Decision making in health care diagnosis: Evidence from Parkinson's disease via Hybrid Machine Learning

D Jinil Persis*
Operations and Supply Chain Management area,
National Institute of Industrial Engineering (NITIE), Mumbai, India
Email Id: jinilpersis@gmail.com

V Raja Sreedharan
BEAR Lab, Rabat Business School, International University of Rabat, Rabat, Morocco
Email Id: raja.sreedharan@uir.ac.ma

Gopalakrishnan Narayanamurthy,
University of Liverpool Management School, Chatham Street, Liverpool, L69 7ZH, UK
Email Id: G.Narayanamurthy@liverpool.ac.uk

Accepted for Publication in
IEEE Transactions on Engineering Management

Corresponding author

Decision making in health care diagnosis: Evidence from Parkinson's disease via Hybrid Machine Learning

Abstract

Health care is a complex system that demands critical decision-making especially in diagnosis of various conditions in patients. To minimize possible errors in diagnosis, an emerging technology, machine learning (ML) is being effectively used. ML classifiers can be used to proactively diagnose the medical conditions, which are identified based on the presence or absence of specific characteristics of the diseases. Therefore, present study demonstrate how ML can be used to determine Parkinson's disease (PD) and thereby provide early diagnosis using non-clinical data of the patients. Novel ensembles are developed in this study to improve the diagnostic capability and the experimental results show that improved versions of artificial neural network (ANN) could yield 13.4% more accurate results compared to the traditional ANN classifier. PD is considered a challenging medical condition owing to its global relevance and complexity in diagnosis. Moreover, early detection of PD is instrumental for patient recovery, and any lapses in diagnosis can lead to an immeasurable loss to patients. Also, study has developed an effective diagnostic tool for PD and detects the disease at an early stage using voice data of individuals, and this would aid making better clinical decisions related to PD, thus rendering better health services.

Keywords: *Machine Learning (ML); Parkinson's Disease (PD); Random Forest; Artificial Neural Network; Support Vector Machine; Hybrid Classifier*

1. Introduction

In health care operations, clinical decision-making aids in the diagnosis and treatment of a disease based on a well-developed knowledge base. Further, decision support systems have also been developed in recent years with a focus on minimising or eliminating human errors in such diagnoses or treatments [1]–[3]. Identification of the presence of a disease and the level of its progression in the human body is a complex process governed by numerous factors, which make the whole procedure manually cumbersome. However, machine learning (ML) techniques are applied in health care to aid the medical practitioners in predicting, identifying and classifying the diseases in their primitive development [4], [5]. The World Health Organisation has defined three types of diagnostic errors, namely, (i) missed, (ii) delayed, and (iii) wrong diagnoses, rendering the health care system highly unreliable and unsafe [6]. Diagnostic errors pose a huge risk to patient health and safety. Healthcare system is affected by influx of new diseases making it difficult to provide diagnosis and makes the decision-making process more ambiguous.Also, different issues such as glitches in data-capturing systems, lack of continuous monitoring systems, inadequate medical equipments, disproportionate doctor–patient ratios, and non-adherence in patients can lead to a poor diagonsis [7], [8]. On the otherhand, a huge volume of clinical data is collected in the health care information system which are hardly used to generate any meaningful insights. This is because data are manually collected, siloed, or at times even made inaccessible to the practitioners appropriately in a suitable form [9]. However, the effective use of emerging technologies such as artificial intelligence and machine learning (ML) could make efficient use of these data to derive constructive insights and knowledge and provide valuable healthcare services. Further, with the increasing population growth day by day, making accurate and precise diagnosis and providing better patient care has become a daunting task. The Institute of Medicine (IOM) released a report on *"To Err Is Human: Building a Safer Health System"* exposing the issues related to patient safety and how diagnostic error affects the patients care [10].

To provide better healthcare services, identifying the diagnostic error is critical for improving the quality of care. However, diagnosis error is underappreciated in the healthcare domain and mostly never surface and further gets concealed within rapid treatments, therapeutic interventions, trauma care and emergency care often given to the patients in priority. Achieving perfect concordance between physician's diagnoses and patient's actual state enables better patient care, as proper diagnosis paves way to patient recovery and wellbeing. Furthermore, the medical professionals are overburdened with different diseases and diagnostic processes with associations, redundancies and commonalities possible, they often inadvertently get into a bias in their decision making known as the cognitive bias [11]. Such bias in the decision-making leads to diagnostic error. Furthermore, data on diagnostic errors is sparsely available making it complex to identify and minimise in the healthcare system. From the patient's perspective, diagnostic error is the failure to identify the patient's condition without any delay and inform them accordingly for early intervention [12], [13]. Moreover, getting the right diagnosis for the patients depends on the involvement of a medical professional in the diagnosis process itself. Further, the accuracy of the diagnosis is influenced by the clinical reasoning, where physician takes dual approach - analytical and non-analytical decision making for the diagnosis. However, poor knowledge about the disease, failing to read the pattern of the diagnosis and overconfidence leads to a cognitive bias in choosing the non-analytical approach for complex diagnosis leading to diagnosis and poor patient care.

To overcome, these issues and provide better care to the patients, the authors have proposed machine learning based medical diagnosis which is demonstrated for Parkinson disease. Here non-clinical data is used for diagnosis which can further be validated with regular clinical diagnosis thereby, minimising diagnostic error by providing enhanced decision-making capability to the physicians. Also, using ML based prediction for different diseases, physicians can observe specific patterns and hence be empowered with better information for effective decision making and provide better patient care thereby, reducing the cognitive bias. Although, identifying degenerative disorders like Parkinson's disease (PD) using MLis being explored in the literature, however, understanding the nature of PD is still challenging due to the different rating scales and complex diagnosis procedures adopted [14]. Therefore, the paper focuses on the issues related to diagnosis of PD and challenges involved in the process. This study is guided by following research question:

RQ: What are the challenges in diagnosis of PD and how proactive measures can be developed for better patient care?

Taking this situation as a motivation for this study, the study focused on developing a diagnostic tool for PD using efficient ML classifiers. The classification model is developed iteratively using training dataset and the ML algorithms attempt to recursively modify the classification rule to minimise the classification error, which is the deviation between predicted class of the data point and the actual class. Firstly, choice of classifiers suitable for this study are made through experimental study [14] and based on the results, artificial neural network

(ANN), support vector machine (SVM), decision tree (DT) and random forest (RF) classifiers were found to be preferable with less detection error (both Type I and Type II). However, robustness of the model is necessary for an accurate clinical diagnosis. Therefore, secondly, hybrid ensemble was developed in this study to increase the performance of PD diagnosis, making it more resilient. Also, the hybrid classifiers can overcome the drawbacks of the traditional classifiers with better performance for PD diagnosis especially when feature intensive non-clinical and unstructured data is used in diagnosis [15].

The study has unique contributions such as (1) Based on results from the traditional classifiers, the classifier with less classification error was chosen and the authors developed hybrid ensembles namely, RF–ANN, DT–ANN and SVM–RF classifiers for the PD diagnosis tool and found that RF–ANN classification model is best suitable for developing the PD diagnostic tool. (2) To enhance the diagnostic practices adopted for PD by medical professionals, the study has proposed a deployment framework for the PD diagnostics. (3) The findings from the study proposed a decision-making model for the identification of PD through voice data. This helps in the early diagnosis of this chronic disease, thereby helping the patient to seek various treatment options. ML based non-clinical diagnosis becomes a precursor for further clinical investigations and timely treatment process. (4) Finally, the PD diagnostic tool can be emphasised as a data-driven assistant for the medical experts not only for early diagnosis but also for efficient diagnosis with less diagnostic error.

The remainder of the paper is structured as follows. The theoretical background of the study is presented in Section 2. Traditional ML classifiers were used to analyse the patient's voice for diagnosis in Section 3. Section 4 highlights the development of the proposed hybrid classification model. Section 5 summarises the study with a deployment framework for the diagnostic tool and concludes with future directions for further research.

2. Theoretical Background

2.1 Clinical Diagnosis: Criticism and Drawbacks

Diagnosis is a process of collecting the patient's medical history and identifying a condition, injury, or disease by assessing the symptom exhibited by the patient. Based on the nature of the symptom, the diagnosis process consists of physical examination followed by laboratory examination. The patient's condition is assessed, the results of the examination are compared with the standard value and the findings from the diagnosis aid the medical professional to draw a conclusion for further medical interventions. However, correctness in diagnosis is easily affected by different errors which is increasing day by day. Oyebode (2013) states that *"Clinical error as the failure of a planned action to be completed as intended or use of a wrong plan to achieve an aim".* Moreover, the healthcare system suffers as the diagnosis error increases the patient's burden and cost of care. According to the Institute of Medicine (IOM), every year clinical error and poor diagnosis contribute to 17-29 billion USD and further, lead to 7000 deaths and is potential of increasing duration of hospitalization for 4-5 days [11], [16]. Therefore, minimising the diagnosis error is instrumental for better diagnosis and patient care.

Diagnosis error has become more difficult to capture due to the shortage of availability of reliable data [17]. Moreover, studies reported different definitions for the diagnostic errors making it more challenging for assessment based on the available information. Also, diagnostic error can often be blurred and not expressed clearly among the other possible errors in medication, therapy, surgery and other medical procedures involved in the entire patient lifecycle [18]. Also, the chance of such diagnostic error reaching patient/relative attention is very less and gets often lost in the process. Furthermore, the complexity of the diagnosis process along with the uncertainty makes the measurement activity a more complex one. Apart from this, it is noteworthy that the diagnosis mostly involves both subjective and objective patient health information examined by the physician with complex cognitive skills. Hence, diagnosis process is a multi-facet, time-dependent and team centric that can chip in to the diagnostic error easily [19]. The chronological component of the diagnosis process can muddle with the measurement because, over the time, signs and symptoms for a condition may vary creating confusion in the diagnosis timeframe and hence patients' actual condition and physician's perceived condition must be synchronous for effective medical treatment. Clinical reasoning contributes to the diagnostic error, which is difficult to assess as it occurs in the physician mind and can't be captured for documentation. Furthermore, there exist a non-linearity in the diagnosis error such that people recover from health conditions irrespective to their treatment or diagnosis, so that the diagnosis error can't be properly identified [20]. Clinical reasoning along with the diagnostics, is underrated in the healthcare training and education [21], [22]. This lack of reasoning and poor focus is a cause for the diagnostic error [23]. Further, the clinical reasoning for the diagnosis is based on the dual procedure, which is known approach for decision making in the healthcare consisting of non-analytical and analytical approaches for decision making. Non analytical models (fast system 1) involve unconscious, intuitive, and automatic pattern recognition [24]. Analytical models (slow system

2) involve a conscious, deliberate process guided by critical thinking [24]. These systems lead to uncertainty in decision making and cognitive bias among the physicians. Moreover, most of these cognitive biases occur due the over-dependency on the system 1 or system 1 overrides the system 2 leading to diagnostic errors. Therefore, poor understanding due to System 1 leads to clinical overconfidence and diagnostic error [25]. Further, the diagnostic reasoning depends on one's understanding and in-depth knowledge of the disease and its attributes. To develop such in-depth knowledge, one has more clinical experience with the patients and reviewing different diseases and patient's cases to develop a knowledge base.

Promoting approaches that creates the possibility to avert, detect and correct the errors in the diagnosis for different diseases is in demand but such approaches need measurement tool which can ingest both the qualitative and quantitative data leading to effective assessment of the patient conditions and avoid diagnostic error through better interventions. Given, the importance of the healthcare operations, its useful to have measurement tools that can capture the diagnostic data and provide actionable insights to the medical professional to avoid diagnostic errors and increase their competency in the clinical diagnosis. For this purpose, the tool should be able to provide feedback and decision support for the medical professionals based on the diagnosis data. Such tools should be able to assess whether a risk can occur or not along with the feedback for training and learning. Therefore, to overcome this issue and empower the medical professional, the researchers use machine learning approach to capture clinical and non-clinical data addressing interdependencies and non-linearity for effective decision making.

2.2. Challenges in PD diagnosis

Parkinson's disease is defined as "movement disorder or progressive neurodegenerative disease" [26]–[28]. This disease affects individuals as early as 50 years normally, but there are some younger victims also. Around 10 million people in the world (0.3%) are affected by this disease and the current prevalence of PD in developing countries is around 300–400 out of 100,000, which is expected to more than double by 2030. Also, it is the second most neurodegenerative disease after Alzheimer's. Those who are affected by this disease have difficulty in coordination and balance of the body posture. Motor organs gradually reduce in functionality and their movement is affected due to stiffness. This disease is caused by a loss of dopaminergic neurons, leading to abnormalities in neurotransmitters due to which there is a loss in control of the motor organs.

Most studies of PD show that the disease is diagnosed in a majority of the cases through clinical examination. Previously, Levy *et al*. (2002) conducted neurological and neuropsychological evaluations such as verbal and non-verbal memory, visuospatial ability, and abstract reasoning to evaluate the impact of PD on patients [29]. Consequently, Wood *et al*. (2002) investigated 109 idiopathic patients with PD [30]. The patients were evaluated for a year and follow-ups are made through telephonic conservations. However, the study focused on whether the patient will be fallers and non-fallers with PD, which made the analysis subjective without considering all aspects of health of the patients in the study. Also, the logistics regression model used in this study had biased confidence interval due to a smaller sample size. Subsequently, Postuma *et al*. (2010) highlighted that identification of the early stages of PD is critical for neuroprotective therapy and the patient's well-being [31]. The authors conducted a clinical study for the prediction of Parkinson's disease using the sleep disorder and obtained a prediction accuracy of 50%. However, it requires complex inputs such as autonomic testing, cardiac MIBG scintigraphy and transcranial ultrasound for primary screening of patients, which in turn requires both time and resources. Similarly, Stern and Siderowf (2010) conducted a clinical study to predict PD through symptoms, such as abnormalities in olfaction, gastrointestinal function, cardiac imaging, vision, behaviour, and cognition [32]. However, the authors found that the chances of false positives exceed the actual identification of the disease. Also, an incorrectly diagnosed individual may be subjected to unnecessary tests and treatment, which places an undue burden on the patients and affects their well-being as well as entails unnecessary cost.

On the other hand, Kerr *et al*. (2010) used a Unified Parkinson's Disease Rating Scale to predict the patient's chance of falling ill due to PD [33]. Around 101 patients underwent the test and 48% were found to fall within the PD range. This method had a sensitivity of 78% and specificity of 84% for predicting the fall rate. Later, Schrag *et al*. (2017) conducted a study on 309 patients with PD to understand the chance of being cognitively impaired. They found that patients who are 50 years and older have a 5% chance and those 70 years and older have a 34% chance of cognitive impairment by employing logistic regression [34].

According to Oguh and Videnovic (2012), the National Parkinson Foundation (NPF) in the United States collected data from 43 Centers of Excellence and 8 Care Consortium Centers, which collectively served 50,000 patients affected by PD [35]. The study recognised a gap that the centers were not confident about the quality of PD-specific inpatient care. Furthermore, the need to create a plan to provide more access to outpatient care to prevent unnecessary hospitalisations was realised. Also, the availiablity of qualified medical consultants was a challenge for the foundation. Then, Krzysztoń *et al*. (2018) conducted a literature review for diagnostic tools for PD and reported their reliability and sensitivity [36]. The study compared different tools, such as Activities-specific Balance Confidence Scale, Balance Evaluation Systems Test (BESTest), Berg Balance Scale (BBS), Fullerton Advanced Balance Scale (FAB), Functional Reach Test (FRT), Mini-BES Test, Timed Up, and Go (TUG) and Tinetti Balance Scale. All these tools are used to measure the level of PD rather than predicting it in earlier stage. Moreover, the tools focused on evaluating the static and dynamic posture control and sensory orientation of the patient to foresee the resultant problems and causes and the processes were found to be reactive in nature.

Through early diagonsis, PD is treatable with minimum impact. However, the diagnosis of PD is challenging as it has a large number of motor and non-motor symptoms. In addition, the brain function loses symmetry, leading to severe non-motor problems and these symptoms include facial expression, existence of tremors, walking pattern, voice, and stiffness. Non-motor symptoms include sleep disorders, smelling difficulties, constipation, postural hypotension, cognitive impairment, urogenital disorders, and mood disorders [29], [37]. Also, the symptoms take 12–14 years to surface, making it difficult to predict the disease at the earlier stages. However, an early diagonsis can ensure better treatment for the patient [38]. There are two types in PD: tremor dominant and non-tremor dominant. Brain imaging methods such as magnetic resonance imaging (MRI), functional MRI, and positron emission tomography (PET) are found to be reactive measures for the identification of PD. Also, DaTscan, an imaging test that measures dopamine function in the brain, is a complex clincial procedure for the identification of PD. Hence, there is an urge to develop a proactive method to achieve early diagnosis for PD. Further, the progression of the disease is slow in the early stages and the cost of treatment is very high after much of the neurons are lost. After a considerable loss of neurons of up to 60%, the disease starts to manifest its motor symptoms. Hence it is highly challenging to identify the disease in early stages and minimize the damaging effect on the patients and help speedy recovery.

3. Methods

Most of the patients exhibit vocal problems in the early stages. Dysphonia is a type of phonation disorder in Parkinson's patients. This is caused due to the disconnection in laryngeal nerve connections. This leads to disturbed articulation and fluency in speech including vocal tremors, roughness and weak voice [39], [40]. If an individual is diagnosed with dysphonia, the patients have to undergo voice tests, such as running speech test, phonation test, and sustained phonation test. The data from these tests have to be collected and then processed using various signal processing techniques, and useful features have to be thereafter extracted from the results. All these tasks constitute a time-consuming and tedious process. The need for deploying a robust early diagnostic model is vital for medical practitioners and therefore the authors are further motivated developing a deployment framework for the PD diagnosis using non-clinical data. Non-clinical data such as voice [38] and handwriting [41] are used to predict or detect PD in an individual. In [38], voice signals of population consisting of affected and not affected individuals are processed and well-known machine learning classifiers are employed to detect PD however, with a maximum accuracy of 83%. In order to improve the reliability of machine learning classifiers in detecting PD, it is vital to construct ensembles to achieve high accuracy, precision, specificity and sensitivity.

3.1.Dataset

The dataset [1] used in this study, contains voice data from 188 patients with PD (107 men and 81 women) in the age range of 33–87 years and 64 healthy individuals (23 men and 41 women) aged between 41 and 82 years. The data were collected from the Department of Neurology, Faculty of Medicine, Istanbul University. The patients were asked to voice the vowels three times and their speech was recorded through the microphone. Further, the patient's approval and signed consent were obtained for data collection. The voice data was obtained from each individual thrice and a total of 756 samples were collected. Using the voice data, their features were extracted by applying voice signal processing technique, namely, the tunable Q-factor wavelet transform (TQWT) method [38]. Through the TQWT method, the following characteristics were extracted: jitter, shimmer, amplitude, signal-to-noise ratio, phonation frequency, intensity, bandwidth parameters, wavelet features and vocal fold features for both the healthy indivduals and patients with PD [26]. The final dataset with the voice data and its features is available in the UCI Machine Learning repository [38].

3.2.PD detection models

Using the final dataset, the authors focused on developing a PD diagnosis. Primary task is to identify ML classifiers suitable for voice data and to accurately classify the individuals as

[1] https://archive.ics.uci.edu/ml/datasets/Parkinson%27s+Disease+Classification

abnormal (with PD) and normal (without PD). Although most of the classification problems can be well solved by ML algorithms [42], choosing the right algorithm is a daunting task [43]. Data-driven solutions are provided to solve variety of problems in industries including condition monitoring and diagnosis in manufacturing systems [44]–[46]. However it is important to make use of proper learning ensemble appropriate to the problem. Therefore, the authors used ML algorithms such as naïve Bayes, artificial neural network, support vector machine, decision tree, and random forest classifiers that are widely used in the literature and further developed hybrid ensembles [56]. These ML classifiers were used in this study for the diagnosis of PD using the voice dataset and learning capabilities of these classifiers were analysed. Further, the ML algorithm gets trained with past experience and takes decisions based on the learning obtianed from previous instances and continuously improves its learning with new instances for maximising its performance [19]. Any new unclassified data point can be classified further with the help of the learned system.

The right choice of appropriate ML classifier for this specific problem and right hyperparamter settings pave the way for the development of a better classification model, yielding good classification results[50-51]. From the given dataset, the preclassified training data points that are characterised by its features and the corresponding class are given as input to the ML algorithm. The classifying rule is iteratively developed with the goal to minimise the error between the predicted and the actual class[52-54]. After developing the classification rule and the model getting well trained, the fitted model can be used to predict the class of unclassified data points characterised by its features alone. This ML approach has been implemented in this study to detect PD using voice data collected from different individuals. The methodology adopted in this phase is depicted in Fig. 1.

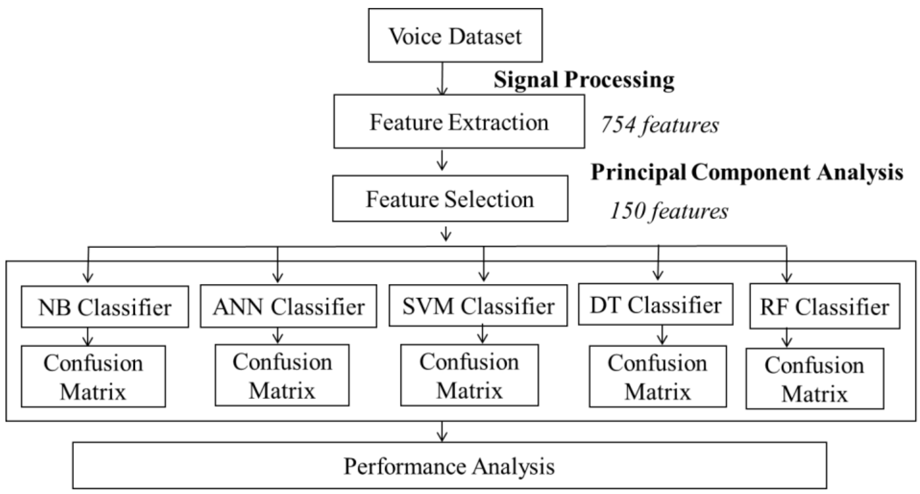

Fig. 1 Workflow of preliminary phase.

In this study, each voice data point is characterised by 754 features that are extracted from the TQWT processing technique [38]. It is known from literature that use of appropriate feature reduction techniques improve the convergence of ML algorithms and fit of the model [47]. Hence principal component analysis was employed to reduce the features [48]. Then the dataset containing 756 data points were divided into a training set and test set where the training set was used to constructively develop the fitted model and test set was used to evaluate the performance of the fitted model from the obtained confusion matrix. From the confusion matrix, the performance of the classification models can be studied and various measures can be evaluated.

3.2.1. Feature Reduction

There were 754 features for each data point and principal component analysis was applied to reduce the features and remove redundant and non-contributory features. The factor loadings for each principal component were obtained and the variance explained by each principal component is shown in Fig. 2a. The cumulative proportion of variance for each principal component is also shown in Fig. 2b.

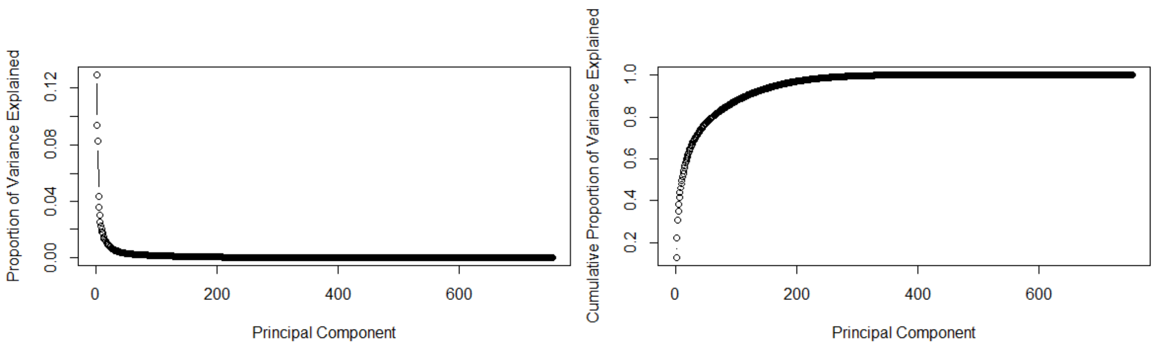

Fig. 2. Variation explained by the features: (a) proportion of variance for each principal component; (b) cumulative proportion of variance for each principal component.

It can be observed that maximum proportion of variation is due to the initial major components. So, in this study, the number of significant principal components taken for further study is fixed at 150.

3.2.2. Experimental results

The traditional classifiers were used to develop diagnostic models for PD and the performance of these classifiers were analysed to investigate the learning capabilities of these classifiers for the diagnosis of PD. Five widely used machine learning classifiers such as NB, SVM, ANN, DT, and RF classifiers are found in the literature. Voice data of the tested individuals was used to detect the presence of PD. In this study, experiments are conducted in R Studio (using R Programming Version 3.6.2) using appropriate R packages.

The classification models developed in this study attempt to classify every individual into either abnormal (with PD) or normal individual. Each model is trained with 567 pre-classifieds (75% as training dataset), each characterised by 150 significant features extracted from the voice data after applying principal component analysis-based feature reduction and the binary class to which the data point belongs to and tested with 189 unclassified data (25% as test dataset). The classification model is developed iteratively using training dataset and the ML algorithms attempt to recursively modify the classification rule so as to minimise the classification error, which is the deviation between predicted class of the data point and the actual class.

After training, the fitted model is tested with the training dataset where each data point is characterised by only the features excluding the class to which it belongs. The predicted classes (positive/negative) of the test dataset obtained from the fitted model were then compared with the actual classes (true/false) and the confusion matrix for binary classification problems comprising four indices – true positive (TP), true negative (TN), false-positive (FP) and false-negative (FN) – were obtained. Several performance measures can be derived from the confusion matrix that show the classification performance of the underlying ML algorithm used. Also, the hyperparameters of the ML classifiers play an important role in determining the performance of the classification models [49]. It is known that the performance of the SVM classifier depends on the regularisation parameter and kernel coefficient that have been fine-tuned in this study through multiple simulation runs. Also, a suitable kernel function has to be selected for better results.

After tuning the SVM model, regularisation parameter is considered as 0.1, kernel coefficient is taken as 3 and polynomial kernel function of degree 3 is used. The test dataset (189 data points) is fed to the fitted SVM model. The predicted class and actual class were compared and the correctly and incorrectly classified data points are shown in Fig. 3a. Out of 189 test points fed to the network, it can produce 144 correct (lower side, at deviation=0) and 45 wrong (upper side, at deviation=1) classifications. The actual and predicted classes were compared for the test dataset and presented in Fig. 3a and various statistics can be drawn from the results.

Similarly, DT-based classification model was developed after pruning and tested with the test dataset, and the classification results are presented in Fig. 3b. In the ANN classifier, the number of hidden neurons, learning rate, activation function and epochs are the hyperparameters to be set. In this study, the smallest learning rate (SLR) neural network algorithm was used and the hyperparameters were set by repeated simulated runs and the classification results are presented in Fig. 3c. Sets of 500 decision trees were developed by the RF classifier and the decisions of all trees were compiled to take a unified decision at the end. The classification error with the increasing number of trees was studied. The deviations are plotted in Fig. 3d. Also, the NB classifier was used to develop a classification model for PD and the results are presented in Fig. 3e.

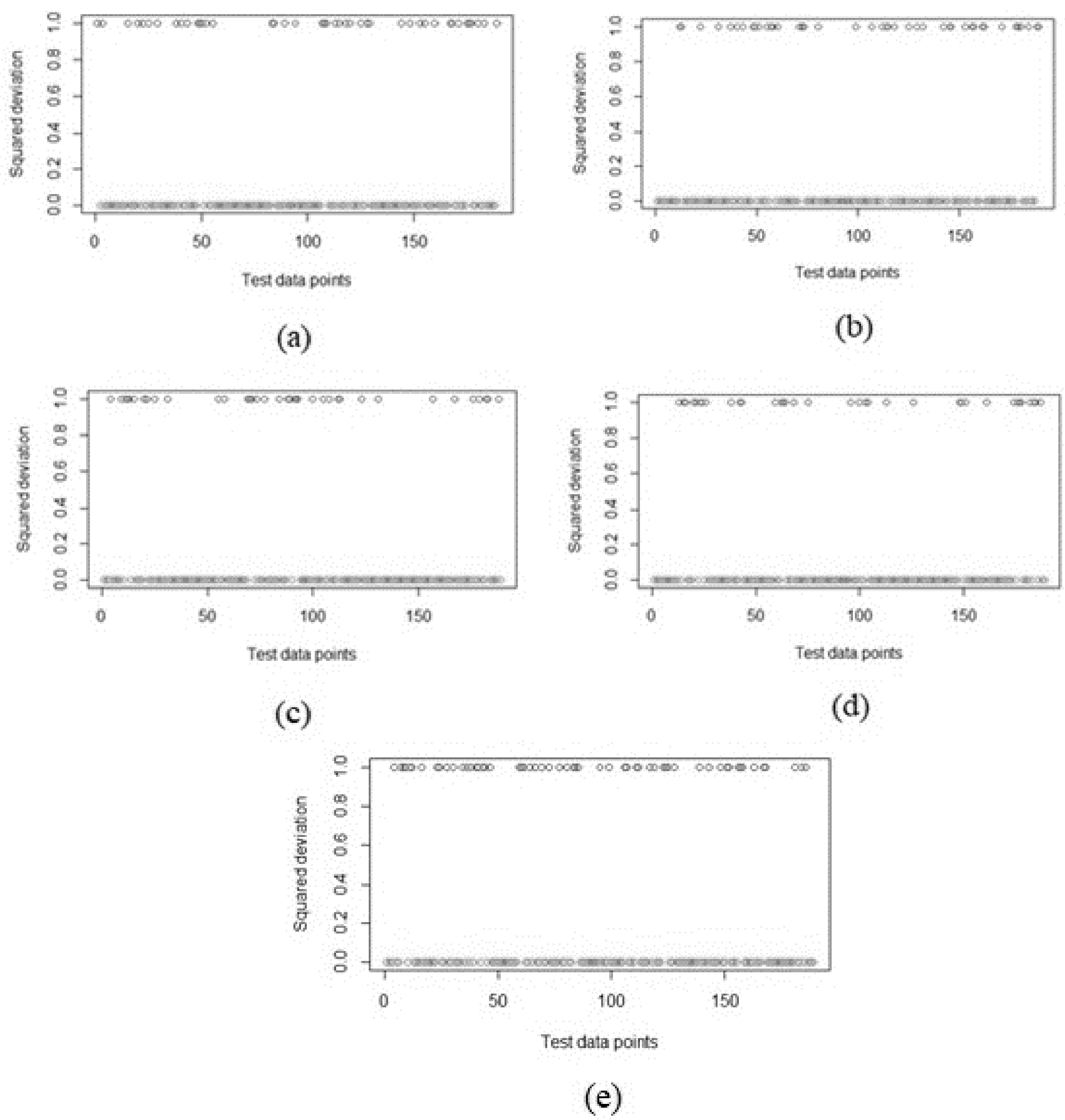

Fig. 3. Deviation between actual and predicted values for the classifier: (a) SVM; (b) DT; (c) ANN; (d) RF; (e) NB

These predicted results are compared with the actual classes viz, PD and not PD to obtain the confusion matrices. The confusion matrices show the number classifications under true positive (TP), true negative (TN), false-positive (FP) and false-negative (FN) categories out of the total test samples. Performance measures such as recognition rate or recognition accuracy, error rate, precision, recall or sensitivity, receiver operating characteristic (ROC), mean average error (MAE), root mean square error (RMSE), and normalised root mean square error (NRMSE) were used to evaluate the performance of the classifiers [50].

Recognition rate or recognition accuracy, which depicts the overall classification performance (in %), is calculated as

$$\textbf{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{Total}} * 100 \tag{a}$$

Precision (in %) is calculated as

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}*100. \tag{b}$$

Recall or sensitivity, which exhibits the classification performance in terms of positive samples (in %), is calculated as

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} * 100 \tag{c}$$

Specificity, which shows the classification performance in terms of negative samples (in %), is calculated as

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} * 100. \tag{d}$$

Root mean square error (RMSE) is calculated as

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{Actual}_i - \mathrm{Predicted}_i)^2}, \tag{e}$$

where *n* is the number of test data points.

The obtained confusion matrices for the traditional classifiers such as NB, SVM, DT, ANN, and RF classifiers are shown in Table 3.

Table 3. Confusion Matrix for the ML Classifiers

ML Classifiers	Confusion Matrix	Predicted Values	
	Actual values	*Positive values*	*Negative values*
NB	True values	118	14
	False values	31	26
SVM	True values	106	40
	False values	5	38
DT	True values	125	22
	False values	27	15
ANN	True values	122	30
	False values	18	19
RF	True values	137	17
	False values	33	2

The confusion matrices (see Table 3) obtained during the testing phase from the adequately trained and fitted classification models were developed using NB, SVM, DT, ANN and RF classifiers. For example, NB-based classification model achieved a true positive (TP) identification of 118 patients with PD, 14 patients having PD were misclassified as normal (TN), 31 normal individuals were misclassified as patients with PD (FP) and 26 normal individuals were correctly classified (FN). TP+FN are correct classifications and the remaining (TN+FP) are misclassified by the fitted NB-based classification model. Similarly other classifiers can also be examined for their correctness in classification. Using equations (a) – (e), the performance for each ML classifier was calculated and the values are tabulated in Table 4. Based on the number of correctly classified data points, various performance measures are calculated and presented in Table 4.

Table 4. Performance Comparison of the Traditional Classifiers for the Diagnosis of PD

Algorithm	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**	**RMSE**
Naïve Bayes	70%	79%	82%	31%	0.5491
Support vector machine	77%	95%	74%	89%	0.477
Decision tree	78%	82%	89%	45%	0.4714
Artificial neural network	80%	87%	86%	63%	0.4281
Random forest	81%	81%	99%	34%	0.4303

Therefore, ML algorithms were used to classify patients with PD using voice data. The observations made from Table 4 are as follows: The accuracy and sensitivity of the RF classifier are higher than all other traditional classifiers; however, in terms of precision and specificity, the SVM classifier excels among the traditional classifiers. Further, the RF classifier is highly accurate, followed by ANN, DT, SVM and NB classifiers. Furthermore, the SVM classifier, though highly precise and specific in identifying PD, is less sensitive and accurate in identifying patients with the disease. Also, the ANN classifier is found to have minimum RMSE followed by RF, DT, SVM and NB classifiers and yields accurate results however at the expense of precision and specificity. Finally, the DT classifier is found to have less RMSE than SVM but is more accurate and sensitive when compared to the SVM classifier. Based on the results, ANN, SVM, DT and RF classifiers were selected to further enhance their performance by developing combined ensembles.

3.3. Proposing a hybrid ensemble for PD detection

Based on performance of the different ML classifiers for the diagnosis of PD (see Table 3), the RF classifier is best suitable for the diagnosis of PD under the assumed test conditions with an accuracy of 81%. Also, ANN exhibited an accuracy of around 80%. However, robustness of the model is necessary for an accurate clinical diagnosis. Therefore, ML classifiers were combined, and hybrid classifiers were developed to increase the performance of PD diagnosis, making it more robust. Also, the hybrid classifiers can overcome the drawbacks of traditional classifiers with better performance for PD diagnosis. It can be seen in Table 4 that SVM is highly precise in identifying the disease and highly specific to normal conditions and DT classifier is accurate and sensitive to the features of the disease. In order to improve the specificity of DT classifier, the feature set (754 features) was reduced based on the relative importance of the features deduced by SVM. SVM based feature reduction has identified 280 important features which are further used by DT, ANN, and RF classifiers to develop a classification rule. DT based feature reduction could identify 85 significant features which are further given as input to ANN and SVM to develop classification rule. Similarly, RF based feature reduction could identify 68 prime features which could be further used by SVM and ANN to develop classification rule. Since RF, DT, ANN and SVM yielded high performance, attempt has been made to further improve different aspects of performance by developing combined classifiers. ANN based models get trained and set weights to features iteratively with experience in which initial weights are set by SVM, DT or RF so as to improve the convergence speed of the algorithm. The results of the hybrid ML classifiers are presented in Table 5.

Table 5. Performance Comparison of Hybrid ML for the Diagnosis of PD

Classifiers	Accuracy	Precision	Sensitivity	Specificity	RMSE
SVM–DT	85.71%	89.61%	92.62%	60%	0.3800
SVM–ANN	85.53%	93.48%	84.31%	88%	0.3469
SVM–RF	90.79%	87.93%	100%	72%	0.3035
DT–ANN	88.15%	92.59%	90.9%	80.95%	0.3022
DT–SVM	80.26%	93.48%	78.18%	85.71%	0.4442

RF–SVM	88.16%	97.87%	85.19%	95.45%	0.3441
RF–ANN	93.42%	92.98%	98.15%	81.81%	0.2197

Table 5 shows that the classification performance of promising traditional classifiers for PD has substantially increased. The comparison of traditional and hybrid ML classifiers in terms of accuracy, precision, sensitivity, and specificity is presented in Fig. 4.

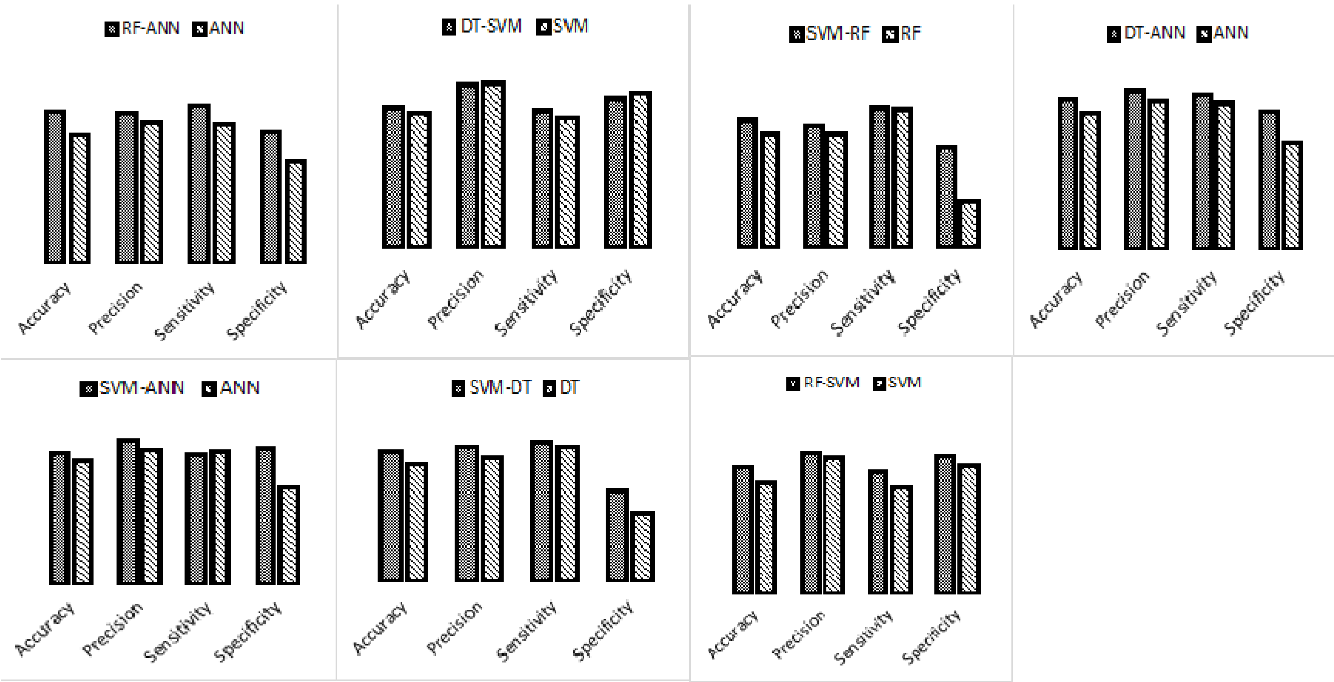

Fig. 4 Comparison of traditional and hybrid ML classifiers for the diagnosis of PD.

From Fig. 4, it can be observed that the accuracy, precision, sensitivity, and specificity of the DT classifier is enhanced by 10%, 9%, 4% and 33%, respectively, by combining it with the SVM classifier. The ANN classifier has yielded promising results in our study (see Table 4). When combined with SVM, the ANN classifier yielded improved results with increased accuracy, precision, and specificity of 7%, 7.4% and 40%, respectively, at the expense of sensitivity. However, when combined with the DT classifier, the accuracy, precision, sensitivity, and specificity of the ANN classifier are enhanced by 10%, 6.4%, 5.7% and 28.5%, respectively, and when combined with the RF classifier, the ANN classifier is enhanced by 16.7%, 6.9%, 14% and 30%, respectively. Hence RF–ANN outperforms DT–ANN and SVM–ANN classifiers in all aspects except sensitivity. The RF classifier, which is less sensitive to the features of the diseased individuals (see Table 4), becomes more sensitive when combined with the SVM classifier, which is highly sensitive (see Table 4) with an increase in value from 34% to 72% in addition to being 100% specific towards normal features as well. Hence the hybrid SVM–RF is also a promising classifier for the diagnosis of PD. When the SVM classifier is combined with DT, the precision and specificity values dropped and hence this combination not suitable for the diagnosis of PD; however, when SVM is combined with RF, the accuracy, precision, sensitivity, and specificity of the SVM classifier was enhanced by 14%, 3%, 15% and 7%, respectively. Therefore, RF–ANN, DT–ANN and SVM–RF classifiers are suitable for developing a PD diagnosis tool; however, RF–ANN and SVM–RF take more computational effort compared to DT–ANN. However, based on the experimental results, it can be concluded that the proposed RF–ANN classification model is best suitable for developing the PD diagnostic tool. ML provides significant health care solutions and mimics the role of a consultant for better health care service. Furthermore, literature clearly shows application of ML for disease detection.

4. Discussion of findings

Majority of the studies on PD show that the disease is predominantly diagnosed with clinical examinations including laboratory and imaging studies. In these studies, it was found that the chances of false positives (Type I) exceed the actual identification of the PD affected patients [51]. This leads to unnecessary human effort and undesirable damages to an healthy patient. It is clear that PD has been well reported in terms of classifying the patients through clinical study. Apart from this, studies requires invasive procedures which will harm the healthy indivduals in the clinical exminations. Many studies have also evaluated the impact of PD on a patient's cognitive movement and sensory orientation through different rating scales [52]. However, the accuracy of many studies range from 65% to 70%, which is not suitable for effective decision-making in health care. Very few studies focused on developing proactive measures for the detection of PD. Further, all the studies are based on the data collected from the patient through clinical examination and both the healthy and PD affected patients has to go through the clinical examination process which can be avoided with the present-day advancements in technology. Further, very few studies used the unstructured patient data such as speech recording and involuntary movements of the body to predict the chance of fallers for PD in their earlier stage of diagnosis.

Therefore, the study used the traditional ML classifiers such as Naïve Bayes, support vector machine, artificial neural network, and random forest which were widely used in the health care for PD diagnosis by analysing the voice data. Here, the authors used the voice data of the patients to develop a model that can classify an individual into affected or unaffected individual. Further, the voice data was processed by the TQWT approach (signal processing) and the findings of the voice data were reported in the open source data repository making it reliable and useable for further analysis. Further the prediction made based on five widely used classifiers such as Naïve Bayes; SVM; ANN; DT and RF for PD diagnosis, analysed the voice data to classify the fallers and non fallers of the PD. In order to make the diagnosis more robust, the authors developed hybrid ensembles. For the hybrid models, high performing classifiers were chosen and Naïve Bayes classifier was dropped due to lower accuracy. The performance of the Hybrid ML classifiers for the PD Diagnosis was estimated (*refer table 5*) and we found that, RF–ANN, DT–ANN and SVM–RF classifiers are suitable for developing a PD diagnosis tool; however, RF–ANN and SVM–RF take more computational effort compared to DT–ANN. However, based on the experimental results, it can be concluded that the proposed RF–ANN classification model is best suitable for developing the PD diagnostic tool. Further the learning capability and the number of instances given for training and performances, for RF–ANN is better than other hybrid models (*refer table 5*). Furthermore to use the diagnostic tool in a health care setting, a deployment framework is introduced to enhance the diagnosis experience of medical professionals. Further, the deployment framework can be used as a primary screening process for the PD diagonsis. Based, on the classifications, the PD fallers can be subjected to detail clinical examinations. Meanwhile non fallers can avoid the clinical examination making the dignostic processes more patient friendly and more proactive. Further, the need for the subject expert is not needed frequently for the deployment framework. As, the deployment framework can be used by the medical and nursing staff with little training in the smart devices in the healthcare facilities.

Further, to validate the generalization capability of the RF-ANN classifier, random sampling of dataset to form train and test datasets is made [44]. This is procedure is repeated (k-folds cross validation) and study has obtained RMSE of the classification error which is plotted in Fig. 6.

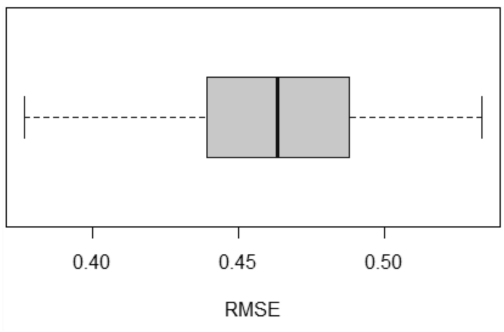

Fig. 6. Cross validation error of RF-ANN model

It could be observed that the mean value of RMSE is found to be 0.47 and the minimum RMSE value achieved for every partition is presented in Table 6.

Table 6. k-Cross validation results

Split	50:50	60:40	70:30	80:20	90:10
K	k=6	k=10	k=15	k=10	k=14
RMSE	0.4290	0.4291	0.4221	0.4176	0.3765

It could be observed from Table 6 that generalization capability of RF-ANN ensemble is found to be best at 90:10 split (90% training and 10% testing) of the dataset in which an RMSE of 0.38 is achieved. Hence, in the diagnostic tool generalized version of RF-ANN model is incorporated for better detection performance.

5. Deployment Framework: PD Diagnostic Tool

Deployment roadmap is very essential to visualize and implement strategies developed especially in a complex and highly dynamic healthcare system [53]–[55]. Employing the traditional classifiers and the author found them to be inadequate for developing a diagnostic model for PD. So, hybrid classifiers, combining two traditional classification systems, were used to yield better results. Further, study tested different hybrid classification models and found that the RF–ANN classification model yields better results with higher accuracy and less error. Therefore, RF–ANN is considered as an appropriate diagnostic model for identifying PD. Further, to use the diagnostic tool in a health care setting, a deployment framework is introduced (see Fig. 5).

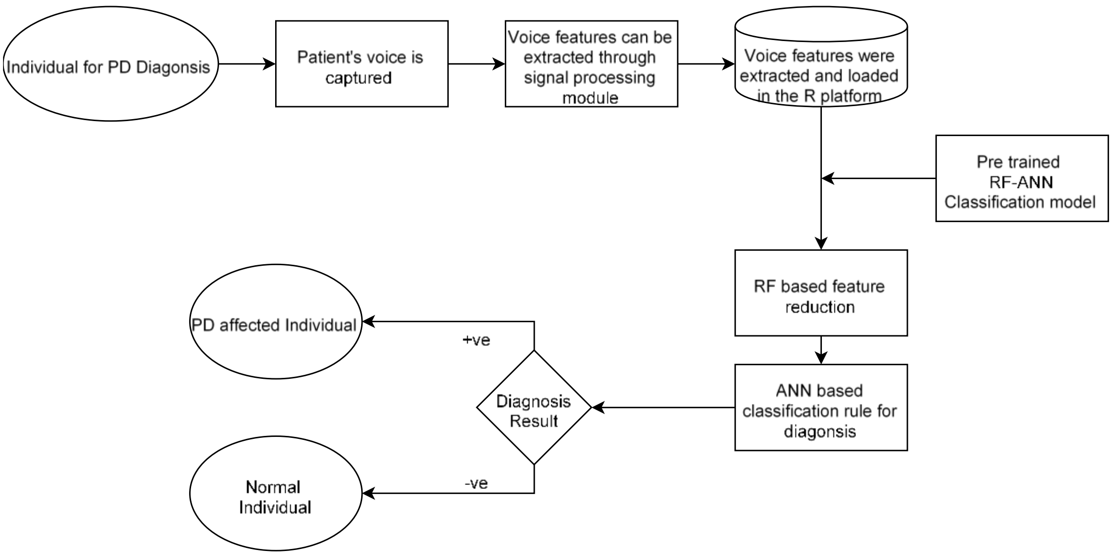

Fig. 5. Deployment framework

For the diagnosis of PD, the process starts with an individual visiting the hospital for evaluation. The individual's voice is captured through a microphone and the voice is recorded. Then, the features from the voice are extracted through the signal processing module. The voice data is analysed and voice-related features such as shimmer, jitter, fluctuation and pitch period entropy, frequency, density, and harmonic parameter are captured. Then, the voice-related features are loaded into the R platform, the pre-trained RF–ANN hybrid classification model is used to analyse the features of the voice. Also, RF focuses on feature reduction and sets initial weight vector for ANN to further classify the individual with high convergence. Then, the diagnostic output is generated through the RF–ANN hybrid classification model and the diagnostic results are categorised as positive (Parkinson-affected individual) and negative (normal individual). The model predicts the chance of a patient being affected by PD. Such a diagnostic tool will prove very useful for taking better clinical decisions as an early detector of PD. Moreover, various classifiers were used in developing the diagnostic tool in order to ensure better transparency and credibility in patient care to all the stakeholders. Also, the diagnostic results ensure that the health care professional can prescribe the necessary treatment for the Parkinson's-affected individual and create a proactive diagnostic approach for the detection of PD. Apart from this, the deployment framework can act as a primary screening process in the diagnosis and prevent the healthy individual from undergoing the clinical examination. Therefore, the deployment framework can serve as a proactive measure for the detection of PD and can be used for better diagnosis with minimum resources and experts input making it more appropriate for the healthcare facilities with limited medical professional for the PD diagnosis.

6. Conclusion

This study focuses on addressing the issues related to health care, particularly for disease diagnosis and patient care. The authors propose an approach to diagnose PD using the ML algorithms. Parkinson's disease is chosen in this study owing to its global relevance and complexity in diagnosis, especially in the early stages. The initial review found that most of the cases of PD diagnosis are reactive and static in nature. The diagnosis not only requires several clinical examinations but also is found to be both time-consuming and expensive. Therefore, the authors focused on developing a proactive and less time-consuming diagnostic method using the ML approach. Traditional ML algorithms shows less accuracy of only around 65–80% and moreover, for the diagnosis of PD the benchmark set in the literature was maximum 83% in terms of accuracy. The collection of non-clinical data of PD patients is also a challenging task. Hence, the authors obtained voice data and its features from the open-source repository. This study utilised the voice dataset given in the UCL Machine Learning repository, which consists of 756 voice data characterised by 754 features. The reduced dataset after applying principal component analysis to remove non-contributing features is fed to the traditional ML classifiers.

The classification accuracy was found to be limited to a maximum of 81% while using the traditional classifiers. Further to this, the authors developed hybrid ensembles based on the findings from the traditional classifiers. The hybrid classification model RF–ANN yields better results for the given voice dataset. Moreover, the accuracy of the diagnosis is found to be 93% sensitive up to 98% with lesser error (RMSE=0.22). Based on the findings, the trained RF–ANN can be used as a diagnostic tool for the detection of PD. So, the authors proposed a deployment framework for the diagnostic tool in the health care environment. Furthermore, the model serves as a diagnostic tool to support medical practitioners to enhance their service level in terms of consultation and diagnosis of PD in the early stages. The tool further incorporates a well-trained model using the RF–ANN hybrid classifier proposed by the authors and can detect the presence or absence of PD in any individual in a negligible amount of time and a non-invasive procedure.

6.1. Theoretical contributions

The study has unique contribution in the following ways. (1) Study contributes to healthcare literatures in term of a robust PD diagnsotic approach. (2) In Machince Learning literatures the present work extension the application of classifers for feature selection in PD. (3) In this study, the performance investigations of state-of-the-art ML classifiers such as artificial neural network, support vector machine, decision tree, and random forest algorithms in predicting PD using voice data, have established benchmark solutions in terms of accuracy, precision, errors, specificity, and sensitivity for future studies. Further, hybrid ML models are developed to further improve the accuracy and efficiency of PD diagnosis. The hybrid model RF–ANN outperforms other models with accuracy, precision, sensitivity, and specificity of 93.42%, 92.98%, 98.15%, 81.81%, and 0.2197 respectively. (4) Finally, study has contributed to the clinical diagnosis literatures for reducing the error in clinical diagnosis using ML.

6.2. Practical implications

The decision-making model for the identification of PD through voice test is developed in this study. This helps in the early diagnosis of this chronic disease, thereby helping the patient to seek various treatment options. Further, the diagnosis of PD is a complex process involving different inferences and knowledge on the disease. However, the PD diagnostic tool can be emphasised as a data-driven assistant for the medical experts to understand the diagnosis efficiently with less diagnostic error. Further, the study proposed a deployment framework which can act as a primary screening process for classifying the healthy individuals from the PD affected patients for further clinical examination. Further, deployment framework can be easily used by the medical and nursing staff, which can minimise the utilisation of medical professional and subject experts as their availability is a constraint in developing economies.

6.3 Limitations and future research

The main limitation of the study is data collection. The availability of the useable clinical data is challenging and not easily obtainable. So, the PD detection model used the voice dataset obtained from the UCI repository. Also, the transparency and credibility of the classifiers are added to a clinical dataset to develop a generalised resilient PD diagnostic tool. However, Tables 4 and 5 serve as a witness for the diagnostic tool’s efficiency. Additionally, [55] addressed the issues related to PD, suggesting that PD is stubborn to most of the treatment regimens except for a few invasive treatments. The present diagnostic tool employs a non-invasive mechanism, which can adopt to and learn from a new patient dataset, leading to a better solution continuously. Further, the study can be extended to improve the accuracy of diagnosis by employing other deep learning algorithms such as Light GBM; XGBoost and CatBoost along with multilevel classification models. The current study has utilised voice data for the detection of PD in the R platform. However, future researchers can develop point-click and drop-down user interface packages for PD diagnosis to enhance the usability of the diagnostic tool among the practitioners. Also, for PD detection, other motor symptoms can be used for early prediction of disease, such as handwriting and walking. These symptoms can be used to build models with multiple symptoms and can be used for early detection, leading to better patient care and well-being. Further, the proposed deployment framework requires less expert support for diagnosis. However, the constant update of new case findings can make the framework more robust which was not included in the study. Further, the model used English language and data from a particular country for the prediction modelling. However, this limitation can be overcome by collecting voice samplings from different geographical locations. Further, authors used the data which was analysed and vetted by previous studies through TQWT methods. However, in the future works authors has to vet the patient data before using it for prediction modelling. Furthermore, the present study focused on the PD, while other neurogenerative diseases like dementia and Alzheimer's requires a primary screening process which researcher can analyse and develop to deliver better patient care and prevent the healthy individuals from clinical examination and non-invasive procedures.

References

[1] J. T. Dudley, R. Miotto, F. Wang, S. Wang, and X. Jiang, “Deep learning for healthcare: review, opportunities and challenges,” *Brief. Bioinform.*, vol. 19, no. 6, pp. 1236–1246, 2017.

[2] Y. Wang *et al.*, “Artificial intelligence in healthcare: past, present and future,” *Stroke Vasc. Neurol.*, vol. 2, no. 4, pp. 230–243, 2017.

[3] D. H. Smaltz, V. Sambamurthy, and R. Agarwal, “The antecedents of CIO role effectiveness in organizations: An empirical study in the healthcare sector,” *IEEE Trans. Eng. Manag.*, vol. 53, no. 2, pp. 207–222, 2006.

[4] J. P. Devarajan, M. A, and V. R. Sreedharan, “Healthcare Operations and Black Swan Event for COVID-19 Pandemic: A Predictive Analytics,” *IEEE Trans. Eng. Manag.*, 2021.

[5] M. Pérez-Ortiz, S. Jiménez-Fernández, P. A. Gutiérrez, E. Alexandre, C. Hervás-Martínez, and S. Salcedo-Sanz, “A review of classification problems and algorithms in renewable energy applications,” *Energies*, vol. 9, no. 8, pp. 1–27, 2016.

[6] S. Makridakis, R. Kirkham, A. Wakefield, M. Papadaki, J. Kirkham, and L. Long, “Forecasting, uncertainty and risk; perspectives on clinical decision-making in preventive and curative medicine,” *Int. J. Forecast.*, vol. 35, no. 2, pp. 659–666, 2018.

[7] M. Elmontsri, R. Banarsee, and A. Majeed, “Improving patient safety in developing countries – moving towards an integrated approach,” *JRSM Open*, vol. 9, no. 11, p. 205427041878611, 2018.

[8] A. Sheikhtaheri, M. Sadeqi-Jabali, and Z. Hashemi-Dehaghi, “Physicians’ perspectives on causes of health care errors and preventive strategies: A study in a developing country,” *Iran. J. Public Health*, vol. 47, no. 5, pp. 720–728, 2018.

[9] Y. Liu, J. W. Bi, and Z. P. Fan, “Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms,” *Expert Syst. Appl.*, vol. 80, pp. 323–339, 2017.

[10] A. Wulff, B. Haarbrandt, E. Tute, M. Marschollek, P. Beerbaum, and T. Jack, “An interoperable clinical decision-support system for early detection of SIRS in pediatric intensive care using openEHR,” *Artif. Intell. Med.*, vol. 89, no. April, pp. 10–23, 2018.

[11] G. Saposnik, D. Redelmeier, C. C. Ruff, and P. N. Tobler, "Cognitive biases associated with medical decisions: a systematic review," *BMC Med. Inform. Decis. Mak.*, vol. 16, no. 1, pp. 1–14, 2016.

[12] D. J. Persis, A. S, V. Sunder M, R. G, V. R. Sreedharan, and T. Saikouk, "Improving patient care at a multi-speciality hospital using lean six sigma," *Prod. Plan. Control*, 2020.

[13] K. Buchan, M. Filannino, and Ö. Uzuner, "Automatic prediction of coronary artery disease from clinical narratives," *J. Biomed. Inform.*, vol. 72, pp. 23–32, 2017.

[14] Y. H. Hsiao *et al.*, "Sleep disorders and an increased risk of Parkinson's disease in individuals with non-apnea sleep disorders: a population-based cohort study," *J. Sleep Res.*, vol. 26, no. 5, pp. 623–628, 2017.

[15] I. Cova and A. Priori, "Diagnostic biomarkers for Parkinson's disease at a glance: where are we?," *J. Neural Transm.*, vol. 125, no. 10, pp. 1417–1432, 2018.

[16] F. Oyebode, "Clinical errors and medical negligence," *Med. Princ. Pract.*, vol. 22, no. 4, pp. 323–333, 2013.

[17] H. Singh and D. F. Sittig, "Advancing the science of measurement of diagnostic errors in healthcare: The Safer Dx framework," *BMJ Qual. Saf.*, vol. 24, no. 2, pp. 103–110, 2015.

[18] R. A. Berenson, D. K. Upadhyay, and D. R. Kaye, "Placing Diagnosis Errors on the Policy Agenda," no. April, p. 19, 2014.

[19] J. M. Mateos-Pérez, M. Dadar, M. Lacalle-Aurioles, Y. Iturria-Medina, Y. Zeighami, and A. C. Evans, "Structural neuroimaging as clinical predictor: A review of machine learning applications," *NeuroImage Clin.*, vol. 20, no. August, pp. 506–522, 2018.

[20] P. Croskerry, "Perspectives on diagnostic failure and patient safety.," *Healthc. Q.*, vol. 15 Spec No, pp. 50–56, 2012.

[21] M. L. Graber, R. M. Wachter, and C. K. Cassel, "Bringing diagnosis into the quality and safety equations," *JAMA - J. Am. Med. Assoc.*, vol. 308, no. 12, pp. 1211–1212, 2012.

[22] Institute of Medicine of the National Academies, "Finding What Works in Health Care: Standards for Systematic Reviews," *Medicine (Baltimore).*, 2011.

[23] S. . Durning, "Submitted input. Input submitted to the Comittee on Diagnostic Error in Health Care," 2014.

[24] G. Klein and D. Kahneman, "Conditions for Intuitive Expertise: A Failure to Disagree," *Am. Psychol.*, vol. 64, no. 6, pp. 515–526, 2009.

[25] A. N. D. Meyer, V. L. Payne, D. W. Meeks, R. Rao, and H. Singh, “Physicians’ diagnostic accuracy, confidence, and resource requests: A vignette study,” *JAMA Intern. Med.*, vol. 173, no. 21, pp. 1952–1961, 2013.

[26] D. Braga, A. M. Madureira, L. Coelho, and R. Ajith, “Automatic detection of Parkinson’s disease based on acoustic analysis of speech,” *Eng. Appl. Artif. Intell.*, vol. 77, no. October 2018, pp. 148–158, 2019.

[27] D. M. Radhakrishnan and V. Goyal, “Parkinson’s disease - A review,” *Neurol India*, vol. 66, pp. 26–35, 2018.

[28] X. Zhang *et al.*, “Data-Driven Subtyping of Parkinson’s Disease Using Longitudinal Clinical Records: A Cohort Study,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.

[29] R. Levy, P. Ashby, W. D. Hutchison, A. E. Lang, A. M. Lozano, and J. O. Dostrovsky, “Dependence of subthalamic nucleus oscillations on movement and dopamine in Parkinson’s disease,” *Brain*, vol. 125, no. 6, pp. 1196–1209, 2002.

[30] B. H. Wood, J. A. Bilclough, A. Bowron, and R. W. Walker, “Incidence and prediction of falls in Parkinson’s disease: A prospective multidisciplinary study,” *J. Neurol. Neurosurg. Psychiatry*, vol. 72, no. 6, pp. 721–725, 2002.

[31] R. B. Postuma, J. F. Gagnon, S. Rompré, and J. Y. Montplaisir, “Severity of REM atonia loss in idiopathic REM sleep behavior disorder predicts Parkinson disease,” *Neurology*, vol. 74, no. 3, pp. 239–244, 2010.

[32] M. B. Stern and A. Siderowf, “Parkinson’s at risk syndrome: Can Parkinson’s disease be predicted?,” *Mov. Disord.*, vol. 25, no. SUPPL. 1, 2010.

[33] G. K. Kerr, C. J. Worringham, M. H. Cole, P. F. Lacherez, J. M. Wood, and P. A. Silburn, “Predictors of future falls in Parkinson disease,” *Neurology*, vol. 75, no. 2, pp. 116–124, 2010.

[34] A. Schrag, U. F. Siddiqui, Z. Anastasiou, D. Weintraub, and J. M. Schott, “Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson’s disease: a cohort study,” *Lancet Neurol.*, vol. 16, no. 1, p. 66, 2017.

[35] O. Oguh and A. Videnovic, “Inpatient Management of Parkinson Disease: Current Challenges and Future Directions,” *The Neurohospitalist*, vol. 2, no. 1, pp. 28–35, 2012.

[36] K. Krzyszton, J. Stolarski, and J. Kochanowski, “Evaluation of balance disorders in Parkinson’s disease using simple diagnostic tests - Not so simple to choose,” *Front. Neurol.*, vol. 9, 2018.

[37] J. Zhang, X. Li, and J. Da Li, “The Roles of Post-translational Modifications on α-Synuclein in the Pathogenesis of Parkinson’s Diseases,” *Front. Neurosci.*, vol. 13, no. APR, 2019.

[38] C. O. Sakar *et al.*, “A comparative analysis of speech signal processing algorithms for Parkinson’s disease classification and the use of the tunable Q-factor wavelet transform,” *Appl. Soft Comput. J.*, vol. 74, pp. 255–263, 2019.

[39] M. A. Little *et al.*, “Suitability of dysphonia measurements for telemonitoring of Parkinson’s disease,” *IEEE Trans Biomed Eng*, vol. 56, no. 4, 2011.

[40] S. Yang *et al.*, “Effective dysphonia detection using feature dimension reduction and kernel density estimation for patients with parkinson’s disease,” *PLoS One*, vol. 9, no. 2, 2014.

[41] M. Thomas, A. Lenka, and P. Kumar Pal, “Handwriting Analysis in Parkinson’s Disease: Current Status and Future Directions,” *Mov. Disord. Clin. Pract.*, vol. 4, no. 6, pp. 806–818, 2017.

[42] M. Pérez-Ortiz, S. Jiménez-Fernández, P. A. Gutiérrez, E. Alexandre, C. Hervás-Martínez, and S. Salcedo-Sanz, “A review of classification problems and algorithms in renewable energy applications,” *Energies*, vol. 9, no. 8. pp. 1–27, 2016.

[43] D. R. Schrider, J. Ayroles, D. R. Matute, and A. D. Kern, “Supervised machine learning reveals introgressed loci in the genomes of Drosophila simulans and D. sechellia,” *PLOS Genet.*, vol. 14, no. 4, p. e1007341, Apr. 2018.

[44] S. S. Ali, R. Kaur, D. J. Persis, R. Saha, M. Pattusamy, and V. R. Sreedharan, “Developing a hybrid evaluation approach for the low carbon performance on sustainable manufacturing environment,” *Ann. Oper. Res.*, 2020.

[45] S. S. Ali, F. Ersöz, R. Kaur, B. Altaf, and G. W. Weber, “A quantitative analysis of low carbon performance in industrial sectors of developing world,” *J. Clean. Prod.*, vol. 284, 2021.

[46] S. Samar Ali, R. Kaur, F. Ersöz, L. Lotero, and G. W. Weber, “Evaluation of the effectiveness of green practices in manufacturing sector using CHAID analysis,” *J. Remanufacturing*, vol. 9, no. 1, pp. 3–27, 2019.

[47] H. L. Chen, D. Y. Liu, B. Yang, J. Liu, and G. Wang, “A new hybrid method based on local fisher discriminant analysis and support vector machines for hepatitis disease diagnosis,” *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11796–11803, 2011.

[48] N. Javaid *et al.*, “Comparative Analysis of Classifiers for Developing an Adaptive Computer-Assisted EEG Analysis System for Diagnosing Epilepsy,” *Biomed Res. Int.*,

vol. 2015, pp. 1–14, 2015.

[49] P. Probst, A. L. Boulesteix, and B. Bischl, “Tunability: Importance of hyperparameters of machine learning algorithms,” *J. Mach. Learn. Res.*, vol. 20, pp. 1–32, 2019.

[50] S. Abe, *Support Vector Machines for Pattern Classification*. Springer verlag, 2005.

[51] B. J. Peterson, W. A. Rocca, J. H. Bower, R. Savica, and M. M. Mielke, “Identifying incident Parkinson’s disease using administrative diagnostic codes: a validation study,” *Clin. Park. Relat. Disord.*, vol. 3, pp. 1–5, 2020.

[52] A. Abraham, R. P. Duncan, and G. M. Earhart, “The role of mental imagery in Parkinson’s disease rehabilitation,” *Brain Sci.*, vol. 11, no. 2, pp. 1–14, 2021.

[53] N. Gerdsri, P. Assakul, and R. S. Vatananan, “An activity guideline for technology roadmapping implementation,” *Technol. Anal. Strateg. Manag.*, vol. 22, no. 2, pp. 229–242, 2010.

[54] R. S. Vatananan and N. Gerdsri, “The current state of technology roadmapping (trm) research and practice,” *Int. J. Innov. Technol. Manag.*, vol. 9, no. 4, 2012.

[55] N. Gerdsri, “Strategic Planning: A Quantitative Model for the Strategic Evaluation of Emerging Technologies,” pp. 97–119, 2016.

Author biography

D Jinil Persis* received her Master's and Ph.D. degrees in Engineering with Industrial Engineering specialization from College of Engineering, Guindy (CEG), Chennai, India, in 2011 and 2015, respectively. She is currently an Assistant Professor working with the Operations and Supply Chain Management area at National Institute of Industrial Engineering (NITIE), Mumbai. Her research interests include multi-objective optimization and its applications in network problems, machine learning models for healthcare, agriculture, and manufacturing applications.

V. Raja Sreedharan V. Raja Sreedharan received the Graduate degree in industrial engineering from the College of Engineering Guindy, Chennai, India, in 2013, and the Ph.D. degree in lean six sigma from the College of Engineering Guindy, Anna University, Chennai, India, in 2017. He is currently an Assistant Professor with Supply Chain Management, BEAR Lab (Business, Economie et Actuariat), Rabat Business School, Universite Internationale de Rabat, Morocco. He has published works in peer reviewed journals. His current research interest focuses on decision making for VUCA in business operations and process improvement for services industries. He also consults in the field of process optimisation with industries in private sectors.

Gopalakrishnan Narayanamurthy received his undergraduate degree in biomedical engineering from Anna University, Chennai, India, and his doctoral degree from the Indian Institute of Management Kozhikode, Kozhikode, India, in 2012 and 2018, respectively. During his doctoral studies, he has been a recipient of the Fulbright-Nehru Doctoral Research Fellowship. He is currently a Senior Lecturer in the Department of Operations and Supply Chain Management at the University of Liverpool Management School (ULMS), UK. His current research interests are in the area of healthcare operational excellence, transformative service research, disruptive technologies, and digitization.

Corresponding author*