Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning

WU WANG, JUNHO LEE, FOUZI HARROU, (Member, IEEE), AND YING SUN

Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

Corresponding author: Fouzi Harrou (fouzi.harrou@kaust.edu.sa)

ABSTRACT Accurately detecting Parkinson's disease (PD) at an early stage is certainly indispensable for slowing down its progress and providing patients the possibility of accessing to disease-modifying therapy. Towards this end, the premotor stage in PD should be carefully monitored. An innovative deep-learning technique is introduced to early uncover whether an individual is affected with PD or not based on premotor features. Specifically, to uncover PD at an early stage, several indicators have been considered in this study, including Rapid Eye Movement and olfactory loss, Cerebrospinal fluid data, and dopaminergic imaging markers. A comparison between the proposed deep learning model and twelve machine learning and ensemble learning methods based on relatively small data including 183 healthy individuals and 401 early PD patients shows the superior detection performance of the designed model, which achieves the highest accuracy, 96.45% on average. Besides detecting the PD, we also provide the feature importance on the PD detection process based on the Boosting method.

INDEX TERMS Parkinson's disease, deep learning, ensemble learning, early detection, premotor features, features importance.

I. INTRODUCTION

Parkinson's disease (PD) is becoming an important degenerative disease of the central nervous system, affecting the quality of lives of millions of seniors worldwide [1]. Symptoms of PD can progress differently from one person to another because of the variety of the disease. Patients with Parkinson may show symptoms including tremors mainly at rest. Different types of tremors are possible: tremors in hands, limb rigidity, and gait and balance problems. Generally, two types of symptoms of PD can be distinguished: movement-related (i.e., motor) and unrelated to movement (non-motor). In fact, patients showing non-motor symptoms are more affected than whose main symptoms are motor. Non-motor symptoms may include depression, sleep behavior disorders, loss of sense of smell, and cognitive impairment. It has been reported by the Centers for Disease Control and Prevention (CDC) that PD complications are ranked as the 14th leading causes of death in the United States. To date, the cause of PD rests principally unknown. Particularly, the economic burden

The associate editor coordinating the review of this manuscript and approving it for publication was Fatih Emre Boran.

due to direct and indirect cost of PD covering treatment, social security payments, and lost income is estimated to be approximately $52 billion per year in the United States alone. Actually, the number of people affected by PD has exceeded 10 million worldwide. It should be noted that the timely detection of the PD facilitates rapid treatment and alleviate symptoms significantly as reported in [2]. Therefore, detection of PD at an earlier stage is certainly a key element to slowing down its progression and could give patients the possibility of accessing to disease-modifying therapy, when available.

Till now, there is no way to diagnose Parkinson's disease (PD) [2]. However, there are various symptoms and diagnostic tests used in combination. Several biomarkers have been investigated by scientists to early identify PD to slow down the disease process. Currently, all therapies used for PD improve symptoms without slowing or halting the disease progression. Various methods have proposed to help detection PD based on different kinds of measurements including speech data [3]–[6], gait patterns [7], force tracking data [8], smell identification data [9] and spontaneous cardiovascular oscillations [10]. In [11], an approach using the sawtooth

inspired pitch estimator (SWIPE) scheme is used to assess speech disorders recorded via smartphone caused by Parkinson's disease. Acceptable results have been achieved by the SWIPE scheme in discriminating PD from healthy patients. However, at a lower signal to noise ratio level, an improved algorithm is needed to obtain efficient robustness to noise. In [12], an early detection algorithm of PD based on reduced vocal features is designed. It has been illustrated that the use of Wrappers subset selection is suitable because of the low dimensionality of the selected feature and improved PD detection capability. In [13], a PD detection system is introduced using a 1D convolutional neural network based on the gait signals. However, the performance of PD detection based on both speech and gait analyses is generally limited by the sensitivity to background noise in speech recording, causing a high number of false alarms and missed detection. Also, gait tracking and inspection need specialized devices and sufficient space for walking [14]. The authors of [15] suggest a method based on wavelet to analyze data collected from smartwatches worn by nineteen patients affected by PD. This method showed good ability in detecting symptoms of tremor, bradykinesia, and dyskinesia. In [16], an approach to detect motor impairment in PD based on mobile touchscreen typing is introduced. Essentially, the proposed algorithm uncovers signs of PD motor by analyzing touchscreen typing features that include descriptive statistics (covariance, skewness, and kurtosis) and time information. In [17], multi-source of data including imaging, genetics, clinical and demographic data are incorporated in developing models for PD prediction. Other approaches employed handwriting measurement for Parkinson's diagnosis [14], [18], [19]. In [14], the PD diagnosis approach has been proposed based on handwriting measurements gathered from patients with PD. It has been shown that improved PD diagnosis is obtained when taking into consideration the age and sex information in the decision process [13], [14].

Accurate and early detection of PD is vital due to its ability to provide crucial information to slow down the progression of PD. All over the years, various data-driven methods have been developed to improve the detection of PD. In contrast to the model-based detection techniques, where prior availability of an analytical model is required, in data-driven techniques, only the availability of historical data is needed. Recently, machine learning (ML) has emerged as a promising field of research in PD diagnosis, both in academia and industry [20]. Owing to its data-driven approaches, ML has brought a paradigm shift in the way relevant information in PD biomarkers are extracted and analyzed. Furthermore, machine learning techniques provide pertinent information that offers guidance related to PD classification and diagnosis to speed up decision making. Various machine learning techniques have been applied in the literature to address the PD detection problem. For instance, in [21], dysphonia measurements have been used to detect patients with PD from healthy people. The support vector machine (SVM) is applied to only four dysphonic features for PD classification due to

its ability to extract nonlinearity by using nonlinear kernels. In [6], three common machine learning algorithms, namely Random Forest (RF) or Support Vector Machine (SVM) and neural network, have been applied to detect Parkinson's disease based on acoustic analysis of speech. It has been shown the promising results of RF an SVM in early PD detection. In [22], the performance of four classifiers, Decision Trees, Regression, DMneural, and Neural Networks (NN), has been compared in detecting PD, and the best accuracy of 92.9% is obtained using NN algorithm. Recently, deep learning-based techniques have gained special attention in PD diagnosis due to their capacity to handle big data and achieving high accuracy with free-assumption on data distribution [3], [23]. Authors in [23] applied a Long short term memory algorithm to detect the Freezing of Gait (FOG), which a good indicator of PD patients that may cause falling. It has been shown that LSTM outperforms the SVM in detecting FOG.

To guarantee early detection of PD, the premotor or prodromal stage in PD should be carefully monitored [24], [25]. This premotor stage is generally characterized by different symptoms than the usual motor symptoms, including Rapid Eye Movement (REM) sleep Behaviour Disorder (RBD) and olfactory loss [24]. The purpose of the present paper is threefold. Firstly, to uncover PD at an early stage, several indicators have been considered in this study, including RBD and olfactory loss, Cerebrospinal fluid (CSF) data, and SPECT imaging markers. Secondly, Still within the data-driven techniques, this paper is aimed at presenting a comparative study between the most advanced data-driven prediction methods in detecting PD. Here, three kinds of data-driven methods are compared: shallow machine learning-based, ensemble learning-based, and deep learning-based. In this work, a deep learning model is designed to discriminate normal individuals and patients affected by PD. Essentially, the aim of this study is to provide a comparative study and throw light on the performance of these advanced prediction methods when applied to small PD data sets. Indeed, the used PD data from the Parkinson's Progression Markers Initiative (PPMI) is relatively small and includes features from 183 healthy individuals and 401 early PD patients, which may make the application the machine learning methods attractive to investigate under this small dataset problem. Results showed that the designed deep learning offers superior detection performance compared to the twelve considered machine learning models in discriminating normal people with patients who have Parkinson's disease. Lastly, feature importance and selection frequency computed based on the Boosting method highlight the largest impact of the imaging markers of SBR's for left and right putamen on the PD detection process.

The next section presents the involved PD dataset and provides a brief description of the proposed deep learning model and the considered machine learning models. Section III discusses the PD detection results and comparisons. Lastly, conclusions are drawn in section IV.
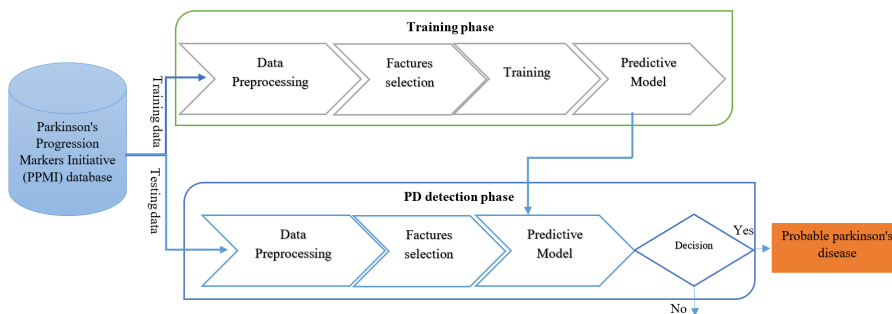
FIGURE 1. Flowchart of the proposed PD detection procedure.

II. DATA AND METHODS

This study proposes a deep learning framework for the early detection of PD. The general framework of the proposed detection approach is illustrated in Figure 1. PD detection is done into two main stages: training and testing. In the first stage, the raw data is preprocessed and standardized and then it is used to construct the deep learning model. The values of parameters of deep learning models are selected such that the loss function is minimized during the training. After that, in the testing stage, the previously constructed model with the selected parameters are used for PD detection.

A. PPMI DATA

For our study, we prepare the data from the Parkinson's Progression Markers Initiative (PPMI) database (http://www.ppmi-info.org/data). The PPMI is a landmark observational clinical study to comprehensively evaluate cohorts of significant interest using advanced imaging, biologic sampling and clinical and behavioral assessments to identify biomarkers of Parkinson's disease progression [26].

The PPMI cohort comprises 401 early PD patients and 183 healthy individuals followed longitudinally for clinical, imaging and biospecimen biomarker assessment. All the PD patients are in their earlier stages of PD (stage 1 or 2 in Hoehn and Yahr scale [27]), and all the health subjects are 30 years or older wihtout PD. To detect the early PD, we consider thirteen features based on information from previous similar studies [24], [25], as provided below.

- RBDSQ score. The REM sleep Behavior Disorder Screening Questionnaire (RBDSQ) is a specific questionnaire for rapid eye movement behavior disorder (RBD) [28]. The higher score indicates the more likely to have RBD.
- UPSIT score. The University of Pennsylvania Smell Identification Test (UPSIT) evaluates olfactory function [29]. It consists of comprehensive 40 questions. The maximum test score is 40 when all the odors are identified correctly.
- α-syn. A Cerebrospinal fluid (CSF) biomarker of α-synuclein.
- $A\beta$1-42. A CSF biomarker of amyloid beta peptide 1-42.

- P-tau$_{181}$. A CSF biomarker of tau phosphorylated at threonine 181.
- T-tau. A CSF biomarker of total tau.
- P-tau$_{181}$/$A\beta$1-42. The ratio between P-tau$_{181}$ and $A\beta$1-42.
- P-tau$_{181}$/T-tau. The ratio between P-tau$_{181}$ and T-tau.
- T-tau/$A\beta$1-42. The ratio between T-tau and $A\beta$1-42.
- CAUDATE_L. The striatal binding ratio (SBR) of the left caudate from Single Photon Emission Computed Tomography (SPECT) imaging.
- CAUDATE_R. The SBR of the right caudate from SPECT imaging.
- PUTAMEN_L. The SBR of the left putamen from SPECT imaging.
- PUTAMEN_R. The SBR of the right putamen from SPECT imaging.

More details about CSF biomarkers and SPECT imaging can be found in the PPMI manuals (Biologics Manual, SPECT Manual, http://www.ppmi-info.org/study-design/research-documents-and-sops/)

B. EXPLORATORY ANALYSIS

Figure 2 shows the histogram of the features. Most of the covariates, except UPSIT scores, are severely right-skewed. Some of the covariates, e.g., α-syn and T-tau, have large outliers. The feature RBDSQ score takes discrete values. To reduce the degree of skewness and reduce the influence of outlier observations on the models to be built, we log-transform all the features except RBDSQ scores in the following analysis. Also note that the scale of the features varies from thousands, e.g., $A\beta$1-42, to less than 1, e.g., T-tau/$A\beta$1-42. We center and scale the features after the log-transformation unifies the scales of the features.

From Figure 2, we can readily identify five features that have significantly distinct distributions between healthy people and patients with Parkinson's disease. These five features are CAUDATE_L, CAUDATE_R, PUTAMEN_L, PUTMEN_R, and UPSIT score. The patients with Parkinson's disease have lower scores in all of the five features. These five features are critical in discriminating healthy people with patients with Parkinson's disease. Note that the marginal relationship between features and disease indicators shown
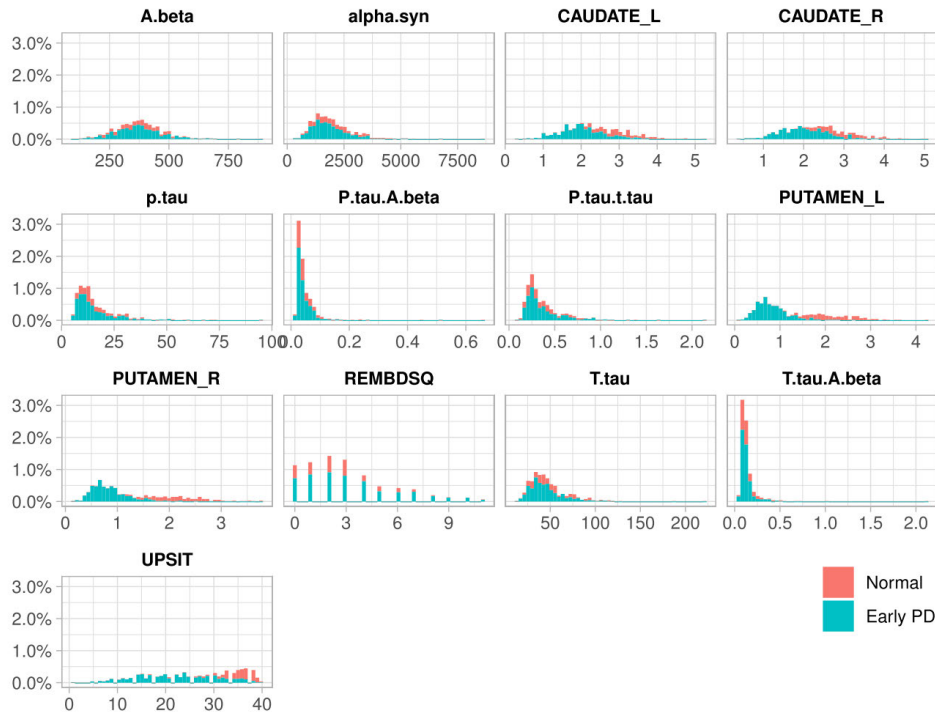
FIGURE 2. The histogram of the features.

in Figure 2 does not exclude the possibility that other features might also be important in identifying Parkinson patients, because the features may have a nonlinear relationship with the disease indicator, and the features may interact with each other and be distinguishable in high-dimensions.

The Pearson correlation coefficient among all the features is shown in Figure 3. The features CAUDATE_L, CAUDATE_R, PUTAMEN_L, and PUTMEN_R are highly correlated with each other.

C. DESCRIPTION OF THE METHODS

In this article, we use machining learning methods to study the link between Parkinson's disease and build a model for early diagnosis of Parkinson's disease. Denote the training data by $\{(\mathbf{X}_1, y_1), \ldots, (\mathbf{X}_n, y_n)\}$, the following classifiers are used in our study.

- Deep Learning (DEEP). In recent years, deep learning algorithms have achieved striking performances in fields such as computer vision [30], natural language processing [31], and speech recognition [32]. Deep learning is also changing other fields such as biology [33] and engineering [34]. The deep learning system is composed of multiple layers of connected artificial neurons. The neurons are information processing modules, which essentially are simple nonlinear transformations of inputs. For the supervised feed-forward neural network (FNN) considered in this article, when raw-data are fed into the network, the deep learning algorithm can automatically extract hierarchical representations of the data which
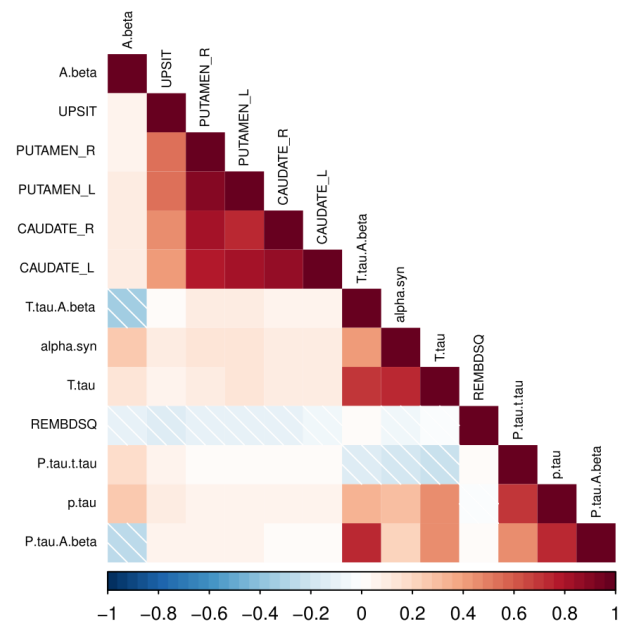

FIGURE 3. The pearson correlation coefficient between the considered features.

are best suited for the underlying learning task, e.g., classification in our paper. In this article, we construct an FNN which has two hidden layers. The first hidden layer has twenty neurons, and the second hidden layer has ten neurons. The structure of the network is shown in Figure 4. The deep learning algorithms also have some weaknesses. Usually, training deep learning
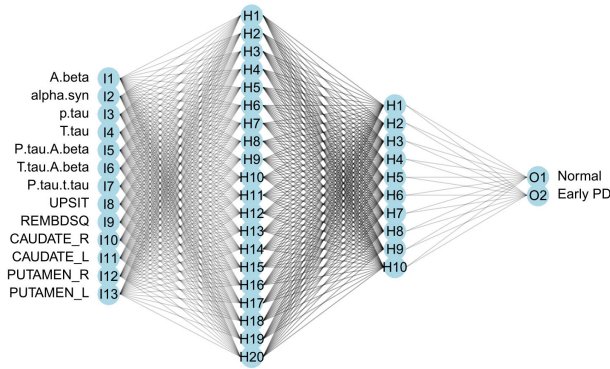
A.beta I1
alpha.syn I2
p.tau I3
T.tau I4
P.tau.A.beta I5
T.tau.A.beta I6
P.tau.t.tau I7
UPSIT I8
REMBDSQ I9
CAUDATE_R I10
CAUDATE_L I11
PUTAMEN_R I12
PUTAMEN_L I13

H1 H2 H3 H4 H5 H6 H7 H8 H9 H10 H11 H12 H13 H14 H15 H16 H17 H18 H19 H20

H1 H2 H3 H4 H5 H6 H7 H8 H9 H10

O1 Normal
O2 Early PD

FIGURE 4. The architecture of the forward neural network model we built.

algorithms needs large amount of data to achieve desired accuracy. Deep learning is routinely used as a black box algorithm, it is hard to interpret the trained neural networks. Also, theoretically, it is still hard to understand why and how deep learning achieves good performances.

- Classification trees (TREE). Classification tree [35] recursively splits the feature space into sub-sets such that the Gini impurity is minimized at each step. The Gini impurity is defined as

$$Gini(p_0, p_1) = 1 - p_0^2 - p_1^2, \qquad (1)$$

where p_0 and p_1 are the proportion of normal people and patients, respectively. The classification tree algorithm first grows a tree to the maximum depth such that each leaf node is pure, then prunes upwards to balance the classification error and the number of terminal nodes of the tree, that is minimizing

$$R(T) + \alpha|T|, \qquad (2)$$

where $R(T)$ is the classification error of a candidate tree T, $|T|$ is the number of terminal nodes of the tree T, and the parameter α which is a trade-off between estimation error and size is selected by cross-validation.

- Boosting. Boosting is an ensemble algorithm designed to convert the performance of weak base learners, such as shallow regression trees, to strong learners [36]. References [37] and [38] discovered the connection between the Boosting algorithm and estimation in functional space, thus opened the way for applications other than classification. In this article, we implemented three Boosting algorithms with different base learners. All the Boosting algorithms use the cross-entropy as the loss function. The first Boosting algorithm adopts classification trees [35] as its base learner, and we abbreviate it as BOOST_TREE. Moreover, we also employ linear models of one variable, i.e., $\beta_j X_j$ where X_j is the jth element of the features $\mathbf{X}$ and β_j is a coefficient, and B-splines of one variable, i.e., $b(X_j)$ where $b(.)$ is a cubic spline function, as the base learners, and we abbreviate the

corresponding Boosting algorithms as BOOST_GLM and BOOST_GAM, respectively. Besides classification, BOOST_GLM and BOOST_GAM have the additional feature of variable selection, i.e., they can automatically select variables during the training process [39].

- Random forest (RF). Random forest [40] is closely related to the boosting algorithm because both algorithms aggregate the results of a cluster of base learners. RF trains a cluster of classification trees using bootstrap samples of the training data. A key feature of random forest is that it de-correlates the trees by randomizing the candidate splits when building the trees. That is, the actual splitting feature is selected from a random subset of the features. Let $\mathbf{X}_s = \{X_{i_1}, \ldots, X_{i_k}\}$ be a random subset of the features $\mathbf{X} = \{X_1, \ldots, X_p\}$, where $k \leq p$ denotes the number of candidates. RF restricts the input feature to be $\mathbf{X}_s$ when minimizing the Gini impurity (1). The number k is a major parameter of RF, and it is selected by cross-validation in our implementation.

- Logistic regression (LOGIS). Logistic regression is a member of the generalized linear model family, which is backed up by rich statistical theory. In logistic regression, we model the conditional probability of having the disease $\pi_i = p(y_i = 1|\mathbf{X}_i)$ as

$$log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{X}_i^\top \boldsymbol{\beta}, \qquad (3)$$

and the coefficient $\boldsymbol{\beta}$ is estimated by maximum the log-likelihood function

$$\sum_{i=1}^{n} \left\{ y_i \mathbf{X}_i^\top \boldsymbol{\beta} - \log\left(1 + \exp\left(\mathbf{X}_i^\top \boldsymbol{\beta}\right)\right) \right\}. \qquad (4)$$

Besides ordinary logistic regression, we also considered penalized logistic regression (LOGIS_PEN), where $\boldsymbol{\beta}$ is estimated by

$$\sum_{i=1}^{n} \left\{ y_i \mathbf{X}_i^\top \boldsymbol{\beta} - \log\left(1 + \exp\left(\mathbf{X}_i^\top \boldsymbol{\beta}\right)\right) \right\} - \sum_{j=1}^{p} |\beta_j|, \quad (5)$$

where p is the number of features. The penalized logistic regression can select features from the feature sets and fit a parsimonious model [41].

- Discriminant analysis (DIS). In linear discriminant analysis, the conditional distribution $p(\mathbf{X}_i|y_i = 0)$ and $p(\mathbf{X}_i|y_i = 1)$ are assumed to be normally distributed with common variance Σ and mean $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$, respectively. For a new input $\mathbf{X}$, if

$$\mathbf{X}^\top \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) > \frac{1}{2}\left(\boldsymbol{\mu}_1^\top \Sigma^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^\top \Sigma^{-1}\boldsymbol{\mu}_0\right), \quad (6)$$

then the case is classified to be a patient, otherwise the case is classified as a normal people.

- K-nearest neighbor (KNN). In KNN, a new case is classified as a patient if more than half of its k-nearest neighbors measured by the distance between the corresponding features are patients. The number of neighbors k used in the algorithm is selected by cross-validation.

- Support vector machines (SVM). SVM maps the features into a high-dimensional space using the kernel trick and builds a hyperplane in the mapped space which optimally separate the patients and normal people. In SVM, a patient is coded as $y = 1$ whereas a normal people is coded as $y = -1$. Let $k(\mathbf{X}_i, \mathbf{X}_j) = \exp\left(-\gamma \|\mathbf{X}_i - \mathbf{X}_j\|^2\right)$ be the Gaussian radial basis kernel where $\gamma > 0$ is a parameter. SVM classifies a new case with feature $\mathbf{X}$ as a patient if $\sum_{i=1}^{n} c_i y_i k(\mathbf{X}_i, \mathbf{X}) > 0$, where c_i, $i = 1, \dots, n$ are determined by the optimization problem

$$\max \sum_{i=1}^{n} c_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j y_i y_j k(\mathbf{X}_i, \mathbf{X}_j)$$

$$\text{subject to } \sum_{i=1}^{n} c_i y_i = 0, \quad 0 \le c_i \le \frac{1}{2n\lambda}, \text{ for all } i.$$

The parameters $\gamma > 0$ and $\lambda > 0$ are determined by cross-validation.

D. IMPLEMENTATION DETAILS AND COMPUTATION COSTS

To evaluate the accuracy of the above algorithms on Parkinson's disease discrimination, we randomly split the data and use 70% as training data and the rest as testing data. The ratio of patients with healthy people in the training and testing data is kept the same as the original data using stratified sampling. We train the machine learning methods on the training data, and predict whether the cases in the test data are Parkinson patients or not using the trained model. The splitting is repeated 100 times. We report the performance measures on the testing data using accuracy, specificity, sensitivity and area under the ROC curve (AUC), precision and F1 in Section III.

We train deep learning models in mini-batches of 16 samples using the stochastic gradient descent (SGD) algorithm with the cross-entropy function as the loss. The algorithm is trained with dropouts [42] to prevent overfitting. We also use batch normalization [43] to accelerate training. The history of the loss function and the accuracy of the network on the testing data in one of the cross-validation data for 100 epochs is shown in Figure 5. For each epoch, the computational cost of a naive implementation of SGD is about $O(nm)$, where n is the number of samples, and m is the number of weights and biases in the network. That is, the computational cost of training a network increases as the number of samples, the depth of the network and the number of neurons in each layer increase. With modern parallel computing devices, the training of deep learning methods can be largely accelerated.

We train three deep learning models with different structures to show robustness of the results with respect to the hyperparameters. The three trained models are all feed-forwad neural networks with two hidden layers. The first network, abbreviated as DEEP1, has 40 and 20 neurons in the first and second hidden layers, respectively. The second
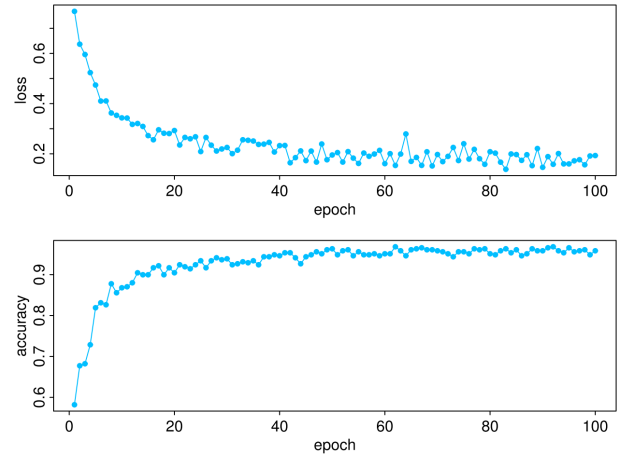

FIGURE 5. The history of the risk function and the accuracy on the testing data in one cross-validation data set.

network (DEEP2) has has 20 and 20 neurons in the first and second hidden layers, respectively. The third network (DEEP3) has has 20 and 10 neurons in the first and second hidden layers, respectively. Lastly, we combine the three deep learning models to form an ensemble (DEEP_EN) by averaging the outputs. The deep learning models are trained with 25 and 50 epochs of data, the results are abbreviated as DEEP1(25), DEEP1(50),…etc.

To make the comparisons as fair as possible, we tune all other methods to achieve their best performance. For example, the number of candidate splits in RF are selected by out-of-bag error rates. This method is recommended by Breiman [40]. The number of trees in the boosting methods, the number of neighbors in KNN and the penalty parameter in penalized logistic regression are all selected by 5-fold cross-validation. For the TREE method, a large tree is first grown to its maximum depth, and the pruned by 5-fold cross-validation [35]. For the SVM method, we use a Gaussian kernel, and the parameters are selected by 5-fold cross-validation.

III. RESULTS
A. METRICS
To evaluate the performance of machine methods for discriminating Parkinson patients, we employ the following criterion:

- Accuracy $= \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}}$,
- Sensitivity $= \frac{\text{TP}}{\text{TP+FN}}$,
- Specificity $= \frac{\text{TN}}{\text{TN+FP}}$,
- Precision $= \frac{\text{TP}}{\text{TP+FP}}$,
- F1 $= 2\frac{\text{Precision.Sensitivity}}{\text{Precision+Sensitivity}} = \frac{2\text{TP}}{2\text{TP+FP+FN}}$.

where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives and FN is the number of false negatives. Besides the five metrics defined above, we also use the area under the receiver operating characteristic curve (AUC). Accuracy evaluates the proportion of correct predictions. A higher accuracy value means a better overall prediction performance.
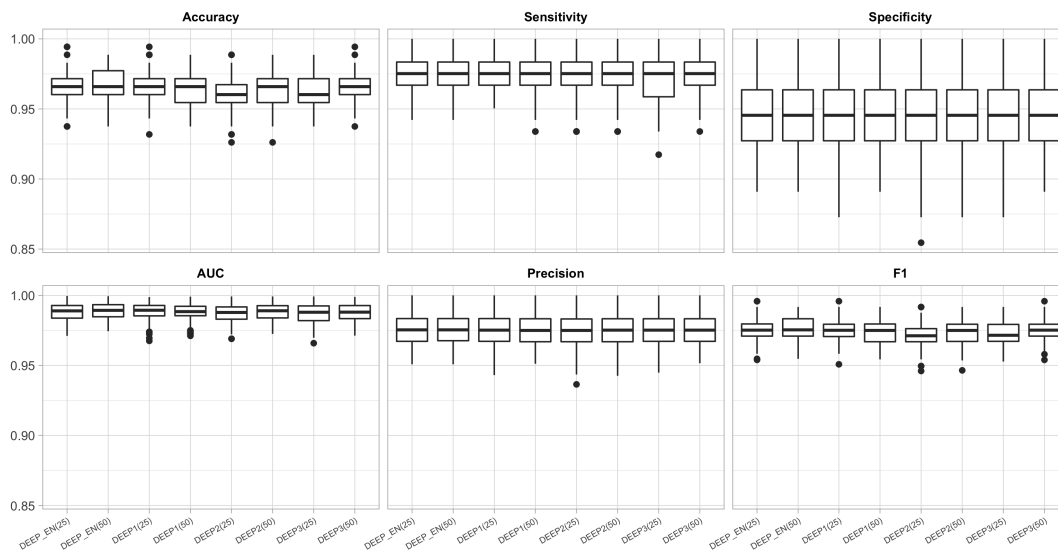
FIGURE 6. The boxplot of performance measures for all the competing methods.

Sensitivity refers to the ability to correctly detect Parkinson patients. Note that the recall is identical to sensitivity in binary classification. Specificity shows the proportion of actual negatives that are correctly predicted. Specificity refers to the ability to correctly detect normal people. Precision refers to the relevance of the predicted positives. F1 score is the harmonic mean of the precision and sensitivity.

B. RESULTS OF THE DEEP LEARNING METHODS

We first show the performance of deep learning methods for discriminating Parkinson patients. We trained all the deep learning models for 25 and 50 epochs on the training data and summarize the prediction results on the testing data. The average evaluation metrics are summarized in Table 1. The distribution of the evaluation metrics over the 100 splits are shown in Figure 6.

We observe from Table 1 and Figure 6 that the deep learning models are robust to the structure of the networks, e.g., number of neurons in the hidden layers, and the number of epochs when training. For example, the accuracy of DEEP1, DEEP2 and DEEP3 are 96.55%, 96.15% and 96.33%, respectively, when trained for 25 epochs, with minor difference to the accuracy 96.43%, 96.44% and 96.53%, respectively, when the number of epochs increases to 50. Larger networks (DEEP1) do not show higher accuracy in discriminating Parkinson patients, and they do not show signs of overfitting either. The second observation is that the ensemble network (DEEP_EN), which combines the results of DEEP1, DEEP2 and DEEP3, effectively boosts the performance of individual networks. The ensemble network, whether trained for 25 epochs or 50 epochs, achieves better performance in every measure compared to any single network.

TABLE 1. Performance measures over the testing sets for all the competing methods.

Method	Accuracy(%)		Sensitivity(%)		Specificity(%)		AUC(%)		Precision(%)		F1(%)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DEEP_EN(25)	96.64	1.11	97.42	1.35	94.91	2.73	98.82	0.64	97.70	1.21	97.55	0.81
DEEP_EN(50)	96.68	1.06	97.52	1.32	94.84	2.53	98.86	0.59	97.67	1.12	97.58	0.77
DEEP1(25)	96.55	1.11	97.40	1.30	94.67	2.75	98.81	0.67	97.59	1.22	97.49	0.81
DEEP1(50)	96.43	1.17	97.36	1.45	94.36	2.53	98.79	0.63	97.45	1.12	97.40	0.86
DEEP2(25)	96.15	1.33	97.09	1.47	94.09	3.14	98.70	0.68	97.33	1.39	97.20	0.97
DEEP2(50)	96.44	1.26	97.37	1.43	94.40	2.86	98.79	0.64	97.47	1.26	97.41	0.92
DEEP3(25)	96.33	1.16	97.16	1.53	94.51	2.69	98.70	0.72	97.51	1.18	97.32	0.86
DEEP3(50)	96.53	1.13	97.42	1.38	94.56	2.77	98.77	0.67	97.54	1.22	97.47	0.83

C. COMPARISONS WITH OTHER MACHINE LEARNING METHODS

In this section, we compare the deep learning methods with other machine learning methods in discriminating Parkinson patients. Since deep learning models are not sensitive to the number of epochs explained in the last section, we only compare the results for deep learning models trained for 50 epochs. The performance measures of the competing methods are summarized in Table 2. The distribution of accuracy, sensitivity, specificity, AUC, precision and F1 are depicted in Figures 7. Overall, all the deep learning models have a highest accuracy than other machine learning methods in discriminating healthy people with patients who have Parkinson's disease. Especially, ensemble network (DEEP_EN) achieves the highest accuracy over all methods, 96.68% on average over 100 splittings of the data. It also achieves a better balance between sensitivity and specificity, and has the highest F1 scores. The boosting methods, including BOOST_GAM, BOOST_GLM, and BOOST_TREE, follow Deep learning closely, and all have accuracy higher than 96.2%. The linear discriminate analysis method performs the best in terms of sensitivity; that is, it has the best chance to distinguish a real patient. However, it has a low sensitivity, 91.09%, and is prone to misclassify a normal people. Tree-based methods, such as BOOST_TREE and
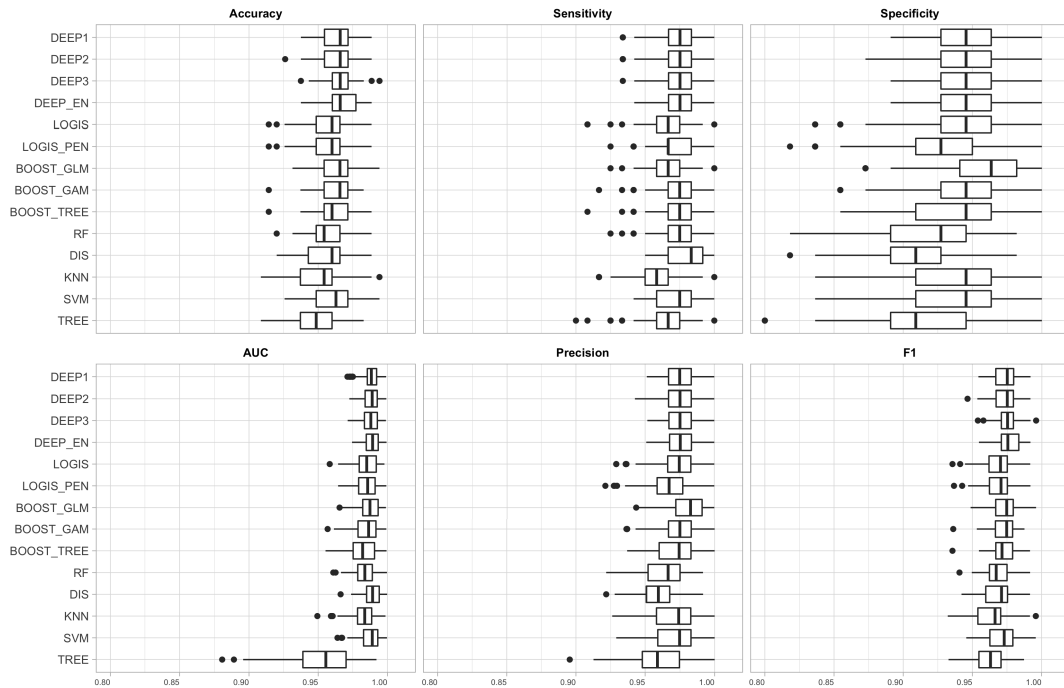
FIGURE 7. The boxplot of performance measures for all the competing methods.

TABLE 2. Performance measures over the testing sets for all the competing methods.

Method	Accuracy(%) Mean	SD	Sensitivity(%) Mean	SD	Specificity(%) Mean	SD	AUC(%) Mean	SD	Precision(%) Mean	SD	F1(%) Mean	SD
DEEP1	96.43	1.17	97.36	1.45	94.36	2.53	98.79	0.63	97.45	1.12	97.40	0.86
DEEP2	96.44	1.26	97.37	1.43	94.40	2.86	98.79	0.64	97.47	1.26	97.41	0.92
DEEP3	96.53	1.13	97.42	1.38	94.56	2.77	98.77	0.67	97.54	1.22	97.47	0.83
DEEP_EN	**96.68**	1.06	97.52	1.32	94.84	2.53	98.86	0.59	97.67	1.12	**97.58**	0.77
LOGIS	95.73	1.44	96.56	1.58	93.91	3.55	98.51	0.81	97.22	1.56	96.87	1.06
LOGIS_PEN	95.72	1.52	97.07	1.36	92.78	3.78	98.48	0.80	96.73	1.65	96.89	1.10
BOOST_GLM	96.36	1.25	96.86	1.40	95.27	2.94	98.70	0.74	97.83	1.31	97.33	0.92
BOOST_GAM	96.23	1.29	97.24	1.42	94.02	3.16	98.50	0.88	97.28	1.38	97.25	0.95
BOOST_TREE	96.25	1.23	97.43	1.52	93.65	2.93	98.23	1.01	97.12	1.28	97.27	0.91
RF	95.61	1.37	97.35	1.41	91.80	3.57	98.36	0.84	96.31	1.53	96.81	0.99
DIS	95.74	1.48	97.87	1.21	91.09	3.64	98.89	0.65	96.02	1.57	96.92	1.06
KNN	94.94	1.70	95.68	1.78	93.31	3.28	98.24	0.92	96.91	1.47	96.28	1.25
SVM	96.12	1.54	97.34	1.46	93.45	3.42	98.76	0.77	97.03	1.52	97.18	1.12
TREE	94.81	1.60	96.53	1.85	91.05	4.39	95.19	2.49	95.97	1.88	96.22	1.17

FIGURE 8. The variable importance and selection frequency calculated using the boosting model with smooth base learners.

random forest, also have sensitivity above 97.3%. Deep learning has 97.17% sensitivity. BOOST_GLM has a specificity of 95.31% on average, which is the highest among all the competing methods. The specificity measures how accurate a method identifies a true healthy people. Deep learning achieves the second largest specificity 94.84%, which is a little bit less than that of BOOST_GLM. BOOST_GLM and Deep learning achieve a good balance in sensitivity and specificity, in the sense that these methods achieve the smallest gap in sensitivity and specificity. Lastly, most of the methods, except TREE, have AUC greater than 98%.

The feature importance and selection frequency calculated using the Boosting method with smooth base learners (BOOST_GAM) is reported in Figure 8. The imaging markers of SBR's for left and right putamen (PUTMEN_L and PUTMEN_R) have the highest importance among all the features, followed by UPSIT score. The importance of other features is insignificant compared to that of the top three features. The variable importance indicates that dopaminergic imaging has a high value in discriminating Parkinson's
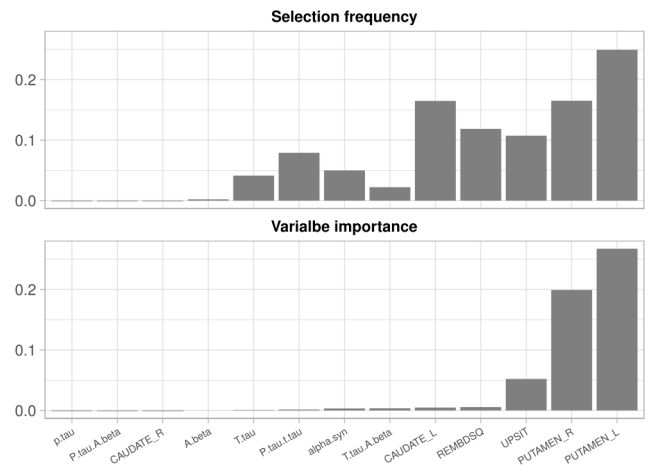
disease. The selection frequency in Figure 8 shows the frequency of features being selected as the base learners in the training process. All important features, e.g., PUTMEN_L and PUTMEN_R, are selected frequently in the training process.

Figure 9 depicts the effect of the features estimated by BOOST_GAM. The important features, e.g., PUTMEN_L, PUTMEN_R, and UPSIT scores, have a monotonic effect in discriminating Parkinson's disease. That is, people with small values of PUTMEN_L, PUTMEN_R, and UPSIT score have a higher tendency in developing Parkinson's disease. This observation is consistent with Figure 2. Several features have highly nonlinear effects, e.g., A.beta, P.tau.t.tau, and
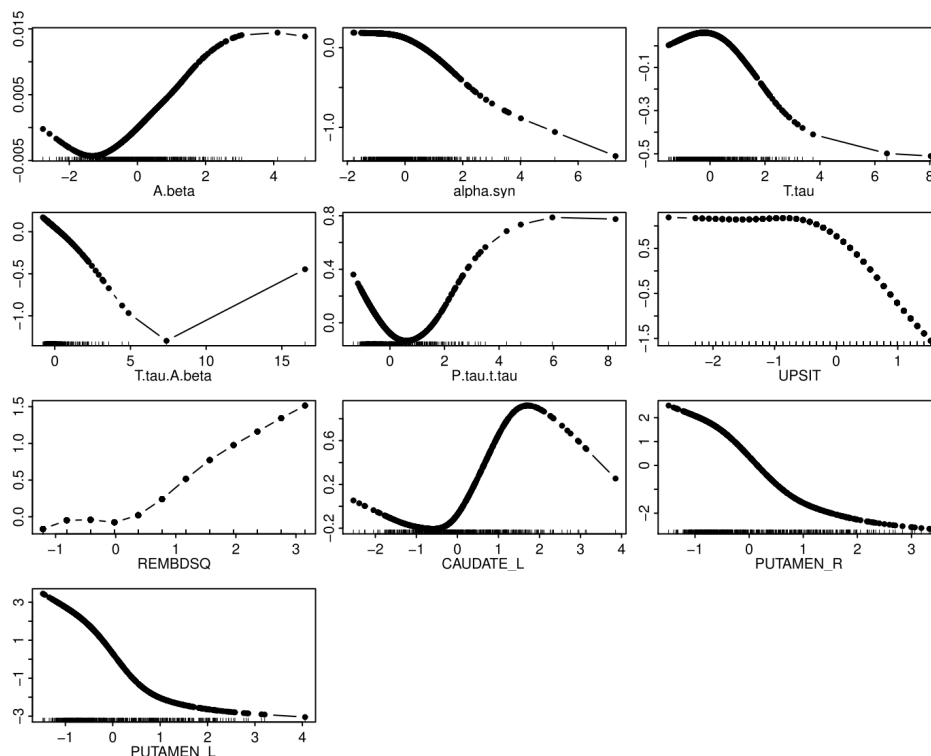
FIGURE 9. The effect of features calculated using the boosting model with smooth base learners.

CAUDATE_L. The effects of these features illustrate the complexity of discriminating patients with the Parkinson's disease. Three features are not selected by BOOST_GAM, which are p.tau, P.tau.A.beta and CAUDATE_R. These features do not contribute in the BOOST_GAM model.

D. PERFORMANCE WITH VARIABLE SELECTION

The features A.beta and P.tau.t.tau are unimportant in the variable importance analysis, and they are marginally uncorrelated with the response. In this section, we show the performance of machine learning methods when these two variables are removed from the analysis. The performance measures of the competing methods are summarized in Table 3. The distribution of accuracy, sensitivity, specificity, AUC, precision and F1 are depicted in Figures 10. The results are quite similar to the results with all features involved in the analysis (Table 2). Especially, ensemble network (DEEP_EN) achieves the highest accuracy over all methods, 96.60% on average over 100 splittings of the data. Although, this accuracy is a little bit less than that when all features are involved in the analysis.

E. COMPUTING TIME

All the methods are run on a workstation with a Intel® Xeon® CPU E5-2680 V4 and 128 gigabytes of memory. For the deep learning methods, we also use a NVIDIA® Quadro® K2200 graphical card with 4 gigabytes of memory for parallel computing. The discriminate analysis (DIS),

TABLE 3. Performance measures over the testing sets for all the competing methods when A.beta and P.tau.t.tau are removed.

Method	Accuracy(%)		Sensitivity(%)		Specificity(%)		AUC(%)		Precision(%)		F1(%)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DEEP1	96.34	1.08	97.21	1.35	94.42	2.97	98.79	0.65	97.48	1.30	97.33	0.79
DEEP2	96.35	1.16	97.18	1.35	94.51	2.80	98.83	0.69	97.51	1.23	97.34	0.85
DEEP3	96.46	1.21	97.34	1.25	94.53	3.04	98.81	0.73	97.53	1.34	97.42	0.88
DEEP_EN	**96.60**	1.14	97.38	1.25	94.89	2.84	98.91	0.65	97.69	1.25	97.53	0.83
LOGIS	96.00	1.29	96.65	1.54	94.58	3.02	98.65	0.73	97.52	1.35	97.07	0.95
LOGIS_PEN	95.71	1.25	97.13	1.24	92.60	3.35	98.55	0.72	96.65	1.46	96.88	0.90
BOOST_GLM	96.42	1.26	96.85	1.42	95.47	2.50	98.78	0.67	97.91	1.13	97.37	0.93
BOOST_GAM	96.11	1.17	97.08	1.37	94.00	2.81	98.60	0.79	97.26	1.24	97.16	0.86
BOOST_TREE	96.39	1.19	97.60	1.16	93.75	3.32	98.30	0.97	97.17	1.45	97.37	0.86
RF	95.66	1.20	97.42	1.20	91.82	3.57	98.41	0.80	96.32	1.52	96.85	0.86
DIS	95.45	1.25	97.71	1.14	90.51	3.53	98.83	0.58	95.76	1.51	96.72	0.89
KNN	95.16	1.41	95.91	1.48	93.53	3.19	97.53	1.06	97.02	1.42	96.45	1.03
SVM	95.96	1.24	97.24	1.32	93.16	3.06	98.70	0.69	96.90	1.34	97.06	0.90
TREE	94.68	1.51	96.37	1.88	91.00	4.23	94.95	2.61	95.94	1.82	96.13	1.11

k-nearest neighbor (KNN), support vector machine (SVM) and tree methods are implemented in Matlab 2020, all other methods are implemented in R. Especially, the deep learning methods are implemented using the keras package with the tensorflow as the computational backend.

Table 4 shows the summary of computing time for all the competing methods, where the interquartile range (IQR) is defined as the difference between 75th and 25th percentiles. All the deep learning methods are trained for 50 epochs. First, we can see that the logistic regression (LOGIS), penalized logistic regression (LOGIS_PEN), random forest (RF), discriminate analysis (DIS), KNN, SVM and TREE methods are very fast. On average, they use less than 1 second to process the data. The deep learning methods and boosting methods take significantly more time to train. The deep learning methods take 11 seconds to train on average. Interestingly, the training time for different networks are approximately the same, irrespective of relatively larges networks (DEEP1) or
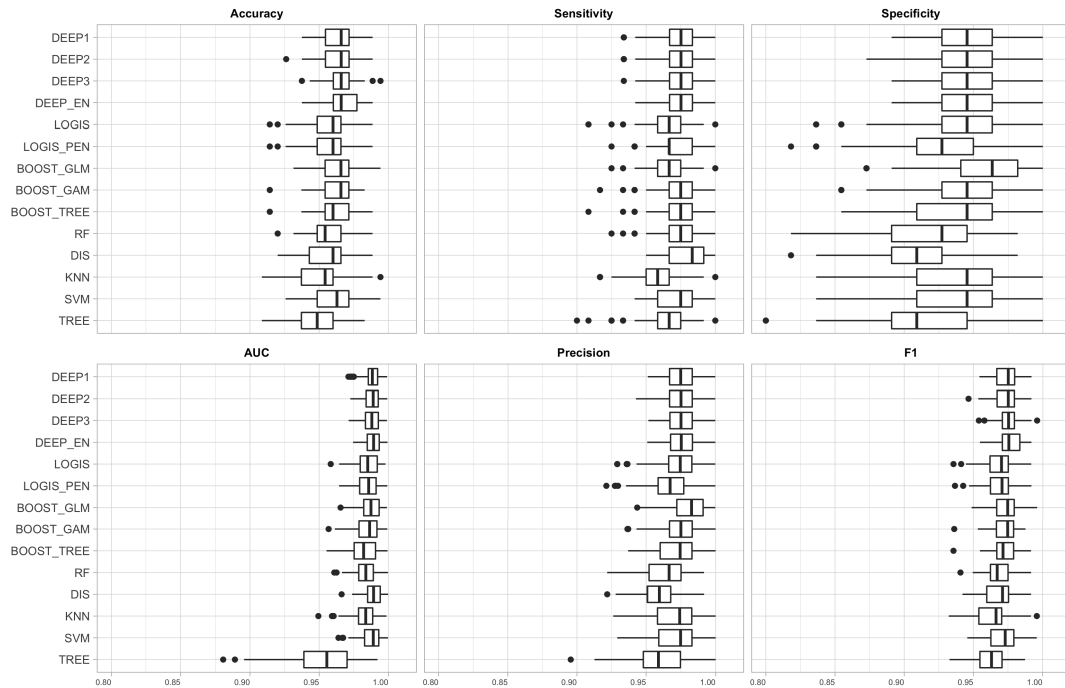
FIGURE 10. The boxplot of performance measures for all the competing methods when A.beta and P.tau.t.tau are removed.

TABLE 4. The summary of computing time (in seconds) for each competing methods.

	Mean	Median	Standard deviation	IQR
DEEP1	11.224	10.971	1.306	1.970
DEEP2	11.178	11.174	1.499	2.559
DEEP3	11.288	11.153	1.478	2.555
LOGIS	0.021	0.020	0.008	0.001
LOGIS_PEN	0.235	0.232	0.056	0.032
BOOST_GLM	6.641	6.498	0.454	0.571
BOOST_GAM	22.333	22.228	0.673	0.733
BOOST_TREE	11.026	11.072	0.214	0.197
RF	0.736	0.727	0.065	0.047
DIS	0.112	0.091	0.177	0.006
KNN	0.111	0.104	0.053	0.008
SVM	0.127	0.120	0.055	0.005
TREE	0.094	0.088	0.039	0.006

small networks(DEEP3). This is mainly because the parallel computing implemented in the tensorflow software.

IV. CONCLUSION

The early detection of PD is essential to a better understanding of the disease causes, initiate therapeutic interventions, and enable developing appropriate treatments. This study proposed a deep learning model to automatically discriminate normal individuals and patients affected by PD based on premotor features (i.e., Rapid Eye Movement (REM) sleep Behaviour Disorder (RBD) and olfactory loss). The proposed deep learning model showed good detection capacity by reaching an accuracy of 96.45%. This is mainly due to the desirable characteristics of the deep learning model in learning linear and nonlinear features from PD data without the need for hand-crafted features extraction. Results showed

that the designed deep learning offers superior detection performance compared to the twelve considered machine learning models in discriminating normal people with patients who have Parkinson's disease. The boosting methods also provide comparable performances. Even though deep learning offers superior performance compared to the machine learning models, it is hard to say that the deep learning dominates the others. This is because we designed deep learning using small PD data that are collected from 584 individuals (183 healthy and 401 early PD). However, it is expected that deep learning will demonstrate its capacity when the data is getting bigger and more complicated as day goes by. Accordingly, this outcome of this work can be viewed as a promising first step towards the application of cutting-edge research for early disease detection.

REFERENCES

[1] R. Prashanth and S. D. Roy, "Early detection of Parkinson's disease through patient questionnaire and predictive modelling," *Int. J. Med. Informat.*, vol. 119, pp. 75–87, Nov. 2018.

[2] N. Singh, V. Pillay, and Y. E. Choonara, "Advances in the treatment of Parkinson's disease," *Prog. Neurobiol.*, vol. 81, pp. 29–44, Jan. 2007.

[3] H. Gunduz, "Deep learning-based Parkinson's disease classification using vocal feature sets," *IEEE Access*, vol. 7, pp. 115540–115551, 2019.

[4] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 884–893, Apr. 2010.

[5] S. Lahmiri and A. Shmuel, "Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine," *Biomed. Signal Process. Control*, vol. 49, pp. 427–433, Mar. 2019.

[6] 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, pp. 148–158, Jan. 2019.

[7] A. Zhao, L. Qi, J. Li, J. Dong, and H. Yu, "A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data," *Neurocomputing*, vol. 315, pp. 1–8, Nov. 2018.

[8] S. Bilgin, "The impact of feature extraction for the classification of amyotrophic lateral sclerosis among neurodegenerative diseases and healthy subjects," *Biomed. Signal Process. Control*, vol. 31, pp. 288–294, Jan. 2017.

[9] L. Silveira-Moriyama, A. Petrie, D. R. Williams, A. Evans, R. Katzenschlager, E. R. Barbosa, and A. J. Lees, "The use of a color coded probability scale to interpret smell tests in suspected parkinsonism," *Movement Disorders*, vol. 24, no. 8, pp. 1144–1153, Jun. 2009.

[10] G. Valenza, S. Orsolini, S. Diciotti, L. Citi, E. P. Scilingo, M. Guerrisi, S. Danti, C. Lucetti, C. Tessa, R. Barbieri, and N. Toschi, "Assessment of spontaneous cardiovascular oscillations in Parkinson's disease," *Biomed. Signal Process. Control*, vol. 26, pp. 80–89, Apr. 2016.

[11] V. Illner, P. Sovka, and J. Rusz, "Validation of freely-available pitch detection algorithms across various noise levels in assessing speech captured by smartphone in Parkinson's disease," *Biomed. Signal Process. Control*, vol. 58, Apr. 2020, Art. no. 101831.

[12] G. Solana-Lavalle, J.-C. Galán-Hernández, and R. Rosas-Romero, "Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features," *Biocybern. Biomed. Eng.*, vol. 40, no. 1, pp. 505–516, Jan. 2020.

[13] I. El Maachi, G.-A. Bilodeau, and W. Bouachir, "Deep 1D-convnet for accurate Parkinson disease detection and severity prediction from gait," *Expert Syst. Appl.*, vol. 143, Apr. 2020, Art. no. 113075.

[14] U. Gupta, H. Bansal, and D. Joshi, "An improved sex-specific and age-dependent classification model for Parkinson's diagnosis using handwriting measurement," *Comput. Methods Programs Biomed.*, vol. 189, Jun. 2020, Art. no. 105305.

[15] A. Wagner, N. Fixler, and Y. S. Resheff, "A wavelet-based approach to monitoring Parkinson's disease symptoms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 5980–5984.

[16] T. Arroyo-Gallego, M. J. Ledesma-Carbayo, A. Sanchez-Ferro, I. Butterworth, C. S. Mendoza, M. Matarazzo, P. Montero, R. Lopez-Blanco, V. Puertas-Martin, R. Trincado, and L. Giancardo, "Detection of motor impairment in Parkinson's disease via mobile touchscreen typing," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 1994–2002, Sep. 2017.

[17] I. D. Dinov, B. Heavner, M. Tang, G. Glusman, K. Chard, M. Darcy, R. Madduri, J. Pa, C. Spino, C. Kesselman, I. Foster, E. W. Deutsch, N. D. Price, J. D. Van Horn, J. Ames, K. Clark, L. Hood, B. M. Hampstead, W. Dauer, and A. W. Toga, "Predictive big data analytics: A study of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations," *PLoS ONE*, vol. 11, no. 8, Aug. 2016, Art. no. e0157077.

[18] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Decision support framework for Parkinson's disease based on novel handwriting markers," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 3, pp. 508–516, May 2015.

[19] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease," *Artif. Intell. Med.*, vol. 67, pp. 39–46, Feb. 2016.

[20] F. Harrou, Y. Sun, A. S. Hering, and M. Madakyaru, *Statistical Process Monitoring Using Advanced Data-Driven and Deep Learning Approaches: Theory and Practical Applications*. Amsterdam, The Netherlands: Elsevier, 2020.

[21] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009.

[22] R. Das, "A comparison of multiple classification methods for diagnosis of parkinson disease," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1568–1572, Mar. 2010.

[23] A. S. Ashour, A. El-Attar, N. Dey, H. A. El-Kader, and M. M. Abd El-Naby, "Long short term memory based patient-dependent model for FOG detection in Parkinson's disease," *Pattern Recognit. Lett.*, vol. 131, pp. 23–29, Mar. 2020.

[24] S. Sharma, C. S. Moon, A. Khogali, A. Haidous, A. Chabenne, C. Ojo, M. Jelebinkov, Y. Kurdi, and M. Ebadi, "Biomarkers in Parkinson's disease (recent update)," *Neurochem. Int.*, vol. 63, no. 3, pp. 201–229, 2013.

[25] R. Prashanth, S. Dutta Roy, P. K. Mandal, and S. Ghosh, "High-accuracy detection of early Parkinson's disease through multimodal features and machine learning," *Int. J. Med. Informat.*, vol. 90, pp. 13–21, Jun. 2016.

[26] K. Marek *et al.*, "The parkinson progression marker initiative (PPMI)," *Prog. Neurobiol.*, vol. 95, no. 4, pp. 629–635, 2011.

[27] M. M. Hoehn and M. D. Yahr, "Parkinsonism: Onset, progression, and mortality," *Neurology*, vol. 17, no. 5, pp. 427–442, May 1967.

[28] K. Stiasny-Kolster, G. Mayer, S. Schäfer, J. C. Möller, M. Heinzel-Gutenbrunner, and W. H. Oertel, "The REM sleep behavior disorder screening questionnaire–A new diagnostic instrument," *Movement Disorders*, vol. 22, no. 16, pp. 2386–2393, Sep. 2007.

[29] R. L. Doty, P. Shaman, and M. Dann, "Development of the University of pennsylvania smell identification test: A standardized microencapsulated test of olfactory function," *Physiol. Behav.*, vol. 32, no. 3, pp. 489–502, Mar. 1984.

[30] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[31] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proc. 25th Int. Conf. Mach. Learn. ICML*, 2008, pp. 160–167.

[32] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[33] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, and D. Hassabis, "Improved protein structure prediction using potentials from deep learning," *Nature*, vol. 577, pp. 706–710, Jan. 2020.

[34] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.

[35] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification Regression Trees*. Boca Raton, FL, USA: CRC Press, 1984.

[36] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods," *Ann. Statist.*, vol. 26, no. 5, pp. 1651–1686, Oct. 1998.

[37] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting (With discussion and a rejoinder by the authors)," *Ann. Statist.*, vol. 28, no. 2, pp. 337–407, Apr. 2000.

[38] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[39] P. Bühlmann and T. Hothorn, "Boosting algorithms: Regularization, prediction and model fitting," *Stat. Sci.*, vol. 22, no. 4, pp. 477–505, Nov. 2007.

[40] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[41] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *J. Stat. Softw.*, vol. 33, no. 1, p. 1, 2010.

[42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[43] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

WU WANG received the B.S. degree in statistics from East China Normal University, Shanghai, China, in 2008, and the M.S. and Ph.D. degrees in probability and mathematical statistics from Fudan University, Shanghai, China, in 2011 and 2017, respectively. From 2014 to 2015, he was a Visiting Scholar with the University of Michigan, MI. Since 2017, he has been a Postdoctoral Fellow with the Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology. Since 2018, he has been a member of the American Statistical Association (ASA) and the International Chinese Statistical Association (ICSA).

JUNHO LEE received the B.A. and M.A. degrees in statistics from Yonsei University, Seoul, South Korea, in 2004 and 2006, respectively, and the Ph.D. degree in statistics from the University of Wisconsin, Madison, WI, in 2017.

From 2005 to 2008, he was a Researcher with Korea Credit Bureau, Seoul. He was also an Academic Instructor with Yonsei University between 2008 and 2009. Since 2017, he has been a Postdoctoral Fellow with the Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology. Since 2015, he has also been a member of the American Statistical Association (ASA) and the Korean International Statistical Society (KISS).

FOUZI HARROU (Member, IEEE) received the M.Sc. degree in telecommunications and networking from the University of Paris VI, France, and the Ph.D. degree in systems optimization and security from the University of Technology of Troyes (UTT), France. He was an Assistant Professor with UTT for a year and with the Institute of Automotive and Transport Engineering, Nevers, France, for a year. He was also a Postdoctoral Research Associate with the Systems Modeling and Dependability Laboratory, UTT, for a year. He was a Research Scientist with the Chemical Engineering Department, Texas A&M University at Qatar, Doha, Qatar, for three years. He is currently a Research Scientist with the Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology. He is the author of more than 130 refereed journals and conference publications and book chapters. He is also a coauthor of the book *Statistical Process Monitoring Using Advanced Data-Driven and Deep Learning Approaches: Theory and Practical Applications* (Elsevier, 2020). His current research interests include statistical decision theory and its applications, fault detection and diagnosis, and deep learning.

YING SUN received the Ph.D. degree in statistics from Texas A&M University, in 2011. She held a two-year postdoctoral research position at the Statistical and Applied Mathematical Sciences Institute, University of Chicago. She was an Assistant Professor with The Ohio State University for a year before joining the King Abdullah University of Science and Technology (KAUST), in 2014. At KAUST, she established and leads the Environmental Statistics Research Group, which works on developing statistical models and methods for complex data to address important environmental problems. She has made original contributions to environmental statistics, in particular in the areas of spatiotemporal statistics, functional data analysis, visualization, and computational statistics, with an exceptionally broad array of applications. She received two prestigious awards: the Early Investigator Award in Environmental Statistics presented by the American Statistical Association and the Abdel El-Shaarawi Young Research Award from the International Environmetrics Society.

• • •