Paper 5: Predicting Parkinson’s Disease Progression: Evaluation of Ensemble Methods in Machine Learning

Publication year: 3 February 2022

Author name: Mehrbakhsh Nilashi , Rabab Ali Abumalloh , Behrouz Minaei-Bidgoli , Sarminah Samad , Muhammed Yousoof Ismail , Ashwaq Alhargan , and Waleed Abdu Zogaan

Summary:

Parkinson’s disease (PD) is the second most common and complex neurodegenerative disorder worldwide. Both polygenic and environmental factors can cause PD. It is found that, in about 1%–2% of the PD cases, the disease development occurs through a single gene main symptoms of PD are bradykinesia (motor features), muscle stiffness, and tremor, along with other symptoms such as sleep disorders (nonmotor features), cardiac arrhythmia, and constipation. Alteration of voice and speech is one of the features of PD. Unified Parkinson’s Disease Rating Scale or UPDRS, which shows symptoms’ presence and severity, is mainly used in tracking PD symptom progression. UPDRS is considered as the well-validated test and the most widely used clinical rating scale for patients with PD . UPDRS includes 4 sections, in which UPDRS I, UPDRS II, UPDRS III, and UPDRS IV are used to evaluate psychiatric symptoms in PD, activities of daily living, reliable motor symptoms measured in PD recognized by physical exam, and complications of treatment . In many studies, this scale is considered

based on Total-UPDRS with the range of 0–176 (176 total disability and 0 representing healthy) and Motor-UPDRS which indicates the UPDRS’ motor section with the range of 0–108 (108 indicating severe motor impairment and 0 indicating healthy state) . Machine learning (ML) approaches have demonstrated the capability of handling large volumes of medical datasets and presented perceptive directions .use of MLbased tools could enhance the safety of individuals , enhance the quality of medical care [16–18], minimize the costs of medical care , and support physicians’ efforts by manipulating big data of patients’ records. ML approaches have been broadly utilized for disorders’ classification and prediction [22–30]. Gadekallu et al. investigated the use of machine learning techniques for the prediction of diabetic retinopathy.authors used the PCA-based Deep Neural Network (DNN) model using the Grey Wolf Optimization (GWO) algorithm for of the extracted features of the diabetic retinopathy dataset. )e method was evaluated through the accuracy, recall, sensitivity, and specificity evaluation metrics and compared with the support vector machine (SVM), na¨ıve Bayes classifier, decision tree (DT), and XGBoost. Overall, their method achieved higher accuracy compared with the SVM, DT, and XGBoost techniques. Bhattacharya et al. developed a method for the classification of imbalanced multimodal stroke dataset. )e authors implemented the Antlion optimization algorithm on the DNN model to select optimal hyperparameters in minimal time consumption. A positive aspect of their method was that it consumed only 38.13% of the training time on the stroke dataset. An artificial neural network is among the most significant approaches for disease classification and prediction . Referring to Berner [39], clinical decision support systems (CDSSs) are special tools that are developed to aid medical specialists in their decision-making, considering particular disorders or diseases. ML approaches can be utilized for designing effective CDSS to aid medical specialists in reaching accurate and timely predictions. CDSSs designed using machine learning approaches have played a significant part in evaluating the existence or the severity of the disease. In machine learning methods, unsupervised approaches are used to lower the dimensionality of data, which allows the detection of the disease. Besides, these approaches allow manipulating the data, removing the noise from data, calculating the similarity, and segmenting the data [40]. On the other hand, supervised learning approaches are used to enable the final classification, prediction, and diagnosis of the diseas. While ML has proven its benefits, the effective deployment of ML needs a great effort from human specialists, considering that no particular approach can present acceptable results in all possible scenarios. Although clinical data are available to researchers to explore, the lack of experience to handle big sources of data might restrict the optimum utilization of these sources. Besides, even though several approaches have been used in disease prediction using various real-world medical datasets, the choice of the deployed approach should consider enhancing the accuracy of the prediction and minimizing the time of computation goal of this paper is to present a comparison of machine learning approaches for remote tracking of Parkinson’s disease progression. )e comparative study is based on clustering and prediction learning approaches. To further improve the accuracy of UPDRS prediction, this study uses ensemble learning in the final stage of the proposed method. Ensemble learning approaches have proven to be effective in prediction tasks . Few studies have incorporated ensemble learning approaches for the development of the diseases diagnosis systems. Further investigations are needed for the effectiveness of these approaches in UPDRS prediction. Accordingly, we use ensembles of support vector regression and different clustering techniques for PD data clustering. )e results are then compared with other prediction learning approaches, deep belief network (DBN), support vector regression, multiple linear regression, and neurofuzzy techniques.

Conclusion

Most of the presented methods for PD prediction depend strongly on human proficiency . )e benefits of deploying the ML in the medical sector are that they provide objective, context-independent, and data-driven analysis . ML approaches have been utilized effectively in disease diagnosis and severity prediction. Particularly, ML has also been utilized in analyzing the data collected from wearable IMU sensors for automated evaluation of motor disorders like PD. Hence, the practical aim of this study entails providing supplementary, quick, and accurate methods that can aid experts in reaching more objective medical decisions considering the PD diagnosis. By deploying these methods in the appropriate systems, several gains can be acquired that entail reducing the expenses of manual diagnosis and minimizing diagnosis time. Continuing this line of research and supporting previous literature, this study uses both unsupervised and supervised learning techniques to diagnose PD through UPDRS prediction. Besides, clustering, dimensionality reduction, and prediction learning techniques are used to create the PD diagnosis method. )e basic aim of this paper is to conduct comparative research of the ML approaches for PD diagnosis. We concentrated on clustering and prediction learning methods to conduct the comparative study. Particularly, several clustering approaches for PD data segmentation and SVR ensembles to predict Motor-UPDRS and Total-UPDRS were used. )e findings are then evaluated based on other prediction learning methods, MLR, neurofuzzy, and SVR techniques based on a real-world PD dataset. )e finding of the study indicated the superiority of deploying EM with SVR ensembles in relation to decision trees, neurofuzzy and SVR combined with other clustering approaches in the prediction of Motor-UPDRS and TotalUPDRS. Many previous works have been conducted focusing on patients’ classifications, severity prediction, and remote monitoring. Still, there are future routes in each field to be investigated. Besides, several sensors such as magnetometer, accelerometer, and gyroscope have been utilized and assessed. Additionally, MRI, EEG signals, f-MRI, and DATSCAN images were utilized to present accurate predictions of the disease. Other research directions can be followed by utilizing other brain signal images such as ECG, EMG, and PCG. Other sensing modalities can be explored and combined to present a more accurate classification of the disease. Even though ML methods in previous literature have presented high classification accuracy for PD detection, still, there are some obstacles related to feature extraction and selection which need to be addressed [104]. )e utilization of several features can increase the computation time [105,106]. On the other hand, if fewer features were utilized, this will increase the complexity of extracting the features, which will accordingly impact the computation time. )is paper has some shortcomings which should be considered in future research. )e study is based on a real-world dataset to assess the proposed approaches, which has one limitation considering the number of features used in the prediction process. Other PD datasets with a larger number of features can be utilized in the evaluation of the deployed approaches. Large datasets can present more generalized outcomes. Emerging technologies can be used to collect data from patients using particular applications, as suggested by Bot et , in which the authors developed an application to collect the data from PD patients using their iPhones. )is approach can ease the data collection from the public because of the availability of smartphones and help to present more generalizable outcomes. Furthermore, this study can be extended by incremental machine learning approaches to improve the computation time of previous PD diagnosis methods in processing large dataset.