IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 7 (July. 2019), ||S (I) || PP 13-18

A Survey on Machine Learning Techniques for Parkinson's Disease Diagnosis and Classification

Sherina T.M[1], Mr. Senthilkumar[2] .B
M.Phil Scholar, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu, India [1]
Assistant Professor, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu, India[2]
Corresponding Author: Sherina T.M

Abstract: Disease diagnosis and prediction using data mining techniques are emerging as a research area with different tools and techniques. Parkinson's disease is one of the main disorders of the central nervous system. Disease diagnosis and classification from the digital data is becoming one of the major concerns of health care data mining. Researches use many technologies like machine learning, image analysis, signal analysis and device based techniques to examine the disease data from various sources. There is a need to analyze the unstructured large data using data mining. There are several tools and techniques to achieve this. This survey focuses on the various tools and techniques used for Parkinson's data analysis and classification, and finally provide an outline to overcome the problems of those techniques.

Keywords: Data Mining, Parkinson's disease, Parkinsonian, Machine Learning, deep learning algorithms, Classification, prediction.

Date of Submission: 22-06-2019 Date of acceptance: 10-07-2019

I. INTRODUCTION

Parkinson's disease is a disorder of the central nervous system. This disease caused by a complaint in the central nervous system which affects the human movement. This affects the human brain. The indications of the disease increase due to the death of dopamine producing cells in the mid brain. This disease does not have any known cause and the signs and symptoms can be different for everyone. The symptoms differ in different stages of the disease but generally they all involve cognitive and behavioral problems. Parkinson's disease is the second most common neurodegenerative. Alzheimer's disease affects many people worldwide. Due to the death of nigral neurons, patients experience both motor and non-motor symptoms, affecting their quality of life. The reasons for the cell death are still poorly understood, and there is currently no cure for Parkinson's disease. Physicians try to manage patients' symptoms by introducing medications therapies, using antiparkinson medications. Physicians need to carefully prescribe medications therapies since the prolonged intake—in particular of higher dosages of antiparkinson medications—can have significant side-effects. Changes of the status of Parkinson's disease patients through time is a result of the natural progression of the disease and the medications that the patients are prescribed in order to keep their status stable as long as possible.

Data mining is the effective way to handle different types of clinical dataset. This survey concentrated on the Parkinson's disease related researches. Various machine learning techniques proposed to handle the Parkinson's disease dataset. However, this is necessary to analyze the specific techniques, which are suitable for the Parkinson's dataset. Because, Parkinson's disease datasets are categorized into two major types such as speech and handwriting datasets. The data set of Parkinson's disease dataset is available in the UCI repository.

1.1 Data mining in healthcare:

Healthcare organizations today are capable of generating and collecting large amounts of data. This increase in volume of data requires automatic way for these data to be extracted when needed. With the use of data mining techniques it is possible to extract interesting and useful knowledge and regularities. Knowledge acquired in this manner, can be used in appropriate area to improve work efficiency and enhance quality of decision making process. The above statements points out that, there is a great need for new generation of computer theories and tools to help people with extracting useful information from constantly growing volume of digital data.

In addition, the use of information technology in healthcare enables comprehensive management of medical knowledge and its secure exchange between recipients and providers of healthcare services.

Widespread use of information technology enables the elimination of manual tasks of data extraction from charts or filling of specialized questionnaires, extraction of data directly from electronic records, transfer on secure electronic system of medical records that will save lives and reduce the cost of health care, early detection of infectious diseases with advanced collection of data etc. Retrieval of information with the help of computers can help the quality of decision making and avoiding human errors. When there is a large volume of data that needs to be classified, decision making by people is usually poor.

II. LITERATURE REVIEW

Machine Learning is a branch of computational intelligence dedicated to the development of algorithms that enable a computer program to improve its performance based on prior (learned) information. Since the very beginning of "Perceptron", new mathematical modeling of the working mechanism of the brain has been pursued daily. Such intense research has motivated a number of works that aimed at using machine learning-oriented techniques to aid Parkinson's disease recognition.

Drotár et al. [1], authors proposed a study with some features based on entropy, energy and intrinsic measures of the handwriting skills of an individual. The main objective of this paper is to present the PaHaW Parkinson's disease handwriting database which contains many handwriting samples from Parkinson's disease (PD) patients. This also includes the healthy person dataset for comparison. Another goal of this study is to show that kinematic features and pressure features in handwriting can be used for the differential diagnosis of PD. The authors used PD database, which contains 37 PD patients and 38 healthy persons. To discriminate between PD patients and healthy subjects, three different classifiers were compared: K-nearest neighbors (K-NN), ensemble AdaBoost classifier, and support vector machines (SVM). The authors also considered applying such measures to in-air movements and pressure to exploit the full potential of the handwriting. From the experimental results, the prediction of PD from kinematic and pressure features of handwriting is analyzed. An SVM-RBF kernel was used for classification purposes, thus achieving a classification accuracy of around 81.3% of prediction performance. The authors performed the comparison of different classifiers for diagnosis of Parkinson's disease from hand-writing datasets. Three metrics were compared such as accuracy, specificity, and sensitivity. SVM achieved 81.3 80.9 87.4, AdaBoost 78.9 79.2 82.4 and K-NN 71.7 70.8 78.5 respectively.

Connolly et al. [2] , in this paper authors applied Linear Discriminant Analysis Support (LDA), Support Vector Machines and k-nearest neighbor (k-nn) upon local field potentials sensed from an implanted deep brain stimulation device. Authors used leave one out technique feature selection and performed analysis. For such analysis, 83 montages were recorded from 15 patients suffering from advanced idiopathic PD, thus obtaining an accuracy rate of 91%. The main advantage of this result is more effective and efficient than other feature selection.

Wahid et al. [3] presented a study with two main contributions: firstly, they used a multiple regression normalization strategy to identify differences in spatial-temporal gait features between PD patients and control (healthy) individuals. Secondly, they evaluated the effectiveness of machine learning strategies in classifying PD gait after multiple regression normalization. The authors argued the study has important implications for the analysis of spatial-temporal gait data concerning the diagnosis of PD, as well as the evaluation of its severity with five machine learning strategies employed to classify PD gait: kernel Fisher Fiscriminant (KFD), Naïve Bayesian Approach (NB), k-nn, SVM, and Random Forest (RF).

Smith et al. [4] employed evolutionary algorithms to provide clinically relevant and objective measures to identify PD both in humans and animal models. The human data were collected from commercial sensors via non-invasive procedures, and the animal data were collected using fruit flies with and without PD genetic mutations. Their work used Cartesian Genetic Programming, thus showing such technique can be successfully applied to the assessment of movements in humans when distinguishing PD patients from healthy controls, as well as to classify severity of dyskinesia in patients.

Hirschauer et al. [5] presented a new method for the diagnosis of PD based on continuous phonation samples, which were used as attributes. A technique called minimum Redundancy Maximum Relevance (mRMR) was applied for the identification of the most relevant attributes, and the results were compared to a variety of feature selection algorithms. The dataset used in this study was originally obtained from the University of Oxford in cooperation with the National Voice and Speech Center, Denver, Colorado.2 After the feature selection process, the data are used to feed to different neural classifiers: a standard Artificial Neural Network (ANN) and a Complex-Valued Neural Network (CVANN). The results were promising: ANN obtained an accuracy of 94.28%, and CVANN achieved an accuracy of 98.12%.

Ahmadlou et al. [6] presented an Enhanced Probabilistic Neural Networks (EPNN), a machine learning technique that make use of local decision circles surrounding training samples to control the spread of the Gaussian kernel. Using the Parkinson's Progression Markers Initiative dataset, the proposed approach obtained an accuracy of 98.6% when classifying healthy people from PD patients, and 92.5% of recognition rate when dealing with data of six clinical exams and functional neuro imaging data for two regions of interest of the brain.

Segovia et al. [7] demonstrated a new method based on SVMs and Bayesian networks to separate IPS from APS (atypical parkinsonian syndromes) that makes use of the 18FFDG PET dataset, that allows assessing the glucose metabolism of the brain. Their methodology achieved an accuracy rate over 78%, a reasonable result between sensitivity and specificity, suggesting the proposed method is suitable to assist the diagnosis of PD.

Cook et al. [8] proposed to employ a combination between smart home and machine learning technologies to observe and quantify the behavioural changes of PD patients. The main focus is to aid the clinical assessment and a better understanding of the differences between healthy older adults (HOA) and older adults with cognitive and physical impairments, also classified by the authors as mild cognitive impairment (MCI). The results indicated that smart homes, wearable devices and ubiquitous computing technologies can be useful for monitoring the activity of PD patients, as well as to pinpoint the differences between HOAs and older adults with PD or MCI. However, the authors described some limitations concerning the devices, such as to operate in settings with multiple residents and interrupted activities.

In 2015, Shamir et al. [9] proposed an approach called Clinical Decision Support Systems (CDSS) to examine the results of the incorporation of patient-specific symptoms and medications into three key functions: (i) information retrieval; (ii) visualization of treatment; and (iii) recommendation on expected effective stimulation and drug dosages. In order to fulfil this purpose, the authors used Naïve Bayes, Support Vector Machines and Random Forest to predict the treatment outcomes. The combined machine learning algorithms were able to accurately predict 86% of the motor improvement scores at one year after surgery.

Tucker et al. [3810 predict a PD patient's adherence to medication protocols based on variations in their gait. Using whole-body movement data readings from the patients, it is possible to discriminate PD patients that are "on" or "off" medication with accuracy of 97% using an individually customized model, and an accuracy of 78% considering a generalized model containing multiple patient gait data.

Procházka et al. [11] presented a novel method of Bayesian gait recognition using a Kinect sensor (data acquisition and spatial modelling) combined with signal processing techniques and Bayesian classifier for gait feature analysis aiming at recognizing individuals affected by Parkinson's disease.

Singh and Samavedham [12] proposed an innovative and effective approach for monitoring the disease progression and clinical diagnosis, which is based on the combination of Self-Organizing Maps and Least Squares Support Vector Machines. The proposed approach can achieve an accuracy of up to 97% concerning the differential diagnosis of PD using the PPMI dataset. The same group of authors used unsupervised learning techniques to identify reliable biomarkers to aid the diagnosis of neurodegenerative diseases [13].

Table 1: summarizes the main works concerning machine learning based solutions for the automatic PD diagnosis.

Paper ID	Dataset	Description	Technique	Accuracy	Merits	Demerits
Drotár et al. [1]	Parkinson's Disease Handwriting Database.	To study features based on entropy and energy concerning the handwriting skills.	SVM with RBF kernel.	81.30 %	Effective feature is identified.	The accuracy level is low.
Connolly et al. [2]	Data from 15 patients with a large set of spectral features.	To employ signals sensed from Deep Brain Stimulation.	leave-one-out Technique	91%	Effective and efficient	Small number of dataset used.
Wahid et al. [3]	Data from 23 PD patients and 26 aged-matched controls.	To identify differences in spatial-temporal gait features between PD patients and control (healthy) individuals using multiple regression normalization.	KFD. Compared with NB, k-NN, SVM and RF.	87.40 %	Feature reduction process improves the performance	Accuracy is low
Smith et al. [4]	Data from 49 PD patients and 41 Age-matched	To apply Evolutionary Algorithms to study motor	Cartesian Genetic Programming	78 %	Helps to find the severity	Iteration and process time is high

	healthy controls.	functions in both humans and animal models.				
Hirschauer et al. [5]	Parkinson's Data Set obtained from UCI Machine Learning Repository.	To present a new feature selection algorithm known as mRMR.	Feature selection: mRMR. Classifier: ANN CVANN	98.12	The feature selection algorithm eases the iteration in classifier.	Not suitable for unstructured dataset
Ahmadlou et al. [6]	Parkinson's Progression Markers Initiative.	To present a new machine learning technique that makes use of local decision circles surrounding training samples to control the width of the Gaussian kernel.	ML	98.6%	Improves accuracy level	Not suitable for other features
Segovia et al. [7]	F-FDG PET dataset	To distinguish Parkinson's disease from APS using machine learning techniques	(Bayesian network) Classifier: SVM, NB	78.16%	this methodology can be used as an alternative to the majority voting strategy and other approaches to deal with multiple classifier decisions	Need diagnosis improvement
Cook et al. [8]	84 Caucasian community-dwelling older adults from a larger sample of 260 individuals	Machine learning technologies combined with smart home	ML	97 %	useful for monitoring the activity of PD patients. improve classifier performance	concerning the devices. Interrupted activities.
Shamir et al. [9]	Clinical data from 10 patients and 89 post-DBS surgery visits were used.	To develop a proof-of-concept implementation of a CDSS that incorporates patient-specific details on both stimulation and medication.	SVM, NB, and RF.	71. 0 %	Effective performance	Poor output
Tucker et al. [10]	Composed of non-wearable multimodal sensors.	To model and predict PD patients' adherence to medication protocols based on	SVM	95 %	customized monitoring	The accuracy and robustness of the models

		variations in their gait.				needs to be validated.
Procházka et al. [11]	Kinect sensor (data acquisition and spatial modelling).	To propose a novel method of Bayesian gait recognition.	Bayesian	94.6 %	Improves accuracy	Iteration is high

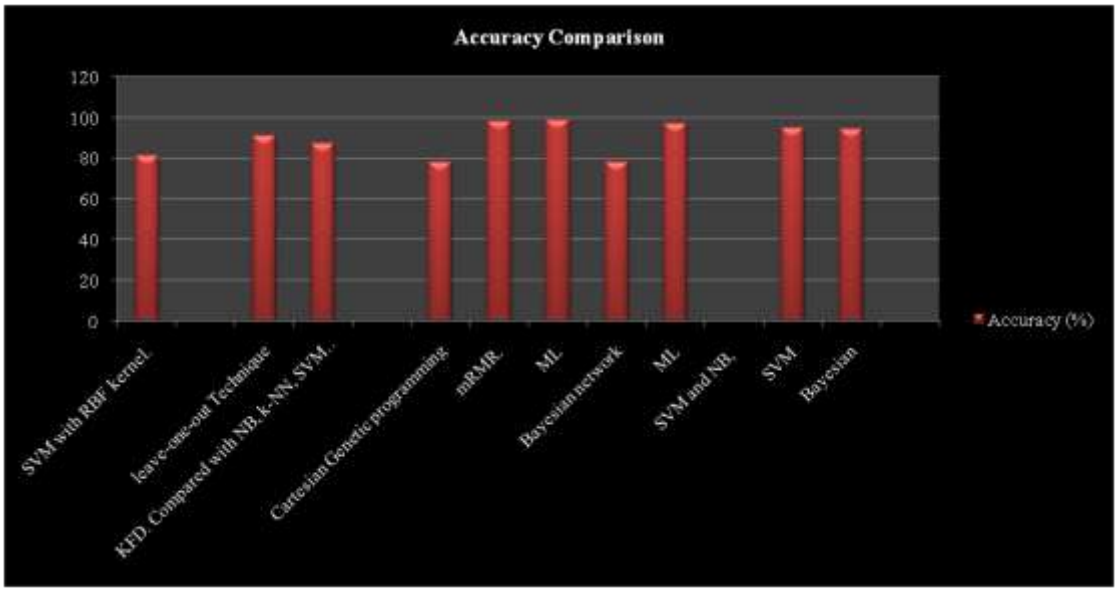
Based on the above different researches, many techniques and algorithms from several papers are found. Some techniques are more suitable for PD analysis. The above mentioned papers shows many authors used machine learning techniques along with several device based applications, and every algorithm is designed with the intension of improving accuracy and reducing iterations. Several machine learning based approaches and feature selection algorithms were used to increase the classification efficiency in different PD datasets. However, the algorithms and techniques could cover a single factor of improvement than all other metrics. The table 1.0 shows the summary and finding from the literature of various PD analysis and diagnosis.

Summary:

The existing techniques and tools for Parkinson's disease analysis and diagnosis are based on the machine learning and image analysis techniques. These techniques need more accurate data samples. The data samples should be proper and the type should be specified correctly. As said, the PD analysis and classification accuracy is based on the effective machine learning and feature selection techniques. However, many authors simply compared the existing with various dataset. So the effective technique for PD analysis needs complete data samples with effective feature selection.

Table 2.0 accuracy comparison

Technique	Accuracy (%)
SVM with RBF kernel.	81.30
leave-one-out Technique	91
KFD. Compared with NB, k-NN, SVM and RF.	87.40
Cartesian Genetic programming	78
mRMR.	98.12
ML	98.6
Bayesian network	78.16
ML	97
SVM and NB,	71. 0
SVM	95
Bayesian	94.6



III. CONCLUSION

The disease analysis and classification plays an important role in the health care industry using data mining. Detection and analysis of disease from various difficult dataset is the most important issue the research trend, since the lack of data samples and sufficient features to detect the Parkinson's disease. These processes create difficulties in the accurate result gain. However, there are several types of disease diagnosis techniques available from sensor to image analysis, every technique need a data mining tool and approach to handle such huge datasets. There are several different methods to diagnosis and classification of Parkinson's disease. This survey offered different machine learning techniques to solve the Parkinson's disease analysis and classification problem. From the analysis, effective feature selection and appropriate classifier will give better result.

REFERENCES

[1]. Drotar P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. Artif Intell Med 2016:39–46.
[2]. Connolly AT, Kaemmerer WF, Dani S, Stanslaski SR, Panken E, Johnson MD, Denison T. Guiding deep brain stimulation contact selection using local field potentials sensed by a chronically implanted device in Parkinson's disease patients. 7th international conference on neural engineering 2015:840–3.
[3]. Wahid F, Begg RK, Hass CJ, Halgamuge S, Ackland DC. Classification of Parkinson's disease gait using spatial-temporal gait features. IEEE J Biomed Health Inform 2015; 19(6):1794–802.
[4]. Smith SL, Lones MA, Bedder M, Alty JE, Cosgrove J, Maguire RJ, Pownall ME, Ivanoiu D, Lyle C, Cording A, Elliott JHC. Computational approaches for understanding the diagnosis and treatment of Parkinson's disease. Inst Eng Technol Syst Biol 2015;9(6):226–33. (7).
[5]. Hirschauer TJ, Adeli H, Buford JA. Computer-aided diagnosis of Parkinson's disease using enhanced probabilistic neural network. J Med Syst 2015; 39(11):1–12.
[6]. Ahmadlou M, Adeli H. Enhanced probabilistic neural network with local decision circles: a robust classifier. Integer Compute-Aided Eng 2010; 17(3):197–210.
[7]. Segovia F, Illán IA, Górriz JM, Ramírez J, Rominger A, Levin J. Distinguishing Parkinson's disease from atypical parkinsonian syndromes using pet data and a computer system based on support vector machines and bayesian networks. Front Comput Neurosci 2015; 9(137):1–8.
[8]. Cook DJ, Schmitter-Edgecombe M, Dawadi P. Analyzing activity behavior and movement in a naturalistic environment using smart home techniques. IEEE J Biomed Health Inform 2015; 19(6):1882–92.
[9]. [Shamir RR, Dolber T, Noecker AM, Walter BL, McIntyre CC. Machine learning approach to optimizing combined stimulation and medication therapies for Parkinson's disease. Brain Stimul 2015; 8(6):1025–32.
[10]. Tucker CS, Conrad S, Behoora I, Nembhard HB, Lewis M, Sterling NW, Huang X. Machine learning classification of medication adherence in patients with movement disorders using non-wearable sensors. Comput Biol Med 2015; 66:120–34.
[11]. Procházka A, Vysata O, Valis M, Tupa O, Schätz M, Marík V. Bayesian classification and analysis of gait disorders using image and depth sensors of Microsoft kinect. Digit Signal Process 2015; 47:169–77.
[12]. Singh G, Samavedham L. Algorithm for image-based biomarker detection for differential diagnosis of Parkinson's disease. Int Fed Autom Control – Papers Online 2015; 48(8):918–23.
[13]. Singh G, Samavedham L. Unsupervised learning based feature extraction for differential diagnosis of neurodegenerative diseases: a case study on early-stage diagnosis of Parkinson disease. J Neurosci Methods 2015; 256:30–40.

Sherina T.M. "A Survey on Machine Learning Techniques for Parkinson's Disease Diagnosis and Classification." IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 07, 2019, pp13-18.