International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

IJIEMR Transactions, online available on 27th Mar 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 03

10.48047/IJIEMR/V12/ISSUE 03/39

Title **DETECTION OF PARKINSON'S DISEASE USING MACHINE LEARNING**

Volume 12, ISSUE 03, Pages: 275-281

Paper Authors
P. SarojaRani, G. Venkata Kartheek, K.V. Durga Prasad, K. Manoj Babu, D.Venkateswarlu

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Detection of Parkinson's Disease Using Machine Learning

P. SarojaRani, Assistant Professor, Department of Information Technology, Seshadri Rao Gudlavalleru Engineering College, Andhra Pradesh.
G. Venkata Kartheek, IV B. Tech Department of Information Technology, Sheshadri Rao Gudlavalleru Engineering College Andhra Pradesh.
K.V. Durga Prasad, IV B. Tech Department of Information Technology Sheshadri Rao Gudlavalleru Engineering College, Andhra Pradesh
K. Manoj Babu, IV B. Tech Department of Information Technology Sheshadri Rao Gudlavalleru Engineering College, Andhra Pradesh
D.Venkateswarlu, IV B. Tech Department of Information Technology Sheshadri Rao Gudlavalleru Engineering College Andhra Pradesh.

Abstract
Parkinson's disease (PD) sufferers are increasing in number alongside the ageing populace. Due to a lack of training and information, PD is often not diagnosed in a timely manner in developing nations. On top of that, not every person with PD experiences the same symptoms, nor do those symptoms always become more noticeable at the same point in time. Thus, the goal of this work is to utilise a cloud-based machine learning system for tele monitoring PD patients in developing countries, combining more than one symptom (rest tremor and voice degradation) to make a diagnosis. Data on rest tremor and vowel phonationare taken from smartphones equipped with accelerometers and fed into the proposed system. For the purposes of developing and refining more effective machine learning models, the data are mainly gathered from individuals who have been identified with PD and from the general population. The accuracy of the taught algorithms in detecting PD is then assessed by collecting data from freshly suspected PD patients. Patients diagnosed with PD are referred to a local physician for evaluation based on the results of these methods
Keywords: — Vocal Recording; Tremor Recording; Machine Learning; Python – keras, Tensor flow; Accuracy.

Introduction

Two to three percent of the over-65 community is affected by Parkinson's disease (PD), making it the second most prevalent neurodegenerative illness after Alzheimer's [1]. Dopaminergic cell death in the substantia nigra is a hallmark of this disorder [2]. Presence of movement signs is required for a determination of PD (bradykinesia, rigidity and tremor at rest [3]). Autopsy and MRI investigations, however, show that movement symptoms of PD only appear when dopaminergic neuron loss has progressed by 50-70%. Clinical evaluation in the area of neurology can be bolstered by a number of different methods. Tests based on images are frequently used, such as single photon emission computed tomography (SPECT) and M-iodobenzyl-guanidine (MIBG) heart scintiscan, but they can be pricey and inconvenient for some patients. Electrical activity of the brain's pyramidal neurons can be recorded non-invasively using electroencephalography (EEG), providing a window into their function that is both indirect and highly time-resolved.

Since it's a simple and inexpensive method, it's been put to good use in the research of epilepsy conditions. Although visual EEG analysis remains the gold standard in the clinical setting, advances in information processing have made it possible to derive additional information useful in describing brain disorders. Since EEG data has a high temporal precision, it can reveal changes in electrical brain activity and coupling over time. Therefore, EEG recordings have been made in a variety of states, including the "base waking condition," during slumber, while performing cognitive activities that require high levels of sensory processing, and at rest. Different neural circuits are anticipated to become active under each of these states, with resting-state coupling typically occurring spontaneously. EEG data are notoriously challenging to analyse due to two primary characteristics: the well-known inter- and intra-subject unpredictability, and the presence of artefacts. These include the randomness of the data and a relatively poor signal-to-noise ratio. Because of the high level of noise in EEG signals (as indicated by a low signal-to-noise ratio), it is necessary to perform pre-processing in order to analyse the signals and obtain results (after removing potential artefacts, like contamination from other biological and non-biological signal sources). The lack of a universally accepted "gold standard" for signal cleansing is the primary hurdle in the way of effective signal filtering. To add insult to injury, removing the signal noise can also mean removing important components in the EEG signals, which could contribute to an incorrect diagnosis.

As opposed to this, events and their states are considered random when they have no bearing on one another. Therefore, sophisticated non-linear approaches are required, along with more costly processing resources, to derive the important features of the data. Machine learning (ML) methods, which are quite powerful and enable non-linear studies as well as the handling of raw EEGs, can be used to evaluate the EEG patterns and thus surmount these obstacles. As per Arthur Samuel, "the field of study that enables PCs to learn without being expressly customized," AI is an area of man-made consciousness that spotlights on making calculations with the capacity to sum up ways of behaving and perceive stowed away examples in a lot of information. ML techniques can be separated into the emblematic assortment, which utilizes formal dialects, sensible cycles, and images, and the subsymbolic assortment, which is made to anticipate utilitarian associations between information.

In the realm of machine learning (ML), artificial neural networks (ANN) are networks whose structure is built on hierarchy models with numerous layers, allowing them to acquire depictions of data at varying degrees of complexity. However, a lot of data is needed for teaching purposes, and the procedure must be handled carefully. In recent years, all of these methods have found widespread use outside of the medical field, where they were originally developed for use in picture analysis. There are a plethora of articles that use ML techniques to study conditions other than epilepsy in which EEG analysis has been applied, including Alzheimer's, schizophrenia, and major depressive disorder. ML-based EEG processing has also been used for therapeutic purposes, such as in the rehabilitation of stroke patients. Even though ML methods in EEG analysis have only been around for the past few years, interest in using EEG to research Parkinson's disease has grown dramatically over the past five years, resulting to a flurry of new publications. The primary target of this study is to survey the present-day impact of ML strategies on EEG investigation of PD patients, and the second is to recognize the most famous techniques and assess those that have yielded the best outcomes. These objectives, which focus on the recognition and movement of PD, will act as a leaping off point for extra investigation into the distinguishing proof of an early, painless, and effectively accessible indicative marker that diminishes the time expected to analyze the illness. The creation of prophylactic treatments that reduce the rate of progression of motor and cognitive decline in PD is aided, hopefully, by the fact that novel diagnostic methods in PD may allow for the disease to be detected in its earlier phases, which is, in pre-motor stage.

2. LITERATURE REVIEW

Background: Motor complaints are the primary basis for a diagnosis of Parkinson's disease (PD), with costly and sometimes unavailable imaging methods like single photon emission computed tomography (SPECT) and M-iodobenzyl-guanidine heart scintiscan (MIBG) providing additional support. In this study, we looked at the research on electroencephalography (EEG) evaluations for quiescent state and muscle activation to identify PD with ML methods. Methods: The study was conducted according to the standards established by the Recommended Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Detailed notes were taken on the key features and outcomes of all releases up until May 2020. The findings included nine research. Seven participants used EEG during repose, and two participated in EEG during movement. The majority of research (83.3%) relied on subsymbolic models. Accuracy ranged from 0.62 to 99.62% when classifying PD. There was a significant deal of variation in the EEG cleansing procedure and the derived features. On the other hand, spectral features were the most prominent. Results showed that the model's success in forecasting the classification was highly dependent on both the characteristics incorporated into the model and the model's design. However, there was no correlation between research with regard to how thoroughly they cleaned their EEG data. Classification of neurological diseases using EEG and ML methods is a relatively new and rapidly developing area of study.

Various movement complaints are typically characterised by doctors in order to make a diagnosis of Parkinson's disease (PD). However, human bias is a risk with conventional diagnosis methods because they depend on evaluating motions that can be difficult to categorise due to their subtlety to the human eye. Meanwhile, early PD non-motor signs may be minor and can be triggered by a wide variety of conditions. As a result, early identification of PD is difficult because these signs are often disregarded. As a result of these challenges, machine learning techniques have been adopted for the classification of PD and healthy subjects or patients with comparable clinical symptoms (e.g., movement disorders or other Parkinsonian syndromes). In this analysis, we combed through studies published in PubMed and IEEE Xplore up until February 14, 2020 to give you a full picture of the data sources and machine learning techniques that have been used in the diagnostic and differential diagnosis of PD. For this study, we looked into the goals, data sources, data categories, machine learning techniques, and related results of 209 studies and pulled the pertinent material to share here. These findings highlight the promising potential for applying machine learning techniques and new indicators in clinical decision making, resulting in a more methodical and accurate identification of PD.

The spreadsheet containing the extracted data includes the following details: The following fields must be filled out: (1) goals, (2) diagnostic type (diagnosis, differential diagnosis, sub-typing), (3) data source, (4) data type, (5) number of subjects, (6) machine learning method(s), dividing strategy, and cross validation, (7) results, (8) year, and (9) reference.

The "year of publishing" was understood to be the year in which the research was first made available online and not the year in which it was stored. If no publishing date was given, the year the piece was published was used. Information was gleaned from research that presented new models and used old models only for comparison. Those with PD who had tests that did not show a dopaminergic deficiency (SWEDD) were considered a subtype (Erro et al., 2016).

We have further classified the studies based on the type of diagnostic and their overall purpose to better describe the varying aims and objectives of the included studies. From a diagnostic standpoint, these studies can be broken down into three distinct categories: (a) the diagnosis or detection of PD (which compares data collected from PD patients and healthy controls), (b) differential diagnosis (differentiating between patients with idiopathic PD and patients with atypical Parkinsonism), and (c) sub-typing (discrimination among sub-types of PD).

The overarching goal of the research included was also examined. We designated as (a) "methodology" studies those whose primary goal was the creation of novel technical approaches for use in the diagnosis of Parkinson's disease. These could incorporate, yet are not restricted to, novel AI and profound learning models and structures, information obtaining gadgets, and component extraction calculations. Clinical application studies were characterized as those that (a) approve and research the utilization of recently distributed and approved AI and profound learning models, or potentially (b) examine the possibility of presenting information modalities that are not normally utilized in the AI based determination of PD, like CSF information.

Here, we used precision as a metric to evaluate the effectiveness of various machine learning algorithms. We outlined the categories of machine learning models that achieved the best precision on a per-study basis for each category of data. On the other hand, some investigations have only tried out a single ML algorithm. As a result, we stipulate that a "model linked with the per-concentrate on most prominent exactness" is either (a) the main model applied and utilized in a review or (b) the model that accomplished the most elevated precision or that was underscored in examinations that utilized various models. The outcomes are introduced as a mean standard deviation (SD).

The precision on tests or validation was taken into account if the study reported both instruction and testing/validation results. Research reporting both confirmation and test precision was taken into account. Accuracy was calculated as an aggregate over multiple datasets or categorization issues in studies that involved more than one (such as HC vs. PD and HC vs. unexplained hyposmia vs. PD). Researchers combined the accuracy recorded for each subject group to get a total for trials that gave categorization accuracy for each group separately. The greatest reported accuracy was used, even when studies reported a range of accuracy or when various cross validation techniques or feature combos reported varying accuracy. Diagnosis of illnesses besides PD or Parkinsonism (like amyotrophic lateral sclerosis) was not taken into account in research comparing HC with other diseases or PD with other diseases. Assessment accuracy was not taken into account.

In the field of neurology, Parkinson's disease is classified. The torso and the limbs begin to shake, and the body becomes rigid. At this late point, there is currently no effective remedy or therapy. Only if treatment is started soon after the illness first appears will it be effective. These may not only help with the financial burden of the illness, but may also save a life. Detection of Parkinson's disease with current techniques is limited to its later stages, when approximately 60% of the dopamine in the basal ganglia has been lost. This region is responsible for coordinating the body's movements with a relatively modest supply of dopamine. It has already expanded rapidly around the globe, affecting more than 145,000 people in the United Kingdom alone. In India, nearly one million people are afflicted by this illness.

Although it typically appears in those over the age of 65, about 15% of cases are discovered in those under the age of 50. We'll use XGBoost, KNN, SVMs, and the Random Forest Algorithm to see which one is most effective at identifying the first signs of illness.

Abstract The neurological illness known as Parkinson's disease worsens over time and is persistent. Damage or death of the brain regions responsible for producing dopamine causes a decline in a person's ability to perform routine activities like speaking, writing, walking, and so on. Over time, patients experience a worsening of these symptoms, which increases the seriousness of the condition. In this article, we suggest a technique for predicting the seriousness of Parkinson's disease using deep neural networks on the UCI Parkinson's Telemonitoring Speech Data Set. We have built our neural network for severity prediction using the python 'TensorFlow' deep learning framework. Our approach yields higher precision values than that of prior studies.

As shown in Fig. 1, the suggested technique for using deep learning to forecast the seriousness of Parkinson's disease is described. As a first move, we compile statistics on PD patients' speech patterns by recording their conversations. Following collection, min-max

-

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

standardisation is applied to the data. The next stage is to create a deep neural network with an input layer, concealed levels, and an output layer. As many neurons are used in the input layer regardless of how many characteristics there are. Two neurons representing the two groups, "severe" and "non-severe," are located in the output layer. The built deep neural network receives the standardised data for training and assessment.

3. PROBLEM STATEMENT

Although there is currently no treatment for Parkinson's disease, medicines can significantly improve symptom management. Treatment and diagnosis must begin as soon as possible. Efforts to consolidate and combine phenomenological studies of the experience and impression of life with Parkinson's disease have been largely unsuccessful. Currently, Parkinson's disease diagnostic methods leave much to be desired. Patients with PD may experience an increase in their health-related quality of life if they receive treatment in a comprehensive outpatient PD paradigm. Using guided machine learning algorithms, determine if an individual has Parkinson's disease based on their speech.

4.PROPOSED SYSTEM

Using supervised machine learning categorization methods, the system suggests a way for identifying Parkinson's disease using speech characteristics - K-Nearest Neighbor, Support Vector Machines, Random Forests, and Decision Trees are some of the most popular supervised learning techniques.

5. IMPLEMENTATION

5.1 Dataset collection:

Information for this unit comes from the Kaggle dataset files. Data on vocal characteristics and the occurrence of Parkinson's disease are included in this collection of records.

5.2 Data Cleaning:

Data that may be inaccurate, partial, copied, or structured incorrectly are removed or modified as part of this module's data cleansing process to make the data ready for analysis.

5.3 Feature Extraction:

This is done to decrease the number of attributes in the dataset, which has the benefit of accelerating training and improving accuracy.

5.4 Model training:

In this module we use supervised classification algorithms like xgboost to train the model on the cleaned dataset after dimensionality reduction. Testing the trained model: In this section, we put the machine-learning model through its paces on an actual testing sample.

5.5 Performance Evaluation:

In this module, we evaluate the performance of trained machine learning model using performance evaluation criteria such as F1 score, accuracy and classification error. In case the model performs poorly, we optimize the machine learning algorithms to improve the performance.

5.6 Prediction of Parkinson's disease:

In this module we use trained and optimized machine learning model to predict whether using the vocal features will reveal whether the patient has Parkinson's disease or not.

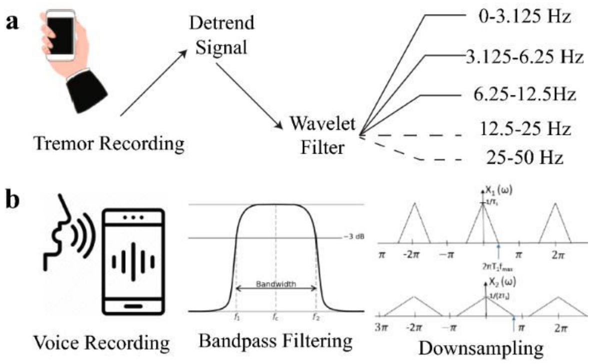

Figure1.Data Collection

Loading model to compare the results

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

6. RESULTS

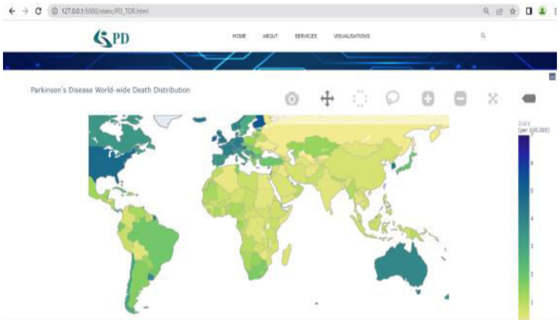

Figure 2. Worldwide Death Rate

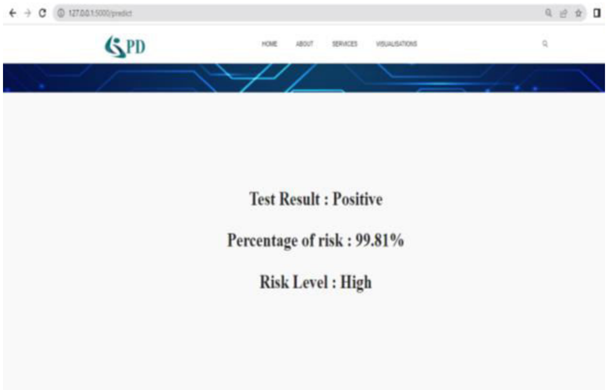

Figure 3. The result for the data input

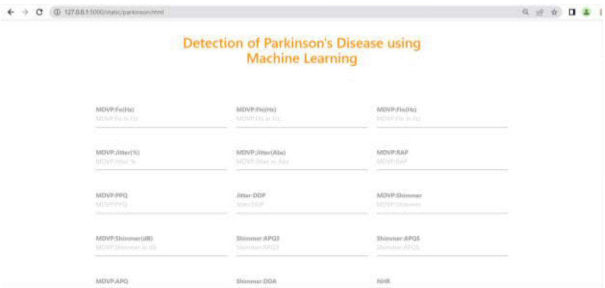

Figure 4. The input parameters

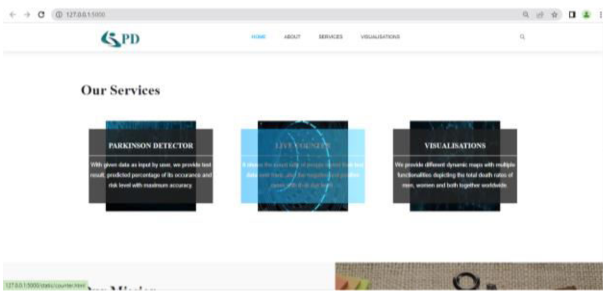

Figure5.The services

Figure6. The Home Page

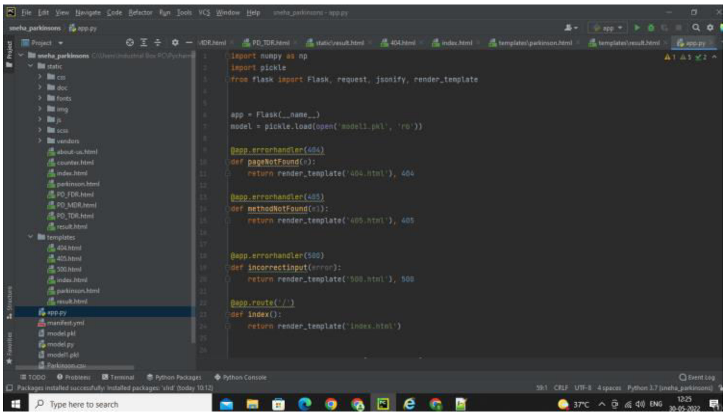

Figure7. The code

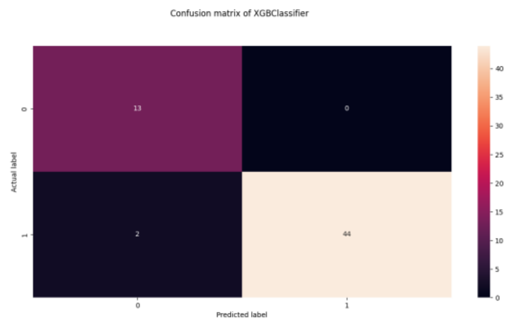

Figure 8. The confusion matrix

7. Conclusion

Parkinson’s disease totally death rate worldwide PD male and female death rate is calculated. The suggested system uses smartphone-captured samples of speech and movement to identify PD, and it does so utilising a cloud-based system for processing, storing, and continuously tracking these samples. When in-person surveillance is scarce, as it often is during an epidemic such as COVID-19, this method can be a useful tool for healthcare officials in poor nations to raise the level of medical care that the senior people can access.

-

8. References

1. B. E. Sakar et al., "Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings," in IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 4, pp. 828-834, July 2013, Doi: 10.1109/JBHI.2013.2245674.
2. Benba, A., Jilbab, A. & Hammouch, A. Analysis of multiple types of voice recordings in cepstral domain
3. using MFCC for discriminating between patients with Parkinson's disease and healthy people. Int J Speech Technol 19, 449–456 (2016).
4. Ö. Eskıdere, A. Karatutlu and C. Ünal, "Detection of Parkinson's disease from vocal features using random subspace classifier ensemble," 2015 Twelve International Conference on Electronics Computer and Computation (ICECCO), Almaty, 2015, pp. 1-4, doi: 10.1109/ICECCO.2015.7416886.
5. Nissar, Iqra & Rizvi, Danish & Masood, Sarfaraz & Mir, Aqib. (2018). Voice-Based Detection of Parkinson's Disease through Ensemble Machine Learning Approach: A Performance Study. EAI Endorsed Transactions on Pervasive Health and Technology. 5. 162806. 10.4108/eai.13-7-2018.162806.
6. Y. Li, C. Zhang, Y. Jia, P. Wang, X. Zhang and T. Xie, "Simultaneous learning of speech feature and segment for classification of Parkinson disease," 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), Dalian, 2017, pp. 1-6, doi: 10.1109/HealthCom.2017.8210820.
7. Caliskan, Abdullah & Badem, Hasan & Basturk, Alper & Yüksel, Meltem. (2017). Diagnosis of the Parkinson disease by using deep neural network classifier. Istanbul University - Journal of Electrical and Electronics Engineering. 17. 3311-3318.
8. Das R. (2010) "A comparison of multiple classification methods for diagnosis of Parkinson disease". Expert Systems With Applications"; 37:1568-1572.
9. Genain N, Huberth M, Vidyashankar R. (2014) "Predicting Parkinson's Disease Severity from Patient Voice Features."
10. Benmalek E, Elmhamdi J, Jilbab A. (2015) "UPDRS tracking using linear regression and neural network for Parkinson's disease prediction." International Journal Of Emerging Trends & Technology In Computer Science (IJETTCS); 4:189-193.

-