A Machine Learning Based Parkinson Prediction System – A Comprehensive Review

Anisha C.D[1], Dr. K.G. Saranya[2],

Research Scholar[1], Assistant Professor (S. Gr)[2],

Dept of CSE, PSG college of Technology, Coimbatore, India.

E-mail: [1]ani.c.dass@gmail.com, [2]kgs.cse@psgtech.ac.in

Abstract

Parkinson's Disorder (PD) is a progressive neurodegenerative disorder, which is incurable. Diagnosis of PD at an early stage aids in the delay of the progression. It requires an accurate and robust system to provide early diagnosis of PD. Machine Learning (ML) based techniques help in developing of PD diagnosis system. This study presents a complete review of the various machine learning techniques along with their working principle that helps for the development of PD diagnosis system. This research highlights the summary of methodologies and also presents a generic framework for the PD diagnosis based on voice signals.

Keywords: Parkinson Disease, artificial neural network, random forest, machine learning, support vector machine, stacking.

1. Introduction

Parkinson Disorder (PD) is a neurodegenerative disorder which is mainly associated with the dopamine receptors. The main generic symptoms related to PD are stiffness, motor related problems and non-motor related problems. There are specific symptoms which are unique for each patient. Speech related issues is one of the common non-motor related problems. PD is a progressive disorder due to the occurrence of death of dopamine receptors. The main diagnosis test is connected with clinical conventional methods which is based on patient history analysis and through examination.

The clinical terminologies associated with symptoms of PD are bradykinesia (slower movement), hypokinesia (absence of movement), rigidity of neck, shoulder and wrist and tremor. The main causes of PD are drugs, multiple cerebral infraction, multiple system atrophy and progressive super nuclear paralysis. PD is characterized by one of the important symptoms known as vocal disorders at the initial stages. [1]

PD symptoms is also associated with mental illness namely depression and dementia along with physiological symptoms. The blend of psychiatric and physiological problems will eventually affect the autonomy of the individual which impacts the living of the patient.

Artificial Intelligence (AI) based diagnosis plays a prominent role in early diagnosis of PD thereby delaying the progression of the disorder. Speech is an important biomarker which acts as an important data to build AI models for PD. Machine Learning (ML) algorithms which is the subpart of AI, helps in the diagnosis of PD using speech parameters; it clearly distinguishes PD patients from healthy subjects.

This study focuses on discussing different ML models along with their properties and working principle, and it also presents various ML models used in PD prediction or classification.

2. Machine Learning Methods

1. Artificial Neural Networks (ANN)

ANN is a ML algorithm which mimics the human brain neurons. Input layer, hidden layer and the output layer are the three layers of ANN. ANN is fully composed of neurons. Neurons is the smallest functional unit in the network. Each neuron in the input layer is associated with weights which is optimized in each iteration to procure the accurate outcome. The weighted inputs are sent to the hidden layer which consists of activation function which produces the outputs to the output layer. The output layer provides the aggregated values based on mathematical operations of averaging. The weights and the activation function play a key role in providing the point of convergence to the problem.

2. K-Nearest Neighbors Classifier (KNN)

KNN is a supervised ML which provides the result purely based on the distance metric computation. The new data is categorized based on its neighbors present in the actual data. The neighbors of the new data are found using distance metrics namely Euclidean distance and Manhattan distance. The value of K determines the number of neighbors to be considered for assigning the label for new data. The K value is an important parameter involved in providing accurate decision. The class labels of K neighbors are taken for arriving at a decision. The maximum occurrence of a particular class label in K neighbors is assigned to the new data.

3. Support Vector Machines (SVM)

The SVM algorithm mainly involves in finding the hyperplane which provides the separation plane for the data. Identification of accurate hyperplane is an important process in the algorithm. The points which are closer to the separation place i.e., hyperplane, is known as support vectors. There are two types of SVM; they are, linear SVM and non-linear SVM. The kernels which are used for non-linear data are polynomial, Radial Basis Function (RBF), Gaussian kernel and Sigmoid kernel.

4. Naïve Bayesian Classifier

Naïve Bayes is a supervised ML algorithm which considers the independence among the features in the data. The assumption of independence among features contributing to the prediction of label reduces the cost of computation. This algorithm is mainly suited for dataset of large dimension.

5. Random Forest

It is a decision tree-based classifier. It constructs various decision trees based on the samples of dataset. The samples can be given either by bootstrapping or without bootstrapping. The predictions of various decision trees are accumulated using voting process. The voting process is of two types and they are, simple majority voting and weighted voting.

6. Bagging

Bagging is a supervised ML which is also known as Bootstrap Aggregation. The decision trees are formed and trained in parallel. Bagging comes under the category of homogenous classifiers, since it trains and combines the predictions of same algorithm type model.

7. Boosting

It is an ensemble-based algorithm. The Boosting algorithm mainly improvises the performance of weak learners like decision tree and SVM. The predictions of week learner are aggregated using simple average method or weighted average method.

8. Stacking

Stacking provides the integrated results of the various base learners. It is also known as stacked generalization.

The summary of the literature survey based on machine learning methods:

1. Single classifier namely ANN, KNN Classifier, SVM, Naïve Bayesian Classifier provides predictions based on single model and provide accuracy at a reliable rate but these models are not as robust as ensemble classifiers.

2. Ensemble classifiers are robust models as they accumulate predictions of various classifiers through averaging or majority voting which makes a highly performing model compared to single classifiers.

3. Machine Learning in The Prediction of Parkinson Disease

C. Okan Sakar et al. [2] have applied Tunable Q-Factor Wavelet Transform (TQWT) along with the learning algorithms to the voice signals of the Parkinson patient and has shown that TQWT is efficient compared to the conventional speech signal processing algorithms. Guruler [3] presented an integration approach for PD diagnosis. The integration framework involves the K-means clustering algorithm and ANN.

Arvind Kumar Tiwari [4] focused on presenting a system which provides PD prediction by incorporating an algorithm for feature selection known as “Minimum Redundancy and Maximum Relevance” which chooses the important features. Sakara et al. [5] presented a system which helps in early diagnosis based on index known as “Unified Parkinson’s Disease Rating Scale”.

Indira R. et al. [6] provided a diagnosis system which uses speech features as the data and fuzzy c-means clustering algorithm for prediction. Shahbakhi et al. [7] presented a system which performs classification of Parkinson’s patients from healthy subjects based on the voice features and using machine learning model namely SVM.

Sellam V. et al. [8] presented a classification system wherein SVM and Radial Basis Function Neural Network (RBFNN) algorithm have been used. The classification system clearly classifies the unhealthy voice and healthy voice based on the acoustic parameters which is extracted from the speech signal. Ma, C. et al. [9] presented a hybrid system based on Extreme Learning (EL) technique that is mainly kernel based. This technique is integrated with feature clustering technique. This hybrid system clearly distinguishes PD patients and healthy subjects.

Yahia et al. [10] proposed a way to detect Parkinson disease using Naïve Bayes and KNN algorithm and has used Parkinson Speech Dataset. Nivedita C et al. [11] classified neurodegenerative disorders according to symptoms using ANN with back propagation. The six major classes of clinical symptoms of neurodegenerative disorders are memory problems, communication issues, changes in personality, idiosyncratic behaviors, loss of voluntary control and common health issues. Farhad.S et al. [12] proposed to discriminate between clinical sample variables (N = 195) that had Parkinson's disease using ANN, RBF, and Multilayer perceptron (MLP).

The summary of literature survey based on ML methods for PD prediction is:

1. Tunable Q Wavelet Transform technique based extracted voice features is the state-of-the-art signal processing technique and provides higher accuracy compared to other voice features.
2. Feature selection techniques frequently used in almost all recent research work are Minimum Redundancy and Maximum Relevance which selects the feature in each iteration based on the relevance factor and tries to minimize redundancy.

3. The prominent ML algorithms used in the literature are SVM and ANN as both the algorithms are efficient in handling higher dimensional data and are predominately used for two class problem.

Table 1 presents the summary of methodologies of the PD diagnosis. The framework highlights four essential modules for automated PD diagnosis. The four modules are:

Table 1. Summary of Methodologies

REF NO.	METHODOLOGIES	CHALLENGES
[2]	Tunable Q-Factor Wavelet Transform	Finding the right feature for convergence of ML model based on 755 features requires lots of trial-and-error experiments.
[3]	K-means Clustering algorithm, ANN	Choosing the right number of clusters for the K means clustering to converge to decision is a tedious and challenging task. ANN has the problem of vanishing gradient which requires early stopping parameter and epochs to be tuned.
[4]	Minimum Redundancy Maximum Relevance	Choosing the score method is critical as it is involved in final decision through its generated score.
[5]	Unified Parkinson's Disease Rating Scale	Time consumption is higher for patients as well as the doctor.
[6]	Fuzzy c-means clustering algorithm	Computation of membership of all data points in every cluster is computationally tedious.
[7]	Support Vector Machine	Choosing the right kernel if the hyperplane, that is the decision boundary, is non-linear.
[8]	Support Vector Machine, Radial Basis Function Neural Network	Data formatting with respect to Gaussian distribution is essential for SVM and RBFNN to make accurate predictions.

[9]	Extreme Learning	Choosing the right activation function to introduce non-linearity to the transformation of input layer.
[10]	Naïve Bayes, K-Nearest Neighbours	Data pre-processing with respect to the distribution of data.
[11]	Artificial Neural Network – Back Propagation	Choosing the right backpropagation weight update algorithm is essential for preventing vanishing gradient problems. The most prominent backpropagation weight update algorithm used is gradient descent.
[12]	Multi-Layer Perceptron	Tuning the hyperparameter values of MLP.

4. Generic Framework of PD Diagnosis

Figure 1 presents the generic framework for PD diagnosis. The generic framework of PD diagnosis is designed based only on voice signal modality. This framework will be useful as a workflow for building the system for PD diagnosis. The framework has been formed from the accumulated knowledge extraction from the review of research papers. The subjects involved for the data collection are healthy subject and PD patients.

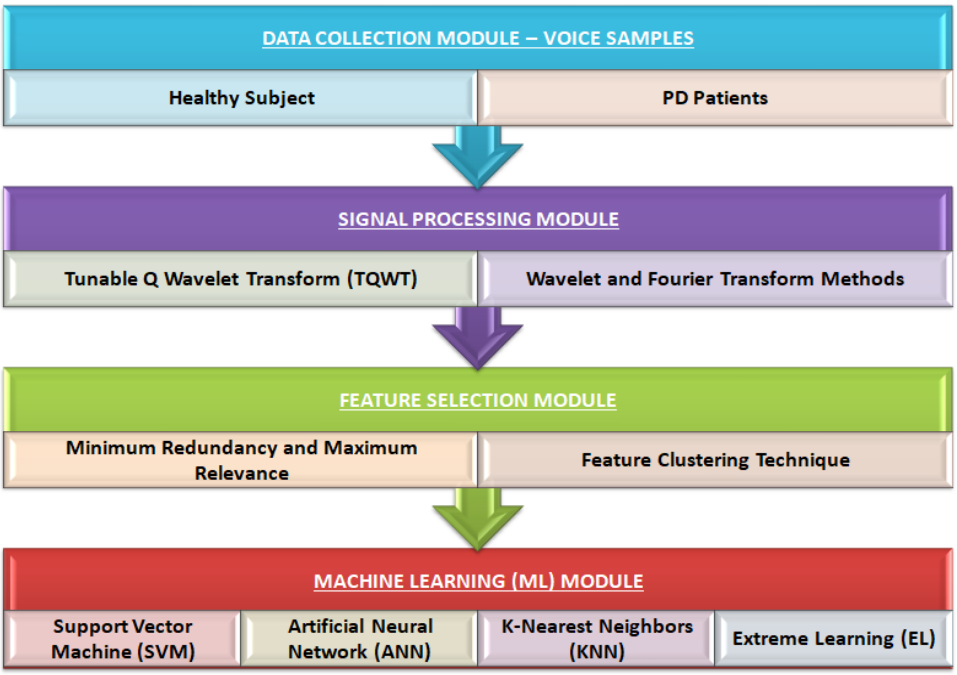

Figure 1. Generic Framework for PD Diagnose

- **Data Collection Module:** The voice samples are collected from healthy and PD patients. There can be repetitive samples to ensure double validation for recording.

- **Signal Processing Module:** The raw voice samples are processed to remove noise and to extract meaningful information which can be useful to train the ML algorithm. The raw voice signal is taken as a spectrum and features are extracted from the voice signal spectrum by applying the Fourier and Wavelet Transform methods. The novel and state-of-the-art signal processing technique namely Tunable Q Wavelet Transform has been employed in the existing system [1] and it provided reliable features which produced higher accuracy compared to the conventional signal processing technique used in other existing system.

  - **Tunable Q Wavelet Transform**

This technique is an upgraded version of the mathematical transform namely wavelet transform. It breaks down the signal into three Q factors namely high Q-factor, low Q-factor and residual components [13]. The signal's oscillatory behavior determines the Q-factor. TQWT possess the property of reconstruction. Tunable Q-factor is tuned based on two parameters namely Q-factor and oversampling rate. The main reason to incorporate TQWT in generic framework is that it is the state-of-the-art technique for signal processing and proven in [2] to be a reliable approach for feature extraction from voice signal spectrum.

  - **Wavelet and Fourier Transform Methods**

The frequency domain signal processing technique is known as Wavelet Transform (WT). The frequency and time-based domain signal processing technique is known as Wavelet Transform (WT) and Fourier Transform (FT). The representation of signal in frequency domain by FT and representation of signal in frequency and time domain by WT is highly efficient. Both transforms when the incorporated provides reliable features for further analysis due to their way of representation. The main reason to include FT and WT in generic framework is due to their representations.

- **Feature Selection Module:** The important features which provide higher discrimination between healthy subject's samples and PD patient's samples are selected using feature selection techniques.

  - **Minimum Redundancy and Maximum Relevance (MRMR)**

MRMR is a minimal optimal method which selects the small set of features that helps in the convergence of decision. The feature set obtained from implementing MRMR produces higher accuracy since the feature set is of high relevance with respect to the class label and very less redundancy with respect to the class label.

- **Feature Clustering technique (FC)**

FC is used for feature selection based on clustering technique.

• Machine Learning Module: The ML algorithm is trained using the features obtained from feature selection module, and the evaluation of ML algorithm prediction is performed and further tuning of algorithms is done based on accuracy parameter [14,15].

- **Support Vector Machine**

SVM clearly discriminates two class labels based on the lagrangian multipliers. The decision boundary known as hyperplane can be of two types namely linear and nonlinear. The Linear SVM hyperplane can be solved using the straight-line equation in 2D plane whereas Non-Linear SVM can be solved by projecting to higher dimensional plane through the mapping function based on Kernels. The generic framework includes SVM as one of the ML models as it is efficient in handling higher dimensional data. SVM model has been included in the framework, since extracted voice features in the benchmark data consists of 755 features.

- **Artificial Neural Network**

In the three layers of the ANN, the hidden layer introduces the non-linearity by incorporating activation functions to the output obtained by linear mapping transformation of input layer with corresponding weights. Neuron is the basic computational unit in ANN. All layers consist of n number of neurons. The two main propagations performed in ANN are forward propagation and backward propagation. During forward propagation, the following steps are performed:

i. Linear Mapping of inputs and weights

ii. Incorporation of activation function.

iii. Computation of error

During back propagation, the following steps are performed:

i. Computation of gradients

ii. Updating the weights

The main reason of using ANN in generic framework is that it introduces non-linearity which aids in solving complex problems that involves complex features with large dimensionality.

- **K-Nearest Neighbors**

KNN is the clustering-based technique, where the model makes prediction based on K nearest neighbors. KNN is simple in nature and therefore it is easier to implement and interpret.

- **Extreme Learning**

EL is simple Feed Forward Neural Network (FFNN) with one hidden layer. The speed of learning is higher and the capability of generalization is also higher.

The techniques highlighted in each module of generic framework are the most accurate techniques which are presented based on the review of research papers. Table 2 presents the techniques used in generic framework along with the reasons for the techniques to be accurate.

Table 2. Generic Framework – Techniques and its Reason

Technique	Reason to be Accurate
TQWT	Captures the oscillatory behaviour of the signal with the tunable parameters Q-factor and oversampling rate.
WT and FT	The representation of signal is efficient in WT and FT as it covers both frequency domain, and time domain with respect to WT and frequency domain with respect to FT.
MRMR	The feature set is small and the feature selection procedure is minimal and optimal. MRMR technique is fast and reliable.
FC	Clustering based feature selection provides data visualization in an easier manner.
SVM	SVM handles higher dimensionality data efficiently and produces best hyperplanes for 2 class problems.
ANN	ANN introduces non-linearity which helps in solving complex problems.
KNN	Simplified implementation. No assumption of data is made hence it is accurate in prediction.
EL	High generalization ability and efficient learning.

5. Conclusion

This study presents concise explanation of various Machine Learning (ML) algorithms. This research also presents the conventional work that utilizes ML algorithms for the prediction of Parkinson Disorder (PD). The work highlights the summary of methodologies from the review and presents the design of generic framework for PD diagnosis which will be useful to develop system for PD diagnosis based on voice samples.

References

[1] J.Jankovic, Parkinson's disease: clinical features and diagnosis, J. Neurol.Neurosurg. Psychiatry 79 (4) (2008) 368–376.

[2] C. Okan Sakar, Gorkem Serbes et al, "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform", Applied Soft Computing Journal 74 (2019) 255–263.

[3] H.Gürüler, A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method, NeuralComput.Appl.28(2016)

[4] Arvind Kumar Tiwari "Machine learning based approaches for prediction of parkinson's disease", Machine Learning and Applications: An International Journal (MLAIJ) Vol.3, No.2, June 2016.

[5] Sakara, Batalu. E., & Kursunb, (2014)O. Telemonitoring of changes of unified Parkinson's disease rating scale using severity of voice symptoms.

[6] Rustempasic, Indira, & Can, M. (2013). Diagnosis of Parkinson's Disease using Fuzzy C-Means Clustering and Pattern Recognition. SouthEast Europe Journal of Soft Computing, 2(1).

[7] Shahbakhi, M., Far, D. T., & Tahami, E. (2014). Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine. Journal of Biomedical Science and Engineering, 2014.

[8] Sellam, V., & Jagadeesan, J. (2014). Classification of Normal and Pathological Voice Using SVM and RBFNN. Journal of Signal and Information Processing, 2014.

[9] Ma, C., Ouyang, J., Chen, H. L., & Zhao, X.H. (2014). An Efficient Diagnosis System for Parkinson's Disease Using Kernel Based Extreme Learning Machine with Subtractive Clustering Features Weighting Approach. Computational and mathematical methods in medicine, 2014.

[10] Yahia A, Laiali A. (2014). Detection of Parkinson Disease through Voice Signal Features. Journal of American Science 2014; 10(10), 44-47.

[11] Nivedita C., Yogender A., R. K. Sinha (2013). Artificial Neural Network based Classification of Neurodegenerative Diseases. Advances in Biomedical Engineering Research (ABER) Volume 1 Issue 1.

[12] Farhad S. G., Peyman M. (2013). A Case Study of Parkinson's disease Diagnosis using Artificial Neural Networks. International Journal of Computer Applications (0975 – 8887) Volume 73– No.19

[13] H. S. Pal, A. Kumar and A. Vishwakarma, "TQWT based Electrocardiogram Compression using Optimized Thresholding," 2021 Advanced Communication Technologies and Signal Processing (ACTS), Rourkela, India, 2021, pp. 1-5, doi: 10.1109/ACTS53447.2021.9708289.

[14] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

[15] API design for machine learning software: experiences from the scikit-learn project, Buitinck et al., 2013.