DETECTING PARKINSON'S DISEASE USING MACHINE LEARNING

LITERATURE SURVEY

INTRODUCTION

Parkinson's disease (PD) is one of the most common neurodegenerative diseases with a prevalence rate of 1% in the population above 60 years old, affecting 1 - 2 per 1000 people. The estimated global population affected by PD has more than doubled from 1990 to 2016 (from 2.5 million to 6.1 million), which is a result of increased number of elderly people and age-standardized prevalence rates. PD is a progressive neurological disorder associated with motor and non-motor features which comprises multiple aspects of movements, including planning, initiation and execution. The diagnosis of PD is traditionally based on motor symptoms. Despite the establishment of cardinal signs of PD in clinical assessments, most of the rating scales used in the evaluation of disease severity have not been fully evaluated and validated

This section of the describes the theoretical background of this project, starting with an explanation of Parkinson's disease, followed by overviews of machine learning, deep learning, related work and finally PD diagnosis problems. The detection of PD is extremely important at the first stage.This entails the literature survey that was conducted for the purpose of better understanding the problem at hand and to explore possible solutions.

Importance of Voice data:

Speech or voice data is assumed to be 90% helpful to diagnose a person for identifying presence of disease. In general, Person with PD suffer from speech problems, which can be categorized into two: hypophonia and dysarthria. Hypophonia indicates very soft and weak voice from a person and dysarthria indicate slow speech or voice, that can hardly be understood at one time and this causes because of damage to central nervous system.

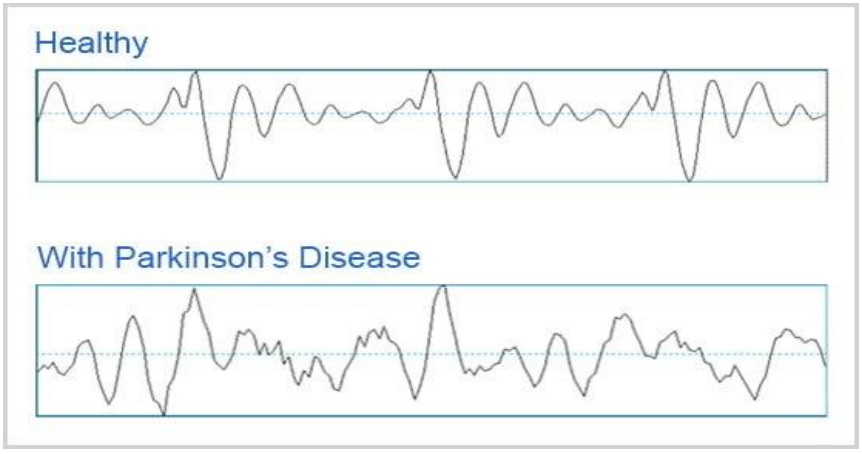

Fig 1.1 Voice data analysis of affected vs healthy person

So, most of the clinicians who treat PD patients observe dysarthria and try to rehabilitate with specific treatments to improvise vocal intensity

Survey carried out for the diagnosis of PD with different algorithms and approaches:

Several strategies are recorded for early detection of PD based on the different ML techniques. But accuracy in detection and classifying within the time is very important or else, it causes development of more symptoms. There are different kinds of data, brain MRI images, Voice data, posture images, senor captured data, handwritten data, using which we can predict whether person is having PD or not. Out of all those, speech or voice data helps in identifying PD accurately.

Jie Mei et al. used all basic algorithms of deep learning techniques for the detection of PD. Like SVM, RF, Decision Tree, ANN, KNN, Radial Basis Function Networks (RBF) and Deep Belief Networks (DBN) etc. The early identification of Parkinson's disease is critical. The identification can be performed with the use of a data mining technique. The techniques for detecting PD, such as Naive Bayes, support vector machine, multilayer perceptron neural network, and decision tree, are theoretically explained in this study. This study uses speech input from acoustic devices to predict Parkinson's disease. People from various areas and speech factors are investigated in this article in order to predict Parkinson's disease among patients. The speech dataset was used to recognize Parkinson's illness using Multi - layer Perceptron and Logistic Regression (LR) frameworks.

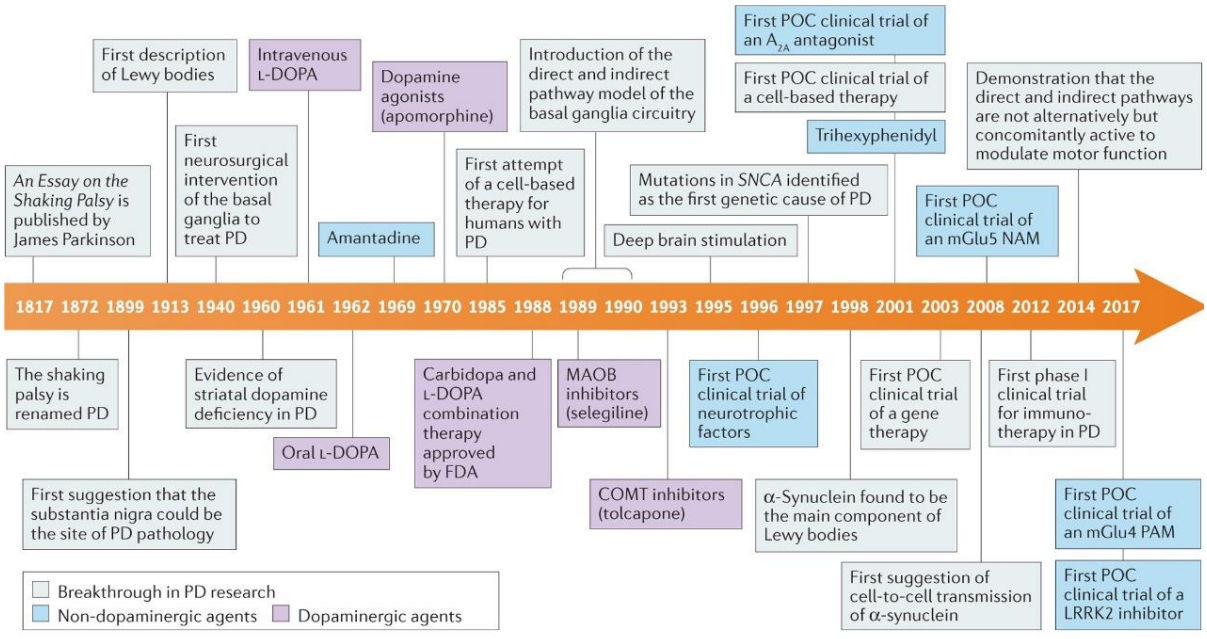

Fig 1.2 Research on Parkinson’s disease

Author Name: DAVID GIL A, MAGNUS JOHNSON B

Title of the Paper: Diagnosing Parkinson by using Artificial Neural Networks and Support Vector Machines

Description: He found that with a smaller number of neurons at hidden layer both training set and test sets performed poorly. With higher number of neurons, the training set performed well with high risk of over fitting. The ideal solution for this layer was found to be 13 neurons.

Author Name: Mohammad S Islam

Title of the Paper: The Mechanistic Role of Thymoquinone in Parkinson's Disease: Focus on Neuroprotection in Pre-Clinical Studies

Description: He compared various ML techniques based on their performance accuracies in determining whether person is having PD or not and mentioned that new classifier may be built to get better accuracies.

Author Name: Kazi Amit Hasan

Title of the Paper: Classification of Parkinson's Disease by Analyzing Multiple Vocal Features Sets

Description: He used different classification methods RF, KNN, Decision Tree, Logistic Regression (LR), SVM, and Naïve Bayes for detection of PD. The best result achieved by Decision Tree and Random Forest (RF) classification methods. The data mining techniques may be a more popular in many field of medical, business, railway, education etc. They are most commonly used for medical diagnosis and disease prediction at the early stage. The data mining is employed for healthcare sector in industrial societies.

Author Name: Shail Raval

Title of the Paper: A Comparative Study of Early Detection of Parkinson's Disease using Machine Learning Techniques

Description: For the detection of PD they include all the aspects such as biological data, chemical data and genetic data. In this paper they mainly focused on the symptoms like rigidity, Tremor at rest, changing voice etc. The secure data transmission is proposed through authentication check, duplication check and faulty node detection. The proposed method is applicable to long ranges of transmission. It is also supporting a retransmission concept.

Author Name: Rajalakshmi Shenbaga Moorthy

Title of the Paper: Freezing of Gait Prediction in Parkinsons Patients Using Neural Network

Description: Novel analytic system for Parkinson’s disease Prediction mechanism using Improved Radial Basis Function Neural Network (IRBFNN). RNNs is during a one

among the deep learning models that are used for modeling the arbitrary length sequences by applying a transition function to all or any it's hidden states during a recursive manner.

Authors name	Machine learning methods	Data description	Performance
			99.39% specificity
Hazan H. (2012)	SVM	speech data	90% accuracy
Sriram T.	Random Forest	voice dataset	90.26% accuracy
Prashanth (2014)	SVM and classification tree	de-novo PD	89.39% accuracy
Amit S. (2014)	Using nonlinear dynamic and SVM	Dyskinesia Data	66% to 77% accuracy
A.H.(2012)	classification and regression tree	Recorded voice signals	93.7% accuracy
Kaya E.(2011)	Entropy-based discretization method	Audio Input	94.87% accuracy
Nivedita C. (2013)	ANN	seven different classes	overall 96.42% accuracy
Farhad S.(2013)	Multi-Layer Perceptron (MLP)	Audio Input	93.22% accuracy
Chen A.(2012)	GA-SVM classifier	microarray dataset	69% to 94% accuracy, 60% to 92% sensitivity and 70% to 95% specificity
Wu D. (2010)	Radial basis function neural network	data recorded	89.91% accuracy
Luukka P. (2011)	fuzzy entropy	voice recording	84.52 % accuracy
Salhi L.(2008)	Multilayer Neural Network	pitch around the expected value (250Hz)	80% and 100%.Results
Max A. (2009)	SVM	speech signals	91.4% accuracy
Zhang, J. (2008)	Increased cerebrospinal fluid(CSF)	AD and PD dataset	95% PD and 75% AD
Saad A. (2011)	Bayesian Believe Network	tremor, poor vocal data	Group or cluster form
Ozcift A.(2011)	Correlation based Feature Selection	PD data samples	89.7% accuracy
Yahia A. (2014)	KNN	speech dataset	93.3% accuracy

Fig 1.3. Literature survey for the detection of Parkinson's disease

DISCUSSIONS

Machine Learning techniques has got prominent role as they are applied in variety of domains especially in the healthcare. Unlike traditional methods, the models generated by applying ML techniques show dynamic outputs as data is fed into it. One shall make note that significant and narrow research is needed to obtain knowledge in diagnosing the disease. Various machine learning algorithms and techniques are being proposed rapidly, out of which some are observed to be promising with the results and few demonstrated their usage in different fields. Advantage with the ML generated models is that when more data is used, the precision values gets increased and the much accuracy in predictions can be gained.

Every circle in the above network represents a neuron at which the inputs and corresponding weights are processed layer by layer.

Input Layer: An input layer accepts large volumes of data as input to build the neural network. The data can be in the form of text, image, audio, etc. In general, the input layer contains features of the dataset, each node of input layer in the above architecture represents one feature

Hidden Layer: Every hidden layer receives the input feature along with their weights, where weights of every feature indicates their contribution towards the decision or prediction. Hidden layer processes the data at each node by performing complex computations and helps in feature extraction. Nodes at first hidden layer receives product of input feature with its weights value

and is passed as input to next hidden layer and so on. Choosing number of hidden layers and number of nodes for every hidden layer varies with the problem as well as dataset.

Output Layer: At output layer Processing of nodes are determined by the functions called as activation functions like tanh, sigmoid, ReLU. Depending on the kind of dataset and criteria, one can decide suitable activation function. Output layer receives the output generated by last hidden layer as input and generates an output in the desired form.

CONCLUSION FROM LITERATURE SURVEY:

Prediction of Parkinson disorder is one of the most important problem that has to be detected in the early phases of the commencement of the disease so as to reduce thedisease progression rate among the individuals .Various researches have been made to find the basic cause and some have reached to the heights by proposing a system which differentiates the healthy people from those with any ND'S (Neurodegenerative disorders) using various machine learning techniques. Lots of pre-processing feature selection and classification techniques have been implemented and developed in the past decades.