

Effective Predictor Model for Parkinson's Disease Using Machine Learning

P Pradeep¹, Kamalakannan J²

¹ Research Scholar, Vellore Institute of Technology University, Vellore

⁵ Sr Associate Professor, SITE, Vellore Institute of Technology University, Vellore

e-mail: pradeep.pathi9@gmail.com

*Corresponding Author: pradeep.pathi9@gmail.com

<https://doi.org/10.22362/ijcert/2023/v10/i04/v10i0410>

Received: 25/03/2023,

Revised: 17/04/2023,

Accepted: 21/04/2023

Published: 15/05/2023

Abstract: - Parkinson's disease is a neural degenerative disease, in which the patients' faces various critical neurological disorders. Thus, the earlier prediction of PD helps to enhance the patients' life. The prediction of PD in the earlier stage is extremely complex and it consumes huge time. Therefore, effectual and appropriate prediction of PD is measured to a challenging factor for the health care experts and practitioners. To deal with this issue and to accurately predict the PD in earlier stage, this work concentrates on machine learning approaches for designing a predictor system. For developing the anticipated model, L1-norm based Genetic algorithm (L1-GA) is applied for predicting PD in the earlier stage. This L1-GA is utilized for selecting the highly related features for accurate classification of Parkinson disease. This L1-GA produces newer feature subset from PD dataset based on its feature weight value. For validation purpose, this work considers 10-fold cross validation (CV) is used. Also, metrics like sensitivity, accuracy, precision, specificity, F1 score and execution time are evaluated. The inputs are taken from The PD dataset which is available online for preceding the feature selection process. The optimal accuracy attained with these newly selected sub-sets are considered for further computation. The experimental findings determine that this study recommends that L1-GA provides better contribution towards PD feature selection and to predict PD in earlier stage. In recent times, Clinical Decision Support System (CDSS) plays an essential role for assisting PD recognition. As well, the anticipated model lays a bridge to fill the gap encountered in feature selection using the available data. The anticipated model gives better trade-off in contrast to prevailing approaches.

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

1. Introduction

Parkinson's disease is a neuron degenerative long term and age dependent disease which affects central

nervous system and influences motor system of the individuals. Specifically the inference and frequency of Parkinson disease over 500 people is estimated with in the 2-4 persons. the occurrence of Parkinson disease over the

individual is assumed to be increase with age. Some common symptoms of PD include walking, shaking, movement, rigidity and some issues over cognitive behavior and thinking. The constant disease progression it turns to be another condition known as dementia. As Well, depression and anxiety are some common issues related to the people under this condition. Anti-Parkinson Medication Levodopa merged with Dopamine Agonists proves that there are some positive responses towards the treatment of PD Patients. It is classified as behavior, cognition mood and thought. The Prediction of PD is made easier in numerous ways. Examination of medical history and neurological analysis are carried out by physics for preliminary evolution of condition. Imaging technologies like computer tomography and Magnetic Resonance Imaging are generally adopted for Predicting PD. MRI is comparatively efficient than CT for appropriate disease prediction.

Diffusion Technique MRU is thoughtful for differentiating PD over the added symptoms of PD. The

arrival for some sophisticated computational diagnosis system, various diseases are analyzed and adopt the intelligent predictor model for enhancing the efficiency. Domestic and industrial applications collect huge data over the days for effectual analysis. some other applications like earth simulator ,protein folding models, weather forecasting. In case if the disease is diagnosed it turns to be a life threatening factor. Similarly the prior diagnoses also drastically enhance the affected condition of every individual rapidly. This investigation attempts to determine the discriminative patterns among the PD control and affected case from vocal features over dataset with operative learning models. PD based dataset are attained from the online available UCI repositories. it is preprocessed missing values and noise. The outcome relies over optimal significance of PD from the anticipated framework. Moreover it enhances the significance of diverse computational model for health care sectors for superior disease prediction.

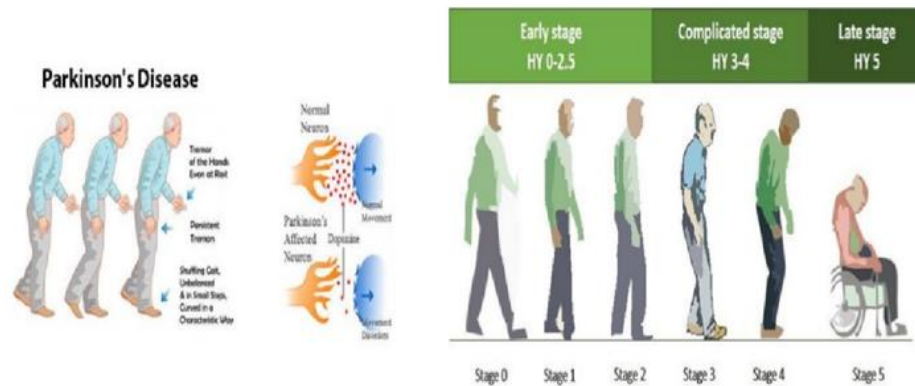


Figure 1: Generic view of PD and its stages

1.1 Types of Parkinson's disease:

Parkinsonism is a term for a disease that has similar symptoms to Parkinson's disease. Mainly includes the following types:

1.1.1 Idiopathic Parkinson's Disease

It is the most common type of Parkinsonism. Idiopathic Parkinson's are tremor, rigidity (stiffness) and slowness of movement. Idiopathic means that the cause is unknown. The main symptoms of

1.1.2 Vascular Parkinsonism

Vascular parkinsonism (also known as arteriosclerotic parkinsonism) affects people with restricted blood supply to the brain - usually older people who have health issues such as diabetes. The main symptoms are difficulty walking, urinary incontinence and memory problems.

1.1.3 Drug-Induced Parkinsonism

A small percentage (about 7%) of people who diagnosed with Parkinson's disease develop symptoms after receiving specific medications.

1.1.4 Multiple system atrophy (MSA)

Multiple system atrophy (MSA) is an adult-onset, sporadic neurodegenerative disease. Its clinical manifestations include autonomic nerve dysfunction, Parkinson's syndrome, cerebellar ataxia and pyramidal tract signs.

1.1.5 Progressive supranuclear palsy (PSP)

Progressive supranuclear palsy affects eye movement, balance, movement, speech and swallowing.

2. Related Work

The prevailing approaches over Parkinson's disease prediction with various available data set. It is explained by various investigators and the researchers pretend to attain better out comes for assisting clinical health care systems. In the past few decades ML Based disease categorization is extensively applied in diverse medical applications and attained notable outcomes. he used various recordings and states language pronunciation comprises of appropriate information for categorizing and attaining better accuracy. it evaluates decomposition and inherent based features from vowels from two diverse data set.

Gillies et al., evaluated the aging parameters in his work. The author considers prosody, articulation features along with his age factors from speech vowels. the author accounts for gender parameters and evaluated the age factor which plays a crucial role in PD Classification. The Investigator modeled multi class and binary SVM and evaluated the outcomes with NN by showing 96% accuracy. Chen provided a novel method for categorizing speech signal towards PD/health patients. The author anticipated a phonological feature based model with monologues speech signal of words and text. The investigator evaluates this model with essential features of PD detection.

O.B Tysnes evaluated an enhanced version of FDA for predicting PD. He Considered Prosody articulation, Phonation and intelligibility features with

Spanish vowels word and sentences. the author considers vowels recordings with online available datasets. He integrates statistical pooling for improving the features and applies gait physionet data set for diagnosing PD.

3. Methodology

This section discusses about the anticipated method used for extracting features from PD for enhancing the classification accuracy. Here L1-Norm is used for performing latent representation of features and the effectual feature extraction analysis is done with Genetic Algorithm. The extracted features are fed as input to the classifier model.

3.1 Dataset Description

The data which is used in this investigation is gathered from 188 patients.(81 women and 107 men) with PD . the age of the patients range from 33 to 87 from the department of Neurology in Istanbul University. Similarly the control groups comprised of 64 healthy individual (41 women and 23 men). During the process of data collection the microphone values is set as 44.1 KHZ .It is based on physician's examination with phonation of vowels collected from every individual with three repetitions. Various speech signal processing algorithms like wavelet transform, Time Frequency Features, Mel Frequency Cepstral Coefficients.

PREDICTOR MODEL

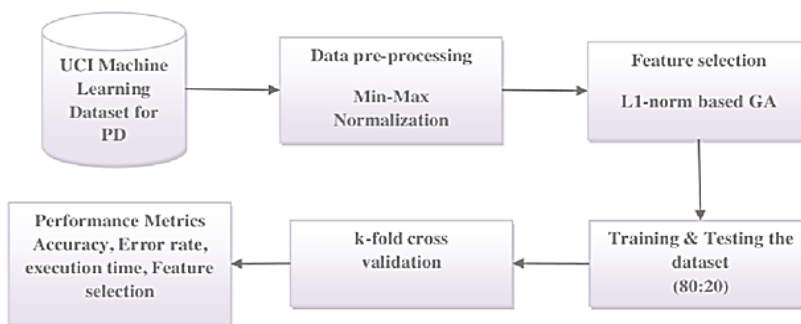


Figure 2 Flow diagram of Proposed Model

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

3.2 Pre-Processing

The appropriate representation of data is extremely an sentimental step before performing classification with ML Approaches. The Preprocessing Technique includes standard scalar, removing of missing values, Min Max Scalar are applied to dataset. the Min-Max normalization method, a feature's values are scaled to fall between 0 and 1. This is accomplished by dividing

each value by the feature's range after subtracting the feature's minimum value from each value.

Table 1 Data Set Descriptions:

Characteristics	Multi-Variate
Attribute Characteristics	Real,Integer
Associated Tasks	Classification
Total Instances	756
Total Attributes	754
Missing Values	N/A

3.3 Genetic Operators with regularization Process:

In evaluation process with regularization process some standard GA operators like crossover, fitness selection and uniform mutation operators are used.

3.4 Cross over

Initially choose the parents randomly from the available population for further breeding process. Subsequently cross over operation is applied to cross over population respectively for producing off spring with inherent characteristics of the parents.

Algorithm:

Step-1.Begin

Step-2.Preprocessing data with min-max normalization

4. Results and Analysis

The Simulation is performed the accuracy of feature selection. The value are mathematically expressed as in

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} * 100$$

Here TP indicate for TRUE Positive

TN indicate for TRUE Negative

FP indicates for FALSE Positive

FN indicates for FALSE Negative

The model accuracy relies overall performances of the L1-Norm based GA. To validate the performance of the diagnostic system various experimentations is related based on feature selection process using L-1 norm GA Algorithm. the performance of the GA over UCI ML repository for PD over the available features performed using k-fold CV where k=2. The experimental outcomes of the anticipated model are compared with various approaches with the experimental results various solution are attained and provided a conclusion towards the research. This model pretends to fill the gap to promote feature selection and classification by an appropriate predictor model.

Step-3 chooses the most influencing feature with the L1-norm based on GA.

Step-4. Initialize population size crossover probability mutation probability and stopping criteria.

Step-5. Perform K-fold CV with available training and testing set.

Step-6.Train Classifier and validate selection with test sets and achieve best set feature selection parameters.

Step-7.Evaluate objective function

Step-8 use genetic operators for regularization.

Step-9 Analyze mutation probability with pm=0.1

Step-10. Generate next selection process with available offspring and parents.

Step-11.Measure the performance of predication model and testing set with candidate process.

Step-12.end the process.

3.4. Mutation

This operator facilities population diversity and larger exploration of search space. The randomly chosen penalized control parameter lambda and alpha with mutation probability are mutating to chosen chromosome

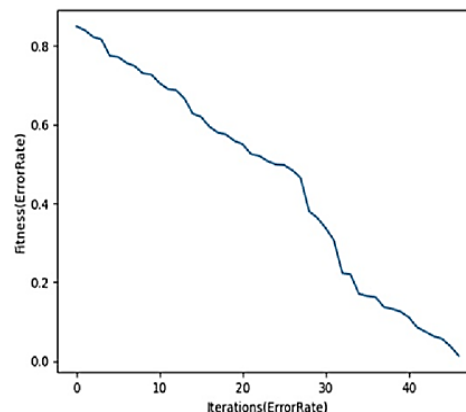


Figure 4 Error Rate Computations

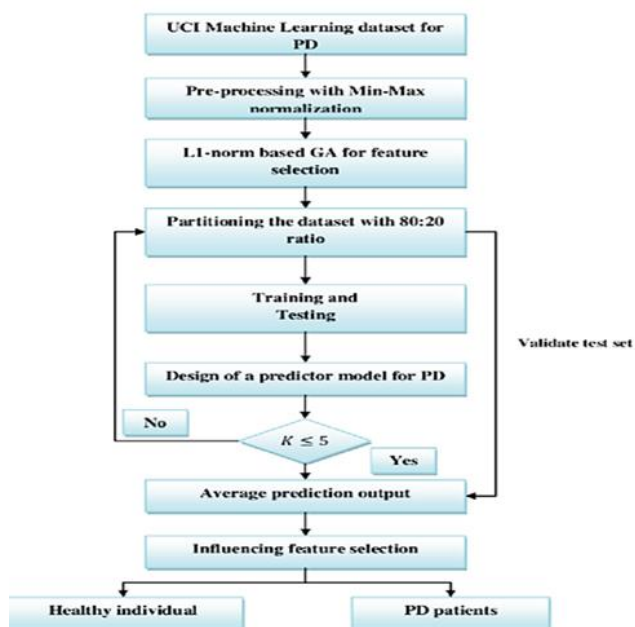


Figure 3 Flow Chart of L1-norm based GA Model

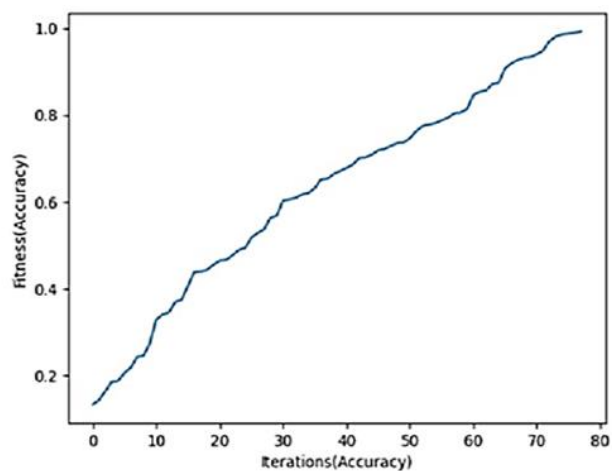


Figure 5 Fitness Computation

5. Conclusion:

An efficient diagnosis system is modeled for predicting PD. In this predictor model a ML Approaches is used for predicting PD using the feature subset selection. Here L1-Norm based GA is used for feature selection with suitable and appropriate feature for accurate prediction of PD. This method generates newer feature subset with reduced complexity. Hence an effectual feature selection process is used for handling critical factors and to choose better feature from the feature space for attaining optimal classifier performance. Feature selection approach discriminate the performance of anticipated model by

differs rent the healthy persons from PD. As well as the anticipated model fills the research gap among the feature selection with voice recordings by appropriate experimental model. Reduced feature subset is attained with L-1 norm based GA which shows highly influencing features that predicts PD more accurately when compared to unique feature spaces. GA Performance is measured for reducing the feature subset compared to other models. In future, feature selection can be performed in a hybridized manner for attaining better classification model and it is used for further enhancement in performance of diagnosis system for PD diagnosis.

References

- [1] Abd Ghani, M. K, Mohammed, M. A., Arunkumar, N. Mostafa, S. A., Ibrahim, D. A., Abdullah, M. K., Jaber, M. M., Abdulhay, Enas, Ramirez-Gonzalez, Gustavo, & Burhanuddin, M.A. (2018) Decisionlevel fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques. *Neural Computing and Applications*.
- [2] Arjmandi, M. K., & Pooyan, M. (2012). An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. *Biomedical Signal Processing and Control*, 7(1), 3–19.
- [3] Can, M. (2013). Neural networks to diagnose the Parkinson's disease. *South East Europe Journal of Soft Computing*, 2(1).
- [4] Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications*, 37(2),1568–1572.
- [5] Davie, C. A. (2008). A review of Parkinson's disease. *British Medical Bulletin*, 86(1), 109–127.
- [6] Doan, S., & Horiguchi, S. (2004). An agent-based approach to feature selection in text categorization. In *Proceedings of 2nd International Conference on Autonomous Robot and Agent* (pp. 362–366).
- [7] Farahnakian, F., & Mozayani, N. (2009). Evaluating feature selection techniques in simulated soccer multi agents system. In *Advanced Computer Control, 2009. ICACC'09. International Conference on* (pp. 107–110). IEEE.
- [8] Ganapathy, S., Kulothungan, K., Muthurajkumar, S., Vijayalakshmi, M., Yogesh, P., & Kannan, A. (2013). Intelligent feature selection and classification techniques for intrusion detection in networks: A survey. *EURASIP Journal on Wireless Communications and Networking*, 2013 (1), 271.
- [9] Gnanapriya, S., Suganya, R., Devi, G. S., & Kumar, M. S. (2010). Data mining concepts and techniques. *Data Mining and Knowledge Engineering*, 2(9), 256–263.
- [10] Gupta, D., Julka, A., Jain, S., Aggarwal, T., Khanna, A., Arunkumar, N., & de Albuquerque, V. H. C. (2018). Optimized cuttlefish algorithm for diagnosis of

Parkinson's disease. *Cognitive Systems Research.*, 52, 36–48.

[8] Gupta, D., Sundaram, S., Khanna, A., Hassanien, A. E., & de Albuquerque, V. H. C. (2018). Improved diagnosis of Parkinson's disease using optimized crow search algorithm. *Computers & Electrical Engineering*, 68, 412–424.

[9] Hariharan, M., Polat, K., & Sindhu, R. (2014). A new hybrid intelligent system for accurate detection of Parkinson's disease. *Computer Methods and Programs in Biomedicine*, 113(3), 904–913.

[10] Kaya, E., Findik, O., Babaoglu, I., & Arslan, A. (2011). Effect of discretization method on the diagnosis of Parkinson's disease. *International Journal of Innovative Computing, Information and Control*, 7, 4669–4678.

[11] M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 56(4), 1015–1022.

[12] Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMedical Engineering OnLine.*

[13] Mandal, I., & Sairam, N. (2013). Accurate telemonitoring of Parkinson's disease diagnosis using robust inference system. *International Journal of Medical Informatics*, 82(5), 359–377.

[14] Mohammed, M. A., Al-Khateeb, B., Rashid, A. N., Ibrahim, D. A., Ghani, M. K. A., & Mostafa, S. A. (2018). Neural network and multifractal dimension features for

breast cancer classification from ultrasound images. *Computers & Electrical Engineering.*, 70, 871–882.

[15] Mohammed, M. A., Ghani, M. K. A., Arunkumar, N., Mostafa, S. A., Abdullah, M. K., & Burhanuddin, M. A. (2018). Trainable model for segmenting and identifying Nasopharyngeal carcinoma. *Computers & Electrical Engineering*, 71, 372–387.

[16] Mostafa, S. A., Ahmad, M. S., Mustapha, A., & Mohammed, M. A. (2017). Formulating layered adjustable autonomy for unmanned aerial vehicles. *International Journal of Intelligent Computing and Cybernetics*, 10(4), 430–450.

[17] Gayathri Nagasubramanian, Muthuramalingam Sankayya, "Parkinson Data Analysis and Prediction System Using Multi-Variant Stacked Auto Encoder", *IEEE Access*, 2018

[18] Laiba Zahid1, Muazzam Maqsood, "A Spectrogram Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson's Disease", *IEEE Access*, 2017.

[19] Haijun Lei, Zhongwei Huang, Feng Zhou, "Parkinson's Disease Diagnosis via Joint Learning from Multiple Modalities and Relations", *Journal of Biomedical and Health Informatics*, 2018.

[20] Yuqian Zhang, "Prediction of Freezing of Gait in Patients with Parkinson's Disease by Identifying Impaired Gait Patterns", *IEEE Transactions On Neural Systems and Rehabilitation Engineering*, Vol. 28, No. 3, March 2020.

[21] Yanhao Xiong, "Deep Feature Extraction from the Vocal Vectors Using Sparse Auto-encoders for Parkinson's Classification", *IEEE Access*, 2020.