International Conference on Intelligent Technologies & Science - 2021
(ICITS-2021)

A Review on Parkinson's Disease Prediction using Machine Learning

Vighnesh Garg[1], Anurag Jaiswal[2]
Computer Science and Engineering,
Goel Institute of Technology & Management, Lucknow, India
vighnesh017@gmail.com, anurag.jaiswal@goel.edu.in

Abstract: Parkinson disease is a non-curable disease which directly affects the central nervous system. Which is on a raise gradually and estimated to affect the most of the lives within a decade. This disease can be controlled if detected at an early stage. The Parkinson disease after effects are it affects the movement of the patients, tremors, and stiffness in body. The exact cause of Parkinson disease is still not known ,it is believed that this disease is liked to damage in the brain cells .generally patients over 50 years are venerable to this disease. Its symptoms appear slowly in the patients. Still no proper methodology is available to detect the disease at an early stage.

Keywords: Parkinson's Disease, Classification, Random Forest, Support Vector Machine, Machine Learning,

1. Introduction

Parkinson disease (PD) is a neurological disorder based on dopamine receptors. Parkinson disease mostly causes problems in moving around. It can cause a person to move very slowly. Parkinson is a progressive neurological condition, which is characterized by both motor (movement) and non-motor symptoms. Apart from many common symptoms each person will experience and demonstrate an individual presentation of the condition. A person with Parkinson disease appears stiff or rigid. At times, a person with Parkinson disease may appear to suddenly "freeze up" or be unable to move for a short period of time. Parkinson disease is a progressive neurodegenerative condition resulting from the death of the dopamine containing cells of the substantia nigra. There is no consistently reliable test that can distinguish Parkinson disease from other conditions that have similar clinical presentations. The diagnosis is primarily a clinical one based on the history and examination.

People with Parkinson disease classically present with the symptoms and signs associated with Parkinsonism, namely hypokinesia (i.e. lack of movement), bradykinesia (i.e. slowness of movement), rigidity (wrist, shoulder and neck.) and rest tremor (imbalance of neurotransmitters, dopamine and acetylcholine). Parkinsonism can also be caused by drugs and less common conditions such as: multiple cerebral infarction, and degenerative conditions such as progressive supra nuclear palsy (PSP) and multiple system atrophy (MSA).

Although Parkinson disease is predominantly a movement disorder, other impairments frequently develop, including psychiatric problems such as depression and dementia. Autonomic disturbances and pain may later ensue, and the condition progresses to cause significant disability and handicap with impaired quality of life for the affected person. Family and carers might get affected indirectly.

Neurodegenerative disorders are the results of the progressive tearing and neurons loss in different areas of the nervous system. Neurons are the functional unit of brain .They are contiguous rather than continuous. A good healthy looking neuron as shown in fig 1 has extensions called dendrites or axons, a cell body and a nucleus that contains our DNA. DNA is our genome and hundred billion neurons contains our entire genome which is packaged into it .When a neuron get sick, it loses its extension and hence its ability to communicate which is not good for it and its metabolism become low so it starts to accumulate junk and it tries to contain the junk in the little packages in little pockets .When things become worse and if the neuron is a cell culture it completely loses its extension, becomes round and full of the vacuoles.

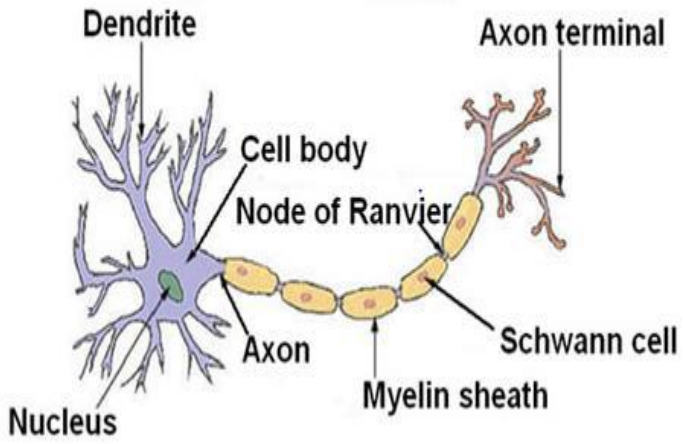

Fig 1: Structure of neuron present in human brain

2. Related Work:

Indira R. et al. (2014) have proposed an automatically machine learning approach and detected the Parkinson disease on behalf of speech/voice of the person. The author used fuzzy C-means clustering and pattern recognition based approach for the discrimination between healthy and parkinson disease affected people. The authors of this work have achieved 68.04% accuracy, 75.34% sensitivity and 45.83% specificity.

Indira R. et al. (2014) have proposed a back propagation based approach for the discrimination between healthy and parkinson diseases affected peoples with the help of artificial neural network. Boosting was used by filtering technique, and for data reduction principle component analysis was used.

Geeta R. et al. (2012) have investigated and performed the feature relevance analysis to calculate the score to classify the Parkinson diseases Tele-monitoring dataset and dataset comparison classes Motor-UPDRS and Total-UPDRS (Unified Parkinson Disease Rating scale).

Rubén A. et al. (2013) proposed a five different classification paradigms using a wrapper feature selection scheme are capable of predicting each of the class variables with estimated accuracy in the range of 72–92%. In addition, classification into the main three severity categories (mild, moderate and severe) was split into dichotomy problems where binary classifiers perform better and select different subsets of non-motor symptoms.

Betala E. et al. (2014) proposed a SVM and k-Nearest Neighbour (k-NN) Tele-monitoring of PD patients remotely by taking their voice recording at regular interval. The age, gender, voice recordings taken at baseline, after three months, and after six months are used as features are assessed. Support Vector Machine was more successful in detecting significant deterioration in UPDRS score of the patients.

A.Tsanas et al. (2011) proposed feature selection, random forest and support vector machine used to discriminate PD from healthy controls. The author achieved overall 99% classification accuracy using only ten dysphonia features.

A.Tsanas et al. (2011) proposed a nonlinear signal approach large dataset (dataset are voice/speech recorded without requiring physician presence in the clinical) apply wide range known speech signal algorithm. This work was performed using nonlinear regression and classification algorithm, and support visibility of frequent, remote, cost-effective, accurate UPDRS telemonitoring based on self-administered speech tests.

A. Sharma et al. (2014) proposed artificial neural network, pattern recognition and support vector machine. It is used to support the experts in the diagnosis of Parkinson disease. The dataset of research was composed of a range of biomedical voice signals healthy people and parkinson disease accuracy was obtained around 85.294%.

Khemphila et al. (2012) proposed a Multi-Layer Perceptron (MLP) with Back-Propagation learning algorithm was used to effective diagnosis Parkinson's disease (PD). Medical diagnosis was done by doctor's expertise and experience. But still cases are reported of wrong diagnosis and treatment. Patients are asked to take number of tests for diagnosis. In many cases, not all the
tests contribute towards effective diagnosis of a disease. The artificial neural networks are used to sort the diagnosis of patients. This work predict the accuracy model training dataset 91.45%, and the validation data set was 80.77%.

Revett et al. (2009) proposed jitter, shimmer, fundamental frequency, harmonics/noise ratios, descriptive statistics, and correlational factors (non-linear dynamic analysis) using all 22 feature, and a binary decision class ('0'is healthy and '1' is IPD decision class). The testing and training set are classified and an ROC and confusion matrix was generated to examine the accuracy of the
classification process. Predict of accuracy shows 100%.

Shahbakhi et al. (2014) presented that a Genetic Algorithm (GA) and SVM were used for classification between healthy and people with Parkinson. Voice signals that 14 features were based on F0 (fundamental frequency or pitch), jitter, shimmer and noise to harmonics ratio, which are main factors in voice signal. Results show that classification accuracy 94.50, 93.66 and 94.22 per 4, 7 and 9 optimized features respectively.

Chen et al. (2012) proposed mainly two classifier Nested-RF and Nested-SVM classifier. Five datasets of cancer (brain cancer, colon cancer, DLBCL, leukemia, prostate cancer) and one disease (Parkinson's) dataset were used to evaluate the performance of the proposed classifiers. Parkinson's disease classification, the Nested-SVM classifier showed the superior performance with the accuracy up to 93% that was 20% more than the results from other three classifiers.

Bocklet et al. (2011) proposed a SVM and Correlation base classification performed to speech/voice of a person was affected by Parkinson disease automatic detection of Parkinson disease based on articulation, voice, and prosodic evaluations. The best results (90.5% recognition rate and 0.97 AUC).

R. Das et al. (2010) have proposed neural networks, Data Mining Neural analysis, and regression analysis and decision trees made a comparative study on Parkinson disease data set with regard to with the Presented results of classification accuracy of 92.9%, 84.3%, 88.6% and 84.3% respectively. To the classification method was diagnosis Parkinson disease based on the SAS software.

Ene M. et al. (2008) proposed a probabilistic neural network (PNN) variant to discriminate between healthy people and people with Parkinson's disease. Three PNN types are used in this classification process, related to the smoothing factor search: incremental search (IS) Monte Carlo search (MCS) and hybrid search (HS). The accuracies reaching run between 79% and 81% for new, undiagnosed patients.

Cam M. et al. (2008) proposed a parallel distributed neural network with two hidden layers, boosted by the use of filtering and a majority voting system to distinguish between the people who have normal vocal signals and who suffer from Parkinson's disease. To perform the boosting by filtering technique, we the Training and Testing stage, the accuracy achieved by > 90.

Caglar et al. (2010) proposed ANN (Two types of the ANNs were used for classification: Multilayer Perceptron (MLP) and Radial Basis Function (RBF) Networks) and Adaptive Neuro-Fuzzy Classifier (ANFC) with linguistic hedges to discriminate between healthy people and people with PD. Adaptive Neuro-Fuzzy Classifier with linguistic hedges gave the best recognition results with %95.38 training and %94.72 testing classifying performance indeed.

Ali Saad et al. (2013) proposed a Bayesian Belief Network (BBN) to find the freezing of Parkinson disease patients and used a Video dataset available online extracted from real Parkinson disease patients though walking and having freezing periods. Each file was composed of a matrix that contains measurement data of the three sensors in x, y and z directions. Weather Freezing of Gait (FoG) occurred or not. These annotations was labelled by synchronizing the data by a video that recorded each patient run and results appeared when testing the models Bayesian Naïve Classifier (BNC) classifier.

Cho, C. at al. (2009) proposed system utilizes an algorithm combining principal component analysis (PCA) with linear discriminant analysis (LDA). We propose a gait analysis system which can detect the gait pattern of Parkinson's disease using computer vision. This system comprises three main parts: pre-processing, training and recognition. Experimental results showed that LDA had a recognition rate for Parkinsonian gait of 95.49%.

Rusz J. et al. (2011) proposed applied support vector machine to find the best combination of measurements to differentiate Parkinson disease from healthy subjects. This method leads to overall classification performance of 85%. Admittedly, we have found relationships between measures of phonation and articulation and bradykinesia and rigidity in Parkinson disease. In the acoustic analysis can ease the clinical assessment of voice and speech disorders, and serve as measures of clinical progression as well as in the monitoring of treatment effects.

Can M. et al. have proposed boosting committee machine to detect Parkinson disease for dataset containing sick and healthy people by the artificial neural network. The filtering techniques used for the neural networks with back propagation, they majority voting scheme. Out of 195 samples, 75.4% are Parkinson's disease type and the remainder was of healthy character.

Kapoor T. et al. (2011) proposed speech recognized by Mel-frequency cepstral coefficients (MFCC) and Vector Quantization (VQ). The MFCC uses speech analysis frames in signal to frequent domain and Vector Quantization was the codebook of lowest distortion was calculated. The 20 phonation's used for normal speech and patient with Parkinson's disease. Vector Quantization result with codebook in normal voice and voice of Parkinson disease rate in classifier 90% and 95% respectively.

Wu, S et al. (2011) proposed regression, decision tree and neural network analysis to analyse the databank of Parkinson disease for error probability calculated. The result was logistic regression, classification and neural network analysis error probability by 5.15%, 8.47% and 23.73% respectively.

Sellam V. et al. (2014) proposed classification of pathological voice from normal voice was implemented using Support Vector Machine (SVM) and Radial Basis Functional Neural Network (RBFNN). The normal and pathological voices of children are used to train and test the classifiers. The speech signal was then analysed in order to extract the acoustic parameters such as the Signal Energy, pitch, formant frequencies, Mean Square Residual signal, Reflection coefficients, Jitter and Shimmer. Show the classification accuracy of RBFNN 91% and SVM 83%.

Chen, H et al. (2013) proposed FKNN-based system was compared with the support vector machines (SVM) based approaches predict to dataset composed of a range of biomedical voice measurements from 31 people, 23 people with Parkinson disease. The best classification accuracy (96.07%) obtained by the FKNN based system using a 10-fold cross validation method can ensure a reliable diagnostic model for detection of Parkinson disease. Salvatore et al. (2014) proposed a supervised machine learning algorithm based on Principal Components Analysis as feature extraction technique and Support Vector Machines to predict of individual differential diagnosis of Parkinson's disease (PD) and Progressive Supranuclear Palsy (PSP) for Magnetic Resonance Images (MRI dataset). Predict of the Parkinson disease (PD) versus Controls, Progressive Supranuclear Palsy (PSP) versus Controls and Progressive Supranuclear Palsy (PSP)
versus Parkinson disease (PD) the Overall Accuracy (Specificity/Sensitivity) were 83.2 (81.9/85.4), 86.2 (92.1/82.9) and 84.7 (87.5/83.8)% for binary labelled groups, respectively.

Przybys Z. et al. (2014) proposed a Reflexive saccades measurements and classifications to predict individual patients and small patient popular significant measure effects are plotted the movement lines in the phase space as changes of the right hip x-angles as a function of the left hip angle changes during three steps of stable walking and found different types of attractor changes as the effect of treatment and motivations.

Morales et al. (2013) proposed naïve Bayes, filter selection naıve Bayes (FSNB), naıve Bayes correlationbased with feature subset selection method (CFS-NB) and support vector machines (SVM) analysing pairs of classes (PDD vs. PDCI, PDD vs. PDMCI, PDMCI vs. PDCI), and (PDD vs. PDMCI vs. PDCI) on the different types comparison symptom of Parkinson disease. CFS-NB for (PDD vs. PDCI) found the highest accuracy, Sensitivity and Specificity of 97%, 93.33% and 100% respectively.

R. Ramani et al. (2011) proposed a many type classification of data mining approaches SVM, KNN, Random tree, Partial Least Square Regression (PLS) etc. to predict dataset biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). To the filtering was applied to the algorithms for better classification purpose, smallest number of qualities with which the better classification was selected and achieved. The Random Tree forms the classification based on three typical features to gain the zero error rates.

Chen, A. et al. (2013) proposed Nested–Random Forest (Nested-RF) classifier and Nested–Support Vector Machine (Nested-SVM) classifier for predict of Five datasets of cancer (brain cancer, colon cancer, DLBCL, leukemia, prostate cancer) and one disease (Parkinson's) datasets. Nested-SVM classifier was applied to the Parkinson's disease dataset

average accuracy, sensitivity, and specificity can reach 93%, 90%, and 93% respectively.

Yadav, G et al. (2009) have proposed classifier, statistical classifier, and support vector machine classifier to discriminate healthy people and Parkinson disease. SVM classifier provides the accuracy of 76%, sensitivity of 97% and specificity of 13%.

Azad, C., et al. (2013) proposed prediction model tree based classification model decision tree, ID3 and decision stumps are used for training and testing the effectiveness many symptoms that lead to Parkinson's disease such ageenvironmental factor, trembling in the legs, arms, hands, impaired speech articulation and production difficuties. Decision tree, ID3 and decision stumps our prediction model provides accuracy 85.08%, 75.33% and 83.55% or classification error 14.92%, 24.67% and 16.45% respectively.

Bouchikhi et al. (2013) proposed Neural Networks (ANN), Data Mining neural, Regression and Decision Tree for effective diagnosis dataset to discriminate healthy people and Parkinson disease. SVM classifier shows performance 97.22% specificity, 95.83% sensitivity and the total classification accuracy of 96.88%. New feature classification optimal Fuzzy k-nearest neighbour (FKNN) model was 96.07% accuracy.

Kihel, B. et al. (2011) proposed Clonclas and Probabilistic Neural Network (PNN) to discriminate between healthy and people with Parkinson's disease (PWD) Taking inspiration from natural immune systems, we try to grab useful properties such as automatic recognition, memorization and adaptation. The developed algorithms have as a base the algorithm of training biomedical inspired Clonclas.

Ma, C. et al. (2014) proposed a novel hybrid method named Kernel-Based Extreme Learning Machine with Subtractive Clustering Features Weighting (SCFWKELM) significantly outperforms SVM, KNN, and extreme learning machine (ELM) approaches for Parkinson disease dataset was to discriminate healthy people from those with Parkinson disease. given the results of various medical tests carried out on a patient achieved highest classification results reported so far via 10-fold cross validation scheme, with SVM-based, KNNbased, and ELM-based accuracy of 99.49%, the sensitivity of 100%, the specificity of 99.39%, AUC of 99.69%, the f -measure value of 0.9964, and kappa value of 0.9867.

Hazan, H et al. (2012) have been A novel hybrid method named Kernel-Based Extreme Learning Machine with Subtractive Clustering Features Weighting Approach (CFW-KELM) to discriminate healthy people from people with Parkinson disease. Experimental results have demonstrated that the proposed SCFW-KELM significantly outperforms SVM-based, KNN-based, and ELM-based approaches and other methods in the literature and achieved highest classification results reported so far via 10-fold cross validation scheme, with the classification accuracy of 99.49%, the sensitivity of 100%, the specificity of 99.39%, AUC of 99.69%, the f -measure value of 0.9964, and kappa value of 0.9867.

Sriram, T et al. (2013) proposed SVM, k-NN, Random Forest and Naïve Bayes voice to dataset for Parkinson disease. The class column represents "status" which was set to 0 for healthy and 1 for PD. SVM, k-NN, Random Forest and Naïve Bayes was the prediction accuracy 88.9%, 88.9%, 90.26 and 69.23% respectively.

Prashanth et al. (2014) proposed Support Vector Machine (SVM) and classification tree methods are use olfactory loss feature from 40-item University of Pennsylvania Smell Identification Test (UPSIT) and Sleep behaviour disorder feature from Rapid eye movement sleep Behaviour Disorder Screening Questionnaire (RBDSQ), obtained from the Parkinson's Progression Marker's Initiative (PPMI) database. Support Vector Machine sleep Behaviour Disorder (SVM-RBD) was predicted of the best accuracy 85.48%, sensitivity 90.55% and specificity 74.58%.

Amit S. et al. (2014) proposed an approach to classification and kind of Parkinson's patients using their postural response and analysing it using a L2 norm metric in conjunction with support vector machines. Twenty four patients were valued before and after medication. Each patient suffered following analysis protocols for the valuation of their postural balance: First, Eyes Open on Force platform (a firm surface) (E0) and second, Eyes Open on Foam placed on Force platform (FO).The classification of subjects with dyskinesia when standing on a firm surface with eyes open was improved from 66% to 77%.

A.H. et al. (2012) proposed Bayesian Networks, Regression, Classification and Regression Trees (CART), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) for proposing a decision support system for diagnosis of Parkinson's disease. Parkinson's disease, the disorder also commonly causes a slowing or freezing of movement. The proposed system achieved an accuracy of 93.7% using classification and regression tree.

Kaya, E. et al. (2011) proposed discretization method, support vector machines, C4.5, k-nearest neighbours and Naive Bayes classier methods are used to classify the dataset. The dataset was classified using the features discretizeted and non-discretizated in order to show the effectiveness of discretization on diagnosis of Parkinson's disease.

Nivedita C et al. (2013) proposed artificial neural network (ANN) with back propagation to classify neurodegenerative disorders according to symptoms. The clinical symptoms of neurodegenerative disorders have been identified as six major classes Memory problems, Communication problems, Personality changes, idiosyncratic behaviours, Loss of voluntary control and Common health problems. Artificial neural network (ANN) was prediction of overall performance of 96.42%.

Farhad S. et al. (2013) proposed Multi-Layer Perceptron (MLP) with back-propagation learning algorithm and Radial Basis Function (RBF) and Artificial Neural Networks ANN) were used to differentiate between clinical variables of samples (N = 195) who were suffering from Parkinson's

disease and who were not. MLP and RBF classification accuracy 93.22% and 86.44% respectively for the data set.

Chen, A. et al. (2012) proposed Random forests (RF) classifier, Support Vector Machine (SVM) classifier, Genetic Algorithm–Random Forests (GA-RF) classifier, and Genetic Algorithm–Support Vector Machine (GASVM) classifier to effect diagnose and classify the Parkinson's disease. GA-SVM classifier significantly improves accuracy (69% to 94%), sensitivity (60% to 92%), and specificity (70% to 95%).

Wu, D. et al. (2010) proposed radial basis function neural network (RBFNN) based on particle cloud optimization (PSO) and principal component analysis (PCA) with Local Field Potential (LFP) data recorded via the stimulation electrodes to predict activity related to tremor onset. RBFNN, PCA + RBFNN and PCA + PSO + RBFNN to the predict accuracies 89.91%, 88.92% and 88.92% respectively.

Luukka P. et al. (2011) proposed fuzzy entropy based feature selection combined with similarity classifier; we achieved to reduce the computational time and simplify the data set. Data set was composed of a range of biomedical voice measurements from healthy people and people with Parkinson's disease (PD). Mean classification accuracy with Parkinson's data set being 85:03%.

Salhi L. et al. (2008) proposed a method that uses wavelet analysis to extract a feature vector from speech samples, which was used as input to a Multilayer Neural Network (MNN), three layer feed forward network with sigmoid activation and Back Propagation Algorithm (BPA) classifier. The classification rate was between 80% and 100%.

Max A. et al. (2009) have been proposed support vector machine (SVM) valuation of the practical value of existing traditional and non-standard measures for discriminating healthy people from people with Parkinson's disease (PWD) by detecting dysphonia.

Zhang, J. et al. (2008) proposed an increased cerebrospinal fluid (CSF) τ and decreased amyloid (A) β42 to validate as biomarkers of Alzheimer disease and no validate biomarker for Parkinson disease. Predicted of all subjects medical history, family history, physical and neurologic examinations by clinicians who specialize in movement disorders or dementia, laboratory tests, and neuropsychological. Analysis for 90 control subjects (95%), 36 patients with likely Alzheimer disease (75%), and 38 patients with likely Parkinson disease (95%).

Saad A. et al. (2011) proposed a based on unsupervised learning of a probabilistic graphical model Bayesian Belief Network (BBN) based on Expectation Maximization (EM) algorithm. Prediction of mixed acquisition system of electronic pen and speech signals are performed through voice and handwriting. This work predict of grouping based analysis of voice and handwriting.

Ozcift A. et al. (2011) proposed computer-aided diagnosis (CADx) systems to improving the accuracy. Rotation forest (RF) collective classifiers of 30 machine learning algorithms correlation based feature selection (CFS) algorithm and Rotation forest prediction to diabetes, heart and Parkinson's datasets. RF classifier predict the accuracy (ACC), kappa error (KE) and area under the receiver operating characteristic (ROC) curve (AUC) of 74.47%, 80.49% and 87.13% respectively.

Yahia A. et al. (2014) proposed classification algorithm based on Naïve Bayes and K- Nearest Neighbours (KNN) using Parkinson speech dataset with multiple types of sound recordings to prediction voice signal find the Parkinson disease or healthy people. K- Nearest Neighbours performed accuracy 80% and Naïve Bayes classifier performed an accuracy of 93.3% sensitivity 87.5%, and specificity 100%.

Shamli and Sathiyabhama proposed multi-classifier system, i.e., based on Big Data analytics to improve predictive performance and efficient time to answer cost-effective actions. The author introduced Big Data with its characteristics and Big Data analytics with their types as Descriptive, Predictive and Prescriptive in healthcare industries. Dopamine, a neurotransmitter, generated by brain cells, is responsible to send signals to other brain cells to control muscle activity. The degeneration of dopamine-producing brain cells causes PD. For analysis purposes, voice dataset of PD is collected from UCI machine learning library. By implementing multiple predictive models to disease datasets, multiple accuracies and results of different classifiers are acquired. C4.5, SVM, and ANN give better results than other machine learning algorithms. After comparing the results of these classifiers, best results are chosen for the final decision. This approach helps organizations to analyze their large datasets quickly and efficiently with maximum accuracy.

Azad et al. explored a predictive model for PD that is based on decision tree algorithm. They introduced PD, a second most common neurodegenerative disease with its symptoms, possible complications, and risk factors associated with it. Various applications of data mining are used for classification purposes that are decision tree, attribute selection measures, ID3 and decision stumps. Their dataset (have 197 instances) is taken from UCI repository and built up from the data of 31 people. For performance analysis, two parameters accuracy and classification error are used. For validation, 10-fold cross-validation technique is used that gives the unbiased outcome. They found that decision tree algorithm performs best and gives the best accuracy and less classification error than other algorithms in their experimental results.

Sriram et al. proposed a method for diagnosis of PD using its voice dataset. This voice dataset is built up from the voice of 31 people among which 23 people are affected by PD. This dataset contains 5875 instances and 26 attributes. In their experiment for statistical analysis, classification, evaluation, visualization, and unsupervised methods Weka V3.4.10 and Orange V2.0b software are used. They achieved the best accuracy 90.2% from Random Forest algorithm.

Richa Mathur et. al (2019), Parkinson disease, the second most common neurological disorder that causes significant disability, reduces the quality of life and has no cure. Approximately, 90% affected people with Parkinson have speech disorders. The medical dataset contains heterogeneous

data in the form of text, numbers, and images that can be mined. Big Data has the potential to give valuable information after processing that can be discovered through deep analysis and efficient processing of data by decision-makers. Data mining is the process of selecting, extracting, and modeling the unknown hidden patterns from large datasets. Machine learning algorithm (MLA) can be used for early detection of disease to increase the chances of elderly people's lifespan and improved lifestyle with Parkinson. In this work, we use various MLAs that can help in improving the performance of datasets and play a vital role in making the early prediction of disease at right time. After comparison of these algorithms, we choose the most effective one in terms of accuracy. From our experimental results, it is analyzed that the accuracy obtained from the combined effect of KNN algorithm with ANN is better as compared to other algorithms.

In this work, the authors propose a convolutional neural network based multimodal disease risk prediction (CNNMDRP) .This algorithm overcome the drawbacks of (CNN-UDRP) convolutional neural network based unimodal disease risk prediction. This algorithm uses both the structured and unstructured data of a hospital. None of the existing algorithm can work on both the structured and unstructured data. Its accuracy is about 94.8 In this work, the researchers present how artificial intelli-gence applied to medical field for the efficient diagnosis. For that purpose they use a k nearest neighbours algorithm and they check the accuracy of the algorithm with the help of UCI machine learning repository datasets. They had to generate patients input and test data for diagnosis. They use a real patient data. They add a additional training sets allow more medical conditions to be classified with the minimal no of changes to the algorithm. [2] In this work, they applying a machine learning techniques by using EMC'S from outpatients department and the algo-rithm are based on a DNN AND DBDT, It can be achieve a high UAR for predicting the future stroke prediction. It provides a several advantages like high accuracy, fastest prediction, and consistency of results. DNN algorithm also requires a lesser amount of data. DNN algorithm can achieves a optimal results by using a lesser amount of a patient data than compared to the GDBT algorithm. [3] In this work, distributed computing environment processing the large volume of a data is done based on Map Reduce. To find the accuracy of a patient data the classification is used. In this work more focused on find out the nearest accuracy of a classifiers. The CART model and random forest is built for the data and accuracy of the classifier is found. By using the random forest algorithm they can found the more nearest accuracy of the prediction. The prediction analysis helps to the doctors to identify the patient's admissions on to the hospital. Predictive model using scalable random forest classification which can accurately give the result rate of risk. [4] In this work for heart disease prediction they use a Neavi Bayes and Decision tree algorithm. They used a PCA to reduce the no of attributes, after reducing the size of the datasets; SVM can outperform a Neavi Bayes and Decision tree. SVM can also be used for prediction of hearts disease. The main goal of this work is to predict the diabetics disease. Using a WEKA data mining tools. Data mining is very useful techniques used by health care sector for classification of disease. The aim of this work is to study supervised machine learning algorithm to predict the heart disease. [5] In this work the data mining and the big data in the healthcare sector is introduced. Machine learning algorithm has been used to study the healthcare data. The continuous increase of data in a healthcare. Several countries are spending a lot of resources, scientist leads to fix the problem of storage and organization of data the data mining will help exploitation complexity of the data and find out the new result this work is based on the use of data mining and big data in the healthcare sector. [6] Traditional wearable devices have various shortcomings, such as comfortableness for long-term wearing, and insufficient accuracy, etc. Thus, health monitoring through tradi-tional wearable devices is hard to be sustainable. In order to obtain healthcare big data by sustainable health mon-itoring, we design Smart Clothing, facilitating unobtrusive collection of various physiological indicators of human body. To provide pervasive intelligence for smart clothing system, mobile healthcare cloud platform is constructed by the use of mobile internet, cloud computing and big data analytics. This work introduces design details, key technologies and practical implementation methods of smart clothing system.

In 2010, Apache Hadoop sharp big data as "datasets which could not be apprehended, succeeded, and managed by general computers within an okay scope." On the basis of this definition, in May 2011, McKinsey & Company, a global accessing help said Big Data as the next edge for improvement, war, and yield. Big data shall callous such datasets which could not

be attained, succeeded and stored by standard database software. This classification includes two associations: First, datasets dimensions that obey to the usual of big data are shifting, and may cultivate over time or with scientific developments. Second, datasets measurements that adapt to the ordinary of big data in unalike submissions contrast from each other.

[2] Clinical data recounting the phenotypes and dealing of patients denotes an underused data font that has much bigger research likely than is currently grasped. Mining of electrical health records has the facility to form a new patient-stratification doctrines and for tight fitting unknown disease links. Mixing EHR data with genetic data will also give a more kind of genotype-phenotype affairs. However, a wide series of permitted, ethical, and methodological reasons presently hold back the organized confession of these data in electrical health histories and their excavating. Here, it consider the likely for furthering medical examination and experimental care using EHR data and the tasks that must be dazed before this is a truth.

[3] The medical resources of many countries are limited. For example, in China, the growth of medical resources is not balanced that 80% people are living in areas with inadequate

medical resources while 80% medical resources are allocated at the big cities. Construction of big health application system by successfully mixing medical health resources using smart depots, health Internet of Things (IoT), big data and cloud computing is the vital way to resolve the above difficulties. Big health is a talented industry, which is characterized by people-center, managing a person's health from birth to decease, from anticipation to rehabilitation and involving industry from administration to market. The field of big health covers health goods field (including the drugs, medical devices, elder goods), health service field (including medical services, income services, mobile healthcare), fitness real estate field (including pension, healthcare) and health finance field (including health protection and other financial products).

[4] Chinese herbal products (CHPs) are commonly developed for patients with hyperlipidemia in traditional Chinese medicine (TCM). Since hyperlipidemia and connected sickness are public topics worldwide, this training discovered the drug shapes and occurrences of CHPs for giving patients with hyperlipidemia. Traditional Chinese medicine (TCM) has become common as a healing for central indicators in patients with hyperlipidemia. This drill likely to study the treatment patterns of TCM for patients with hyperlipidemia. The study population was recruited from a random-sampled troop of 1,000,000 folks from the National Wellbeing Insurance Exploration Record between. It recognized 30,784 fatality visits linked with hyperlipidemia judgment and collected these medical records. Overtone rules of facts withdrawal were led to moveable the co-prescription plans for Chinese herbal products (CHPs).

[5] In this work, it witness the use of recurrent neural networks (RNNs) with the situation of search-based operational publicity. It practice RNNs to map equally queries and ads to real valued vectors, by means of which the significance of a given (query, ad) couple can be simply calculated. On upper of the recurrent neural networks, it familiarize a novel consideration network, which studies to assign attention scores to different word locations according to their intent importance (hence the name Deep Intent). Later by this method, the path output of a arrangement is computed by a weighted sum of the hidden states of the RNN at each word according their attention scores. The system achieve end-to-end exercise of together the RNN and attention system below the guidance of user click logs. These worker click logs are sampled from a commercial search engine. It demonstrate that in most cases the attention network improves the quality of learned vector representations, evaluated by AUC on a physically labeled dataset. And furthermore, it highlight the effectiveness of the learned attention nicks from two aspects as: query rewriting and a modified BM25 metric. The system illustrate that using the learned attention scores, one will be able to produce sub-queries that would be of better qualities than those of the state-of-the-art methods. In count, by regulating the term occurrence with the care scores in a normal BM25 formula, one is bright to improve its performance evaluated by AUC.

[6] Traditional wearable devices have various drawbacks, such as uncomfortableness for long-term wearing, and insufficient accuracy, etc. Thus, health monitoring through traditional wearable devices is hard to be sustainable. In order to obtain and manage healthcare big data by sustainable health nursing, the system design "Smart Clothing", enabling unobtrusive collection of various physiological indicators of human body. To offer persistent cleverness for smart clothing erection, mobile healthcare cloud stand is constructed by the usage of mobile internet, cloud computing and big data analytics. This work announces design facts, key tools and applied implementation methods of smart dress system. Typical claims powered by smart clothing and big data clouds are presented, such as medical backup response, emotion care, disease diagnosis, and real-time tangible interaction.

[7] In this it extant a new deep learning manner Bi-CNN-MI for paraphrase identification (PI). Created on the vision that PI needs associating two sentences on many heights of granularity, it learn multigranular decree images using convolutional neural network (CNN) and model boundary features at each level. These topographies are then the input to a logistic classifier for PI. All limits of the model (for embeddings, convolution and classification) are straight optimized for PI. To address the lack of training data, the system pretrain the network in a novel method using a language modeling task. Results on the MSRP corpus surpass that of earlier NN competitors.

[8]Does the estimate of lung cancer using the double dispensation system. The image dispensation system is familiarized into the double for early prophecy. The challenging in this progression is recognition of tiny nodes which comprehends early cancer finding. The unstipulated knobs in lungs can be spotted using ridge recognition algorithm.

[9] It proposed a system that integrates different datum such as gene information, DNA methylation, and miRNA. In this work, the model has combined multiple kernel learning methods and dimensionality reduction.

[10] On the available data mining algorithms to classify the data and extract the knowledge from it. It discusses about the difficulties in classification, segmentation, extraction and selection. It compares the different algorithms like Support Vector Machine, Naïve Bayesian classification, Rough set theory, Decision Tree.

Big Data analytics plays a huge role in the healthcare industry, as these data are scattered everywhere, big, and complex in nature. In this work, we discuss early stage prediction of Parkinson disease for that we presented a methodology of data

mining using Weka tool for classifying disease dataset. We use various MLAs for classifying our experimental data that indicate the combined effect of ANN algorithms with KNN which is better as compared when we use other algorithms. The system detects the maximum accuracy of the multi-classifier, and their result predicts the disease at its early stage. We discuss the comparative analysis and calculate the overall performance measures in terms of precision, recall, and f-measure. In future, effective optimization techniques can be used to achieve better accuracy and cost-effective interventions for Parkinson disease. Also, limited data is available that describes the real potential of early PD treatment which requires more research to explore the real impact of early treatment.

3. Conclusion:

Parkinson disease directly affects the central nervous system and mainly affects the motion of the patients. The main cause of the Parkinson disease is the drop of dopamine due to the damages in the nerve cell. The primary symptoms of the disease are stiffness, impaired movement, tremors etc. it is a non-curable disease only can be controlled by proper medical observation.

References:

[1] 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).

[2] Rustempasic, I., & Can, M. (2013). Diagnosis of Parkinson's disease using principal component analysis and boosting committee machines. SouthEast Europe Journal of Soft Computing, 2(1).

[3] Ramani, D. R. G., Sivagami, G., & Jacob, S. G. (2012). Feature Relevance Analysis and Classification of Parkinson Disease Tele- Monitoring Data Through Data Mining Techniques. International Journal of Advanced Research in Computer Science and Software Engineering, 2(3).

[4] Armañanzas, Ruben, Bielza, C., Chaudhuri, K. R., Martinez-Martin, P., & Larrañaga, P. (2013). Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach. Artificial intelligence in medicine, 58(3), 195-202.

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

[6] Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., & Ramig, L. O. (2012). Novel speech signal processing algorithms for highaccuracy classification of Parkinson's disease. Biomedical Engineering, IEEE Transactions on, 59(5), 1264-1271.

[7] Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2011). Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. Journal of the Royal Society Interface, 8(59), 842-855.

[8] Sharma, A., & Giri, R. N. (2014) Automatic Recognition of Parkinson's disease via Artificial Neural Network and Support Vector Machine.

[9] Khemphila, A., & Boonjing, V. (2012). Parkinsons disease classification using neural network and feature selection. World Academy of Science, Engineering and Technology, 64, 15-18.

[10] Revett, K., Gorunescu, F., & Salem, A. M. (2009, October). Feature selection in Parkinson's disease: A rough sets approach. In Computer Science and Information Technology, 2009. IMCSIT'09. International Multiconference on (pp. 425-428). IEEE.

[11] 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.

[12] Chen, A. H., Lin, C. H., & Cheng, C. H. New approaches to improve the performance of disease classification using nested– random forest and nested–support vector machine classifiers.
Cancer, 2(10509), 102.

[13] Bocklet, T., Noth, E., Stemmer, G., Ruzickova, H., & Rusz, J. (2011, December). Detection of persons with Parkinson's disease by acoustic, vocal, and prosodic analysis. In Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on (pp. 478-483). IEEE.

[14] Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with Applications, 37(2), 1568-1572.

[15] Ene, M. (2008). Neural network-based approach to discriminate healthy people from those with Parkinson's disease. Annals of the University of Craiova-Mathematics and Computer Science Series, 35, 112-116.