PARKINSON'S DISEASE DETECTION

Seminar (IT290) Report

Submitted in partial fulfilment of the requirements for the degree of

BACHELOR OF TECHNOLOGY In

INFORMATION TECHNOLOGY

By

RAGHAVENDRA PARIPELLI (201IT244)
ADITYA PRAJAPATI (201IT104)
P SAI MANOJ REDDY (201IT241)

DEPARTMENT OF INFORMATION TECHNOLOGY

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALORE -575025

MAY, 2022

D E C L A R A T I O N

I hereby declare that the Seminar (IT290) Report entitled **PARKINSON'S DISEASE DETECTION** which is being submitted to the National Institute of Technology Karnataka Surathkal, in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in the department of Information Technology, is a bonafide report of the work carried out by me. The material contained in this seminar report has not been submitted to any University or Institution for the award of any degree.

Signature of the Student

RAGHAVENDRA PARIPELLI (201IT244)

ADITYA PRAJAPATI (201IT104)

P SAI MANOJ REDDY (201IT241)

Place : NITK, Surathkal
Date : 05-05-2022

CERTIFICATE

This is to certify that the Seminar entitled "**PARKINSON'S DISEASE DETECTION** " has been presented by RAGHAVENDRA PARIPELLI (201IT244), ADITYA PRAJAPATI (201IT104),P SAI MANOJ REDDY (201IT241) students of IV semester B.Tech. (IT), Department of Information Technology, National Institute of Technology Karnataka, Surathkal, on , during the even semester of the academic year 2021 - 2022, in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Information Technology.

Guide Name: BHAWANA RUDRA MA'AM

Signature of the Guide with Date
Place: NITK, Surathkal
Date: 05-05-2022

ABSTRACT

Parkinson's disease is a type of disease caused by the loss of dopamine-producing cells in the brain. As the amount of dopamine decreases, the symptoms of Parkinson's disease emerge. Parkinson's disease is a slow-developing disease, and symptoms such as hands, arms, legs, chin and face tremors are increasing over time. As the disease progresses, people may have difficulty in walking and speaking. There is no definitive treatment for Parkinson's disease; however, with the help of some drugs, the symptoms of the disease can be reduced. Although there is no definitive treatment for Parkinson's disease, the patient can continue his normal life by controlling the problems caused by the disease. At this point, it is important to prevent early detection and progression of the disease. In this study, classification methods likeSupport Vector Machine are compared in order to predict Parkinson's disease. It is concluded that classification results which are attained with expanded features outperform the classification results attained with the original features of the data.

TABLE OF CONTENT

INTRODUCTION

A. Background

Parkinson Disorder, the second most common disease after Alzheimer’s and a major public health problem.

Parkinson's disease is a progressive nervous system disorder that affects movement. Symptoms start gradually, sometimes starting with a barely noticeable tremor in just one hand. Tremors are common, but the disorder also commonly causes stiffness or slowing of movement.In the early stages of Parkinson's disease, your face may show little or no expression. Your arms may not swing when you walk. Your speech may become soft or slurred. Parkinson's disease symptoms worsen as your condition progresses over time.

Although Parkinson's disease can't be cured, medications might significantly improve your symptoms. Occasionally, your doctor may suggest surgery to regulate certain regions of your brain and improve your symptoms.

The main deficits of PD speech are loss of intensity, monotony of pitch and loudness, reduced stress, inappropriate silences, short rushes of speech, variable rate, imprecise consonant articulation, and harsh and breathy voice (dysphonia). The range of voice related symptoms is promising for a potential detection tool because recording voice data is non-invasive and can be done easily with mobile devices.

B. Technological Approach

At the present time, computer based decision and diagnosis systems, called Computer Aided Systems, have become popular with high accuracy, consistent and efficient results. Computer

Aided Systems mainly uses machine learning, optimization, fuzzy logic methods in the numerical data stored. As with various biomedical applications, the diagnosis of Parkinson's disease is an important classification problem. The methods here may produce different results according to the

regulated data. In this sense, various machine learning algorithms should be tested in order to find a useful method for Parkinson speech data. Here, we have developed a model. In our model we are using a Data Set. This study proposes a decision system that contributes to the diagnosis of Parkinson's disease in real time. This study is organised as follows; Current machine learning studies on the diagnosis of Parkinson's disease are presented in this report. The methods and data set information used in the tests are explained briefly.

Experimental results and findings are given in the paper below .

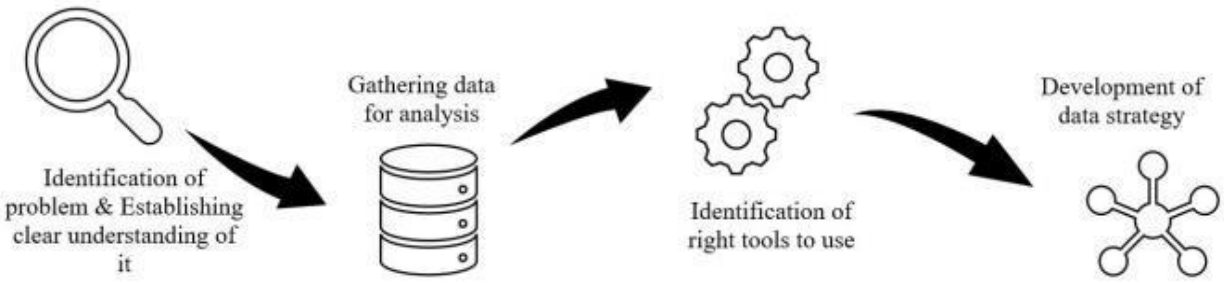


Figure 1. How Data Science works

LITERATURE REVIEW

In the base paper, classification methods such as Adaboost, Support Vector Machines, Naive Bayes, K-Nearest Neighbour, Artificial Neural Network and Community classifier are used. Also, the classification of sentences and continuous conversations is done by linear regression models on the acoustic feature from the middle of the vowels. The proposed framework demonstrated that 89% accuracy could be achieved in the detection of Parkinson's disease. In order to detect patients with Parkinson's disease, the proposed system for evaluating voice disturbances was compressed by calculating the mean squared values of sound to extract the audio from each sound sample. For

differentiation, different vector types are used together. Deficiencies in the production of emotional speech in Parkinson's patients were investigated. Biomedical sound measurements obtained from continuous funding samples were used. The minimum redundancy maximum selectivity feature selection algorithm has been applied to identify the features. The resulting features were classified by Artificial Neural Networks. In another paper, data obtained by a smart pen that can receive handwritten information from the Parkinson patient was used for automatic identification via the Convective Neural Network. Parkinson's disease-based classification algorithm that combines a community learning algorithm with the CART (Classification and Regression Tree) algorithm is proposed. First, the CART algorithm was repeatedly applied to the selection of speech samples so that samples with high discriminability were obtained. Second, a community learning algorithm combining random forest (RF) and Extreme Learning Machine (ELM) is trained based on optimised training examples. When the proposed algorithm is combined with CART and RF algorithms, the average classification accuracy has reached 86%. The paper has shown that in human beings with Parkinson's disease, dysarthria can be diagnosed using the classification of their properties. Multinomial and Bernoulli Naive Bayes were used to select the most suitable trait parameters for the diagnosis of Parkinson's disease. It has been found that the proposed algorithm's closest neighbouring algorithm to multiply the classification accuracy by up to 45%. It has been found that the proposed algorithm shows a higher success, particularly when combining the nearest neighbour and random forest methods to multiple editing. A study has proposed a deep multi-layer sensor (DMLP) classifier for behaviour analysis to predict the progress of Parkinson's disease using smartphones. It analyses the severity of Parkinson's patient’s actions by analysing speech and movement patterns measured by smartphone accelerometers at different times of the day. Linear regression is classified by random forests, the nearest neighbours, popular machine learning algorithms such as M5P and DMLP.A study used C4.5, C5.0, Random Forest and CART algorithms to produce decision trees on the R studio interface. Decision tree model is determined in the most accurate classification by using individual speech signals.

TECHNICAL DISCUSSION

1. Methodology

i. Dataset

The dataset is obtained from kaggle consisting of numerous patients with Parkinson's disease as well as healthy individuals. Multiple voice recording types were taken from all subjects. A group of 23 linear and non-linear properties were extracted from each sound sample. The dataset consists of the voice frequencies and its various components of people who have already suffered from Parkinson's disease and also contains numbers of fit and healthy people. **The dataset henceforth will be referred to as 'Parkinson's data'**.

ii. Support Vector Machine

Support Vector Machine (SVM) uses the idea of core based learning. It aims to separate high-dimensional attribute data with a kernel function. The Support Vector Machine classifies the different classes on the most appropriate hyper plane by creating a decision plane from the samples. In the case of the One-dimensional Support Vector Machine, the data is first moved to the attribute space using an appropriate kernel function, and then the two classes are separated by a hyperplane.

iii. Data Pre-processing

The data collected from the dataset undergoes data pre-processing i.e. preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating the machine learning model. 'Parkinson's data' generally contains noises, missing values, and is in an unusable format which cannot be directly used for machine learning models. Parkinson's data undergoes data preprocessing and cleaning, it is made suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

iv. Importing Libraries

One of the key and first things we have to do for data preprocessing is to import the required libraries. **Pandas** is one of the most famous Python libraries and used for importing and managing the datasets. It is an open-source data manipulation and analysis library. Numpy Python library is used for including any type of mathematical operation in the code. It is the fundamental package

for scientific calculation in Python. It also supports adding large, multidimensional arrays and matrices.

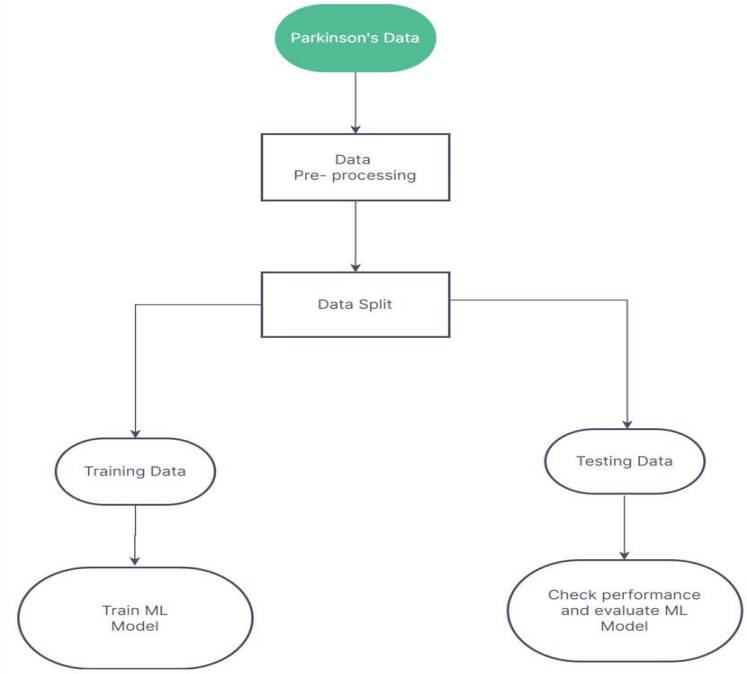
v. Finding missing data

The next step of data preprocessing is to handle missing data in the datasets. If our dataset contains some missing data, then it may create a huge problem for our machine learning model.

Hence it is necessary to handle missing values present in the dataset.

vi. Splitting the data

After the data is preprocessed, it is split into training data and test data in the last step of the data preprocessing. This is one of the crucial steps of data preprocessing as by doing this, we enhance the performance of our machine learning model. The data split is described in the figure below.

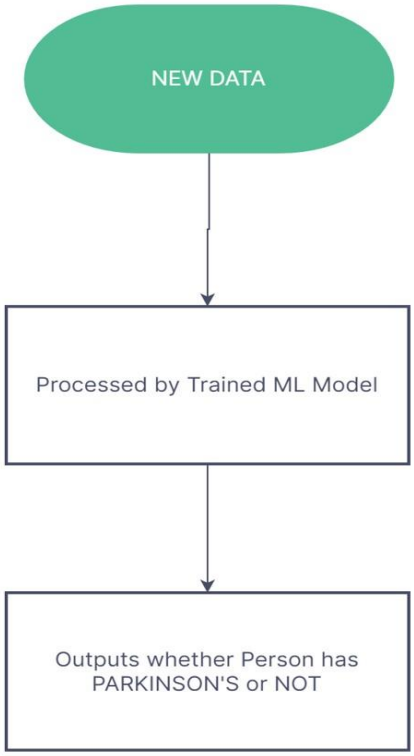


vii. Training the ML Model

After the data is split into two datasets, namely the training data and test data, the ML model undergoes training. The training data is used to train the Support Vector Machine (SVM) Model so that the model is ready for understanding the data of the patient. The test data is used to check the efficiency of the model and test its working. It checks the performance of the working of the model.

viii. Detection of Parkinson's Disease

The final step of the methodology describes the detection of Parkinson's disease. In the previous step, the Support Vector Machine Model is completely trained and is fully functional. This trained model is fed the data of the patients who want to know their status of Parkinson's data. The model then understands, examines the data and finally prints the output. The output tells us whether the person is suffering from Parkinson's data or not. The detection of the disease is described in the figure below.



2. Experimental Results

The most common machine learning methods are used to classify the speech data set of Parkinson's patients. The Support Vector Machine has been tested with different activation functions. Test Accuracy is considered as a benchmarking measure. The accuracy was recorded to estimate the effectiveness of machine learning methods in large data problems of many people with Parkinson's disease. The classification stage, in the training process 80% of datasets, in the testing process 20% of datasets are used. It is preferable to have a high accuracy rate to indicate the best method. It is seen that more accurate results are achieved by using expanded feature spaces. One of the most prominent core-based methods, the Support Vector Machine has been tested with different cores and parameters. The Support Vector Machine gave the best result

81.42%. With the help of Principal Component Analysis 23 features are reduced to 5 features. The expanded features with PCA provide more accurate classification results due to 5 PCA features.

CONCLUSIONS AND FUTURE TRENDS

1. Conclusion

Parkinson’s disorder is a disorder whose diagnosis is complex because of its symptoms similar to other disorders. Moreover, the lack of awareness increases the vulnerability of the patient’s health. This often leads to misdiagnosis of the disorder. The diagnosis of Parkinson’s Disease is not a straight-away-process which implies, a single test like ECG or blood test alone cannot determine PD in a person. Doctors need to study the patient’s medical history followed by some neurological tests . With the high rate of misdiagnosis of PD, due to indefinite tests, leads to a crisis. Technology such as Data Science and Machine Learning tend to utilise this crisis as an opportunity to make diagnosis and treatment of PD patients easy. Thus, medical history of people holding values of Central Nervous System related features can be used to predict Parkinson’s Disorder at early stages. Since PD has not been found with any cure till date, its early detection makes early diagnosis possible. In this study, Support Vector Machine in machine learning are tested for prediction of Parkinson's disease, which gave an astounding accuracy of 81.42%. For better classification accuracy, the different feature extension schemes are presented. The core features are composed as the features obtained by PCA and IG respectively and as the original features. It is concluded that, for the prediction of Parkinson’s disease, the extended feature spaces improved the classification accuracies for all classification algorithms.

2. Future Trends

In future studies, different feature selection or reduction methods can be examined to improve the classification accuracy. Since using various machine models result in different accuracies, using a better model can result in giving the best accuracy in the near future. Also, the data which is being tested by the model to detect the status of Parkinson’s disease int the person, can be appended to the Parkinson’s data which will enhance the chances of a better accuracy and could be helpful in proper early diagnosis of PD.

REFERENCES

1. Parkinson's Disease Detection by Using Machine Learning Algorithms and Hand Movement Signal From LeapMotion Sensor, Anastasia Moshkova, Andrey Samorodov, Natalia Voinova,AlexanderVolkov Bauman, Ekaterina Ivanova, Ekaterina Fedotova,April 2020

2. Li, Y., Yang, L., Wang, P., Zhang, C., Xiao, J., Zhang, Y., & Qiu, M., "Classification of Parkinson's Disease by Decision Tree Based Instance Selection and Ensemble Learning Algorithms", Journal of Medical Imaging and Health Informatics, 7(2), 444-452, 2017

3. Cantürk, İ., & Karabiber, F., "A Machine Learning System for the Diagnosis of Parkinson's Disease from Speech Signals and Its Application to Multiple Speech Signal Types", Arabian Journal for Science and Engineering, 41(12), 5049-5059, 2016

4. Sztahó, D., Kiss, G., & Vicsi, K., "Estimating the severity of parkinson's disease from speechusing linear regression and database partitioning", In Sixteenth Annual Conference of the
International Speech Communication Association, 2015

5. Alhussein, M., "Monitoring Parkinson's Disease in Smart Cities", IEEE Access, 5, 19835-19841, 2017

6. Benba, A., Jilbab, A., Hammouch, A., & Sandabad, S., "Voiceprints analysis using MFCCand SVM for detecting patients with Parkinson's disease", In Electrical and Information Technologies (ICEIT), pp. 300- 304, IEEE, 2015

7. Zhao, S., Rudzicz, F., Carvalho, L. G., Márquez-Chin, C., & Livingstone, S. R., "Automaticdetection of expressed emotion in Parkinson's Disease", In ICASSP, pp. 4813-4817, 2014

8. Pereira, C. R., Weber, S. A., Hook, C., Rosa, G. H., & Papa, J. P., "Deep learning-aided

Parkinson's disease diagnosis from handwritten dynamics", 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), (pp. 340-346). IEEE, 2016

9. Meghraoui, D., Boudraa, B., Merazi-Meksen, T., & Boudraa, M., "Parkinson's DiseaseRecognition by Speech Acoustic Parameters Classification", In Modelling and Implementation of Complex Systems, pp. 165-173, Springer, 2016

10. Zhang, H. H., Yang, L., Liu, Y., Wang, P., Yin, J., Li, Y., & Yan, F., "Classification ofParkinson's disease utilizing multi-edit nearest- neighbor and ensemble learning algorithms with speech samples", Biomedical engineering online, 15(1), 122, 2016

11. Wan, S., Liang, Y., Zhang, Y., & Guizani, M., "Deep multi-layer perceptron classifier forbehavior analysis to estimate parkinson's disease severity using smartphones", IEEE Access, 6, pp. 36825-36833, 2018

12. Vadovský, M., & Paralič, J., "Parkinson's disease patients classification based on the speechsignals", In Applied Machine Intelligence and Informatics (SAMI), 2017 IEEE 15th International

Symposium on pp. 321-326, IEEE, 2017

13. Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., ... & Kursun, O., "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings", IEEE Journal of Biomedical and Health Informatics, 17(4), 828-834, 2013

14. Ekinci, E. Omurca, S.I. and Acun, N., "A Comparative Study on Machine Learning Techniques using Titanic Dataset", 7th International Conference on Advanced Technologies, pp. 411-416, 2018

15. Su, M., & Chuang, K. S., "Dynamic feature selection for detecting Parkinson's diseasethrough voice signal", In RF and Wireless Technologies for Biomedical and Healthcare Applications

(IMWS-BIO), 2015 IEEE MTT-S 2015 International Microwave Workshop Series on pp. 148-149, IEEE, 2015