,, Mangal Sain,, Satyabrata Aich,, Hee-Cheol Kim,,, "A Mixed Classification Approach for the Prediction of Parkinson's disease using Nonlinear Feature Selection Technique based on the Voice Recording,," pp. 959-962, 2017.
7. Z.H. Liu et al,, K.Li, B.S. Tang,,, "LRRK2 A419V variant is a rsk factor for Parkinson's disease in Asian populations,,," vol. 36, pp. pp. 2908 e11-2908 e15, 2015.
8. Aleksander,, Sadikov, et al,,, "Feasibility of spirography features for objective assessment of motor function in Parkinson's disease,,," Artificial Intelligence in Medicine,, vol. 13, pp. 54-62,, 2017.
9. D.Karthiga, Dr.P.Sumithra,, "A Survey of Predicting Parkinson's & A typical parkinson's disease in the Primordial stage by using classification techniques in Data Mining,,," vol. 5, pp. 794-797, 2018.
10. Ahmed Abdulhakim Al-Absi,, Hee-Cheol Kim, Kim younga, Kueh Lee Hui, Mangal Sain, Satyabrata Aich,, , "A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson's Disease,," vol. 7, pp. 1116-1121, 2018.
11. I.L.Guan, , H.Lee, I.Lee,,, "Video analysis of Human Gait and Posture to determine Neurological disorders,," vol. 1, 2008.
12. C.Harris, M.S.Nixon,, P.Huang,, , "Human gait recognition in canonical space using temporal templates,," vol. 146, pp. 93-100, 1999.
13. Mujammed Hammad Memon, Shah Nazir,, Tanvir Ahmad, Asad Malik, Jian Ping Li, Amin ul Haq, Jalaluddin Khan, Mohammad Shahid, Ijaz Ahad, Amjad Ali,,, "Feature Selection based on L1-Norm Support Vector Machine ans Effective Recognition System for Parkinson's Disease Using Voice Recordings,,," vol. 7, pp. 37718-37734, 2019.
14. Minhan Yi, Jifeng Guo, Yuan Zhang, Beisha Tang, Qiying Sun, Xinxiang Yan, Qian Xu, Xun Zhou,,, "Genetic Analysis of LRRK2R1628P in Parkinson's Disease in Asian Populations,," pp. 2-6, 2017.
15. Y. Campos-Roca, L. Naranjo, J. Martin, C. Jperez,,, " Addressing voice recording replications for Parkinson's disease detection,," vol. 46, pp. 286-292, 2016.
16. J. Martin, L. Naranjo, C. J. Perez, Y. Campos-Roca,,, "A latent variable-based Bayesian Regression to address recording replications in Parkinson's disease,," pp. 1447-1451, 2014.
17. R. Cmejla, M. Novotny, E. Ruzicka, J. Rusz,, , "Automatic Evaluation of Articuatory disorders in Parkinson's disease,,," vol. 22, pp. 1366-1378, 2014.
18. F. Karabiber,, I.Canturk,, "A Machine learning system for the diagnosis of Parkinson's disease from speech signals and its application to multiple speech signal types,.," vol. 41, pp. 5049-5059, 2016.
19. Prashant, Shrivastava, et al,,, "A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease,,," vol. 139, pp. 171-179, 2017.
20. C.B. Schneider, A.Storch, L.Klingelhofer et al,,, "Quantitative assessment of non-motor fluctuations in Parkinson's disease using the Non-Motor Symptoms ,,," vol. 122, pp. 1673-1684, 2015.
21. L.M. Shulman, R.von Coellnand ,, "Clinical subtypes and genetic heterogeneity: of lumping and splitting in Parkinson disease,,," vol. 29, pp. 727-734, 2016.
22. Dr.Hariganesh S,Gracy Annamary S,, "A Survey of Parkinson's Disease Using Data Mining Algorithms,," vol. 5, pp. 4943-4944, 2014.
23. R. et al, Prashanth,, "High-accuracy detection of early Parkinson's disease through multimodal features and machine learning,,," pp. 13-21, 2016.
24. Tomas Arias-Vergara, Juan Camilo Vasquez-Correa, Bjorn Eskofier, J.R. Orozco-Arroyave, Elmar Noth, Jochen Klucken, , "Multimodal Assessment of Parkinson's disease: A Deep Learning Approach,,," vol. 23, pp. 1618-1630, 2019.
25. D. Dougherty, J.Westin, K.Taha,,, "Classification of speech intelligibility in Parkinson's disease,," vol. 34, pp. 35-45, 2015.
26. M.Venkateswara Rao, Tarigoppula V.S Sriram, DSVGK Kaladhar, G.V.SatyaNarayana, , "A Comparative And Prediction Analysis for the Diagnosis of Parkinsons Disease using Data Mining Techniques on Voice Datasets,,," vol. 11, 2016.
27. R.B.Postuma, D.Berg, C.H.Adler et al,, , "MDS research criteria for prodromal Parkinson's disease,,," vol. 30, pp. 1600-1611, 2015.
28. P.Porwik, W.Froelich, K.Wrobel,,, "Diagnosis of Parkinson's disease using speech samples and threshold-based classification,,," pp. 1358-1363, 2015.
29. J.Martin, Y.Campos-Roca, L.Naranjo, C.J.Perez,,, "A two stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications,,," vol. 142, pp. 147-156, 2017.

30. 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,,” pp. 101-106, 2018.
31. B.S. Tang, K.Fan, Y.Q. Wang et al,,, “The GBA, DYRK1A and MS4A6A polymorphisms influence the age at onset of Chinese Parkinson Patents,,” vol. 621, pp. 133-136, 2016.
32. B.S.Tang, Y.Yang, L.Weng et al, , “Genetic identification is critical for the diagnosis of parkinsonism: a Chinese pedigree with early onset of parkinsonism,;,” p. 10, 2015.
33. N.Bregman, A.Thaler, T.Gurevich et al, , “Parkinson's disease phenotype is influenced by the severity of the mutations in the GBA gene,,” vol. 45, pp. 45-49, 2018.
34. C.Y.McLean, M.A.Nalls, J.Rick et al, , “Diagnosis of Parkinson's disease on the basics of clinical and genetic classification: A populaion-based modelling study,,” vol. 14, pp. 1002-1009, 2015.

AUTHORS PROFILE

Ranjith N pursuing M. Tech on Software Engineering in ISE department from M S Ramaiah Institute of Technology. Interested in subjects related Data Mining techniques and machine learning. Worked on web development domain for project. Working as Associate Engineer in Unisys Organization.

Lincy Mathews pursuing PhD at Vellore Institute of Technology, Vellore, India. Her interests include Pattern recognition, Data Mining, Data Science. She is working as Asst Prof in M S Ramaiah Institute of Technology, Bangalore. She has published papers in various reputed journals and conferences. She has coordinated projects for the differently abled students.