IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 9, 2018, accepted June 19, 2018, date of publication July 6, 2018, date of current version July 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2851382

Deep Multi-Layer Perceptron Classifier for Behavior Analysis to Estimate Parkinson's Disease Severity Using Smartphones

SHAOHUA WAN[1], (Member, IEEE), YAN LIANG[1], YIN ZHANG[1], (Senior Member, IEEE), AND MOHSEN GUIZANI[2], (Fellow, IEEE)
[1]School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China
[2]Electrical and Computer Engineering Department, University of Idaho, Moscow, ID 83844, USA

Corresponding author: Yin Zhang (yin.zhang.cn@ieee.org)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702553 and in part by MOE (Ministry of Education in China) Project of Humanities and Social Sciences under Grant 17YJCZH252.

ABSTRACT Although the preclinical detection of Parkinson's disease (PD) has been explored, a practical, inexpensive, and overall screening diagnosis has yet to be made available. However, due to the large variability and complexity in progress of PD and the difficulties in gathering a single time-point measurement of a single sign, the goal of precision treatment and assessment severity would be impossible to achieve. Hence, the repeated monitoring and tracking of individuals during their daily living activities at different times would also be of great importance for treating this chronic disease. We propose a deep multi-layer perceptron (DMLP) classifier for behavior analysis to estimate the progression of PD using smartphones. This paper aims to identify severity in PD patients' actions by analyzing their speech and movement patterns, as measured with a smartphone accelerometer in their pocket at different times of the day. Popular machine learning classification algorithms, such as logistic regression, random forests, k-nearest neighbors, M5P, and DMLP, are applied on one dataset from the University of California Irvine and another dataset collected by the authors to classify each patient as being Parkinson positive or negative. We further measure the success of each method for their ability to correctly classify the patients into one of these categories. Of the experimental models, it is demonstrated that DMLP performs the best in both datasets.

INDEX TERMS Parkinson's disease, behavior analysis, DMLP, classification.

I. INTRODUCTION

Parkinson's disease (PD) is a degenerative disorder of the nervous system and thus affects behavioral patterns. The characteristic clinical picture and motor symptoms of the disease include resting tremors, stiffness, slow shuffling gait, slowness of voluntary movement and impaired speech. There is no fully effective cure for PD, with most treatments targeted at mitigating symptoms. In the early phases of PD, the face may reveal little or no expression, or the arms may not shake while walking. The physical condition deteriorates increasingly, sometimes beginning with a barely obvious tremor in the fingers of just one side of the body. The most apparent sign of PD is tremors, but lesions also generally stimulate the aggravating effects of stiffness, swinging and slowness of movement. Speech may also become soft or slurred. Speakers with PD are characterized by speech impairments such as decreased loudness, restricted pitch variability, inaccurate articulation, abnormal speech rates and intermittent ratios. The speech symptoms of PD become worse as the situation progresses, as evidenced by the participants with or without PD who are recognized in our speech experiment.

Although PD patients have obvious motor symptoms, they are not prominent in the preliminary phases of the disease's quiescence. The condition is aggravated after slow movement, gait disorders and other clinical manifestations; therefore, patients with motor signs are an important indicator of clinical diagnosis. The early detection of preclinical patients exerts a significant function in the subsequent prevention, treatment and improvement of patients' quality of life. Although the precise cause of PD is generally not yet clear, genetics, age and living environment may have an positive effect on the development of PD. According to the clinic experience, Parkinson's patients have no obvious symptoms in the early stages, where their situation is deteriorating slowly. As a result, few patients seek treatment during this period. With the progress of the disease, the patient's aforementioned motor symptoms become increasingly obvious. PD poses a severe challenge today, as its prevalence accounts for nearly 160 cases per 100,000 population, and its incidence is approximately 20 cases per 100,000 population. Therefore, the patient's symptoms are an important basis for the clinical diagnosis of PD. Therefore, how to effectively use wearable sensors to detect the movement of the elderly and how to identify and assess the severity of PD according to the appearance of motor symptoms are worthwhile and significant research questions. Early detection allows early treatment, which is crucial because early identification and treatment can effectively prevent and delay impairment.

Although PD cannot be cured, medications may markedly improve the symptoms. Regularly monitoring a patient's status contributes greatly to better treatment completion. The Unified Parkinson's Disease Rating Scale (UPDRS) is worldwide adopted scoring system to clinically assess and track the processing of the disease. Approximately 90% of PD patients manifest speech impairment [1]. UPDRS collects multiple aspects of PD, including mentation, manner, emotional problems (depression and anxiety), daily activities, comprehensive motor examination and complications of chronic therapy. For the repeated monitoring of a patient's UPDRS score, telemonitoring has been proposed. The basic idea is to evaluate the UPDRS score based on previous UPDRS scores and biomedical video recordings, which can potentially be obtained as the patient makes telephone or Skype calls on his/her smartphone. The high-resolution activity data collected from smartphones can aid in the quantitative analysis of PD symptoms with non-invasive, continuous and uninterrupted monitoring during the participant's daily life, which ultimately may lead to a thorough understanding of smart health and mobile health to promote human health.

Recently, researchers have begun to explore the use of applying smartphones with smart learning to detect the severity of PD before the symptoms deteriorate further, as well as in other healthcare context [2]. PD is a neuron degenerative disorder of the central nervous system, which is the main cause of partial or full impairment in postural reflexes, speech, movement, meditation, and other vital signs. Smartphones are the perfect electronic devices for monitoring an individual's health because they are carried around everywhere we go, althogh there are security issues that need to be considered [3]–[7]. Smartphones improve users' experiences by equipping them with a wide range of sensors. A normal smartphone sensor that measures movement, known as an accelerometer, can differentiate between patients with PD and healthy people by measuring their walking patterns. Measuring the fluctuations in PD symptoms over the course of a day with a smartphone can reduce the expenses and inconvenience from interacting with physicians. The ability of smartphone sensors to detect these subtle movement problems has attracted many researchers. PD, a brain disease, results in much less voluntary movement, including speech impairments, slowness of movement, trembling of the hands and legs, muscles resisting movement, and loss of balance. Currently, only when the obvious movement problems of the disease appear during the late stages of the disease can PD be identified. If the disease is detected in the early stages, a cure or treatment that may slow the progression of disease could be found. It is interesting that people who are at risk of developing PD but do not show any symptoms may show very subtle movement problems in their walking patterns, which further aggravates the difficulties in identifying PD. First, since nearly 90% of people with PD show some degree of vocal lesion, we employ UCI Parkinson's Telemonitoring Dataset (Speech Dataset) to detect whether a patient has Parkinson in our paper. Second, we collect our own smartphone data and manually label walking, sitting, and standing actions for accelerometers and gyroscopes as the other data sets. Our approach is proposed to apply Deep Multi-Layer Perceptron (DMLP) to smartphone observations, supplemented with inertial accelerometers, gyroscopes, and microphones, to recognize the severity of PD without interventions from physicians. Profiting from a large scope of enrollment and continuous monitoring on many human activities, these measurements may be greatly helpful in building the baseline variability of real-world activity observations observed via smartphones and may provide a quantification of PD symptoms before these are considered in clinical decision-making. Without requiring physical interactions with physicians, this detection could thoroughly assess the aspects of PD change by using many scope measures from performed activities and passive inference, in contrast to a single signal from mechanical and electronic devices. The contributions of this paper are summarized as follows:

- The behavioral data is collected with smartphones and manually labeled as walking, sitting, and standing actions for accelerometers and gyroscopes. Then, the data is Butterworth low-pass filtered, discretized, denoised and normalized. Finally, the feature vector composed of feature values is used as training data.
- The proposed DMLP classifier for behavior analysis is compared to the other classification methods on the data collected via smartphones and UCI speech data set. In addition, the trained DMLP method is employed to identify actions in the unlabeled data.
- We analyze the frequency and distribution of PD symptoms and assess the severity and possibility of PD.

The remainder of this article is organized as follows. Section II presents related works on the classification methods used to analyze Parkinson datasets from wearable sensors. Section III introduces UCI data sets and individuals' activity data from smartphones. Section IV provides overviews of the k-Nearest Neighbor (KNN), Random Forests (RF), Linear Regression, M5P regression tree, and DMLP classification methods. We theoretically analyze their performance. Section V presents the experimental settings and evaluation metrics. Finally, we conclude this article in Section VI

II. RELATED WORK

In the literature, studies that focus mainly on speech measurement for general voice lesions can be found in [8]–[11], particularly considering PD detection. Some studies use a regression approach to detect the level of PD by utilizing the UPDRS measurements, while other studies approach the problem as a classification problem to detect whether a patient has Parkinson [12]–[16]. Little *et al.* [17] tried to diagnose the lesion by measuring the dysphonia that arose from PD with a dataset containing sustained vowel "a" phonations of 31 attributes. They presented a SVM with RBF that had better performance (91.4%). Furthermore, Sakar *et al.* [18] designed a PD diagnostic device and achieved a 92.75% classification accuracy by combining mutual information into SVM. Additionally, Tsanas *et al.* [19] used linear and nonlinear regression algorithms to estimate the progression of the disease (UPDRS level) on a set of 6000 samples of 42 PWP. Tsanas *et al.* [20] also posed a classification problem with the extended version of the dataset [21] and achieved the best results when using a non-linear SVM, with 97% accuracy, compared with LASSO and Random Forests. The previous research demonstrates the successful use of SVMs in the classification analysis of the Parkinson speech dataset. Comparing the PD detection capabilities of an artificial neural network (ANN), DMneural, with regression with decision trees, Das [22] experimentally demonstrated that the ANN performs the best and produces more correct classification results, with an overall accuracy of 92.9%. Integrating the fuzzy k-nearest neighbor (FKNN) method and a principle component analysis (PCA) (PCA-FKNN), Emary *et al.* [23], Zhao and Yin [24], Hariharan *et al.* [25], and Jiang *et al.* [26] presented the advantages of both FKNN and PCA and developed a PD detection device that produces a classification accuracy of 96.07%. Zuo *et al.* [27], Meng and Shao [28], Cai *et al.* [29], and Li and Ren [30] achieved a mean accuracy of 97.47% when using an efficient PD aided detection tool that improved on the performance of a FKNN based on PSO.

Currently, PD is diagnosed by having the patient meet face-to-face with medical experts, but recent studies have also utilized motion sensors (dedicated or on smartphones) to identify features of the disease, typically by securing the sensors to areas such as the waist and legs [31]–[45]. This creative research, as innovated in the USA, attempts to track PD patients and control interventions in the trial using a smartphone app called MPower, which could provide information about the symptoms, as obtained through finger tapping, voice recording, balance and walking. Critically, the data for this project is publicly available to researchers who have completed the required accreditation process. Our research is to predict PD severity with classification algorithms on smartphone sensor data in pockets from the UCI repository and from our collected actions.

III. DATASET PREPROCESSING

A. UCI DATASET

A series of biomedical voice recordings [46] from 42 people who have early-stage PD and were engaged in a six-month test of a telemonitoring tool for remote signs and symptoms process monitoring form this publicly available dataset. In this period, the voices of these patients are automatically recorded in their home. The dataset details the columns included, such as subject number, subject age, subject gender, time interval from preliminary enrollment date, motor UPDRS, total UPDRS, and 16 biomedical speech measures. Each row matches one of the 5875 voice recording captured from those patients. The major purpose of the recordings is to forecast the motor and total UPDRS scores ("motor_UPDRS" and "total_UPDRS") from the 16 voice recordings. The measurement is in ASCII CSV format. The rows of the CSV file include a living example in accordance with one voice recording. Each patient has approximately 200 recordings, and the first column of the measurements provides the subject number of the patient.

The training data captured in the context of this study is from 20 PD patients, including 6 females and 14 males, and 20 healthy individuals (50% of each gender). Twenty-three voice samples are collected from each subject, and these voice measurements are characteristic of continuous vowel, word, number, idiom and short sentence. When capturing these recordings, the 28 PD patients are only required to speak two persistent vowels 'a' and 'o' three times each. This process gives a total of 168 samples. We extract the 23 feature parameters from the voice sets of these recordings, which can be viewed as an independent trial example to corroborate the results derived from training example. A list of the features used in this experiment is given in Table 1.

The other human activity dataset is created from the daily movement activities of 50 volunteers while wearing a waist-mounted smartphone with installed inner sensors. The task is to designate the performed activities into one of three classifications. A class of 50 volunteers between the ages of 19 and 45 years participates in this experiment. Each person performs 3 activities (WALKING, SITTING, STANDING) while carrying a smartphone on the waist. Since the frequency of human activity is between 0 and 15 Hz and the frequency of a tremor is between 4 and 6 Hz, 3-axial linear acceleration signals are collected with the phone's built-in accelerometer and gyroscope at an invariable rate of 30 Hz, which means that the captured signal not only has integrity but also has no redundancy. The video recordings are marked manually and randomly divided into two sets, where 70% of the available data is allocated to the training example and the remaining 30% are used for the test example.

B. THE DATASET COLLECTED FROM SMARTPHONES' SEGMENTATION AND PREPROCESSING

In addition to the acceleration signal generated by human motion, there are a variety of interference signals, such as the

TABLE 1. PD features extracted from voice samples.

Group	Features
Fundamental Frequency	**Jitter(%)** **Jitter(ABS)** **Jitter(RAP)** **Jitter(PPQ5)** **Jitter(DDP)**
Avg/Max/Min	**Fo(Hz)** **Fhi(Hz)** **Flo(Hz)**
Variation Amplitude	**Shimmer** **Shimmer(dB)** **Shimmer(APQ3)** **Shimmer(APQ5)** **Shimmer(APQ11)** **Shimmer(DDA)**
Nonlinear Measures	**Spread1** **Spread2** **PPE**
Status	**1 - Parkinson's** **0 - healthy**
Dynamic Measures	**RPDE** **D2**
Signal Exponent	**DFA**
Harmonicas Parameters	**NHR** **HNR**

inherent noise of the sensor, the vibration caused by changes in external conditions, and the sensor not being completely fixed, which causes friction due to acceleration and other interference signals. The noise can be reduced by selecting the sensor with greater accuracy and stability when the sensor component is worn. For the noise that has been generated, we process the original acceleration signal by using Butterworth low-pass filtering to denoise and interfere to obtain a relatively smooth acceleration curve. There are various types of filters to choose from. Median filtering for sharp noise and Butterworth low-pass filtering have the flattest frequency response curves, so these methods are used in combination to preprocess and denoise the raw data. The accelerated velocity signal of sensors, which intertwines gravitational signals with body motion components, is used in Butterworth low-pass filtering to split the signals into two parts: gravity and acceleration. With the assumption that the gravitational force contains only low-frequency constituents, filtering with a 0.25 Hz cutoff frequency might be denoised.

So far, a large number of discrete data sequences that are continuous in time have been captured. The huge amount of data contains unrelated details and is not suitable for direct processing. Therefore, we divide these large-scale data sequences into a large number of data segments that are of the same length in time as the basic processing units for subsequent analysis and calculation. The data in each time period is analyzed, and the useful feature values are extracted to identify the training model or used as input for the trained recognition model to determine whether there is an abnormality in this period.

The body's movement is a complicated process in time. In daily activities, an individual's movement consists of a large number of random and regular movements. According to related research on human kinematics, most random actions, such as turning, bouncing and standing, are completed in approximately 2 seconds. The regular cycles of walking, running and going up and down are also approximately 2 seconds. In the meantime, there are intermittent and random seizures in patients with early PD. There is no specific pattern to follow. The frequency of tremors is approximately 4 to 6 times a second, and the duration of symptoms is approximately 2 to 10 seconds. Therefore, we select the time window ΔT to be 2 seconds, and the data are collected every 2 seconds to form a data sequence as the basic unit of information processing. From each time window, a feature vector formed by computing variables of the time and frequency range is naturally created.

After filtering the data for noise reduction and segmentation, the clean and reliable data is employed. Next, we select the feature value and form the feature vector as the input term of the machine-term classifier. Normalization is also called standardization. That is, by scaling the data, all of the data to be processed is mapped into a range of numbers. This is done to improve the convergence of the data and make it easier to process the data. With the normalization method, the problems of dimension and order of magnitude can be eliminated. A change in large-magnitude data will not completely affect the final result, and a change in small-magnitude data will not be completely neglected. When the data is uniform and of the same order of magnitude, it is more suitable for dimensionless learning. In this paper, mean normalization is used to normalize the data according to the mean and divide the variance, so it is suitable for a more scattered data set of data and may be formulated as follows:

$$\hat{X} = \frac{X - \bar{X}}{\sigma} \tag{1}$$

In the field of machine learning, the range of feature values can be confined in the same area through normalization, which accelerates the convergence speed and improves the training speed when training the classifier. The pipeline of the deep MLP classifier for behavior analysis to assess PD severity can be seen in Figure 4.

IV. CLASSIFICATION ALGORITHMS

Two kinds of data sets are normally used for a classification algorithm: the training set and the testing set. Classification

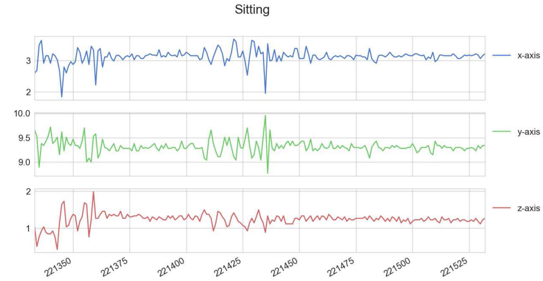


FIGURE 1. Sitting readings from smartphones.

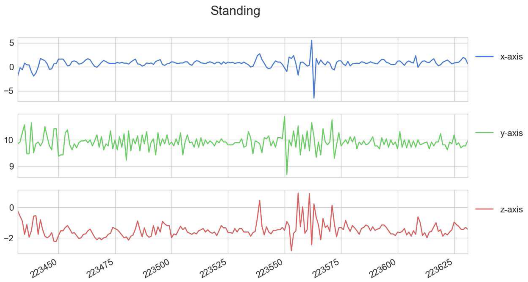


FIGURE 2. Standing readings from smartphones.

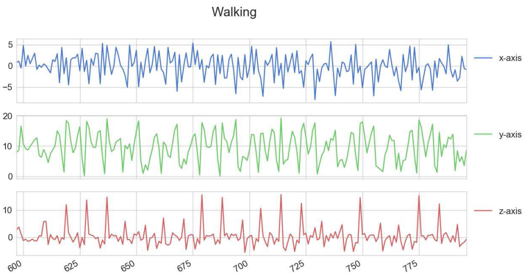


FIGURE 3. Walking readings from smartphones.

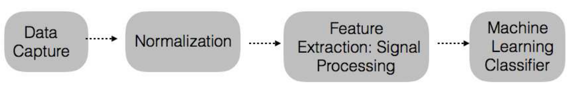


FIGURE 4. Classification pipeline.

algorithms find a model that can be used for a training set according to the testing set. That is, the training set is utilized to build a model, and the testing set is used to evaluate the model. The model can divide the data by class attributes such that each class has the same types of class attributes for the data. An overview of the algorithms used to classify the Parkinson dataset is discussed here.

A. K-NEAREST NEIGHBOR REGRESSION

The KNN classifiers find the k closest nodes to a new node in the training set and identify its classification based on the primary classifications of these k neighbors. There are three integral factors: the training set, the distance or similarity of the k neighbors to the new node, and the value of k. For a given data record, the first step for a KNN classifier is to calculate all the distances between any two nodes in the training set and then to rank and select the k closest from them, so the target record can be classified by its k neighbors based on the primary neighbors that have a common classification. The distance is often Euclidean distance, but cosine similarity may be better for text data sets. Choosing an appropriate value of k is vital and difficult and is always determined by cross-checking (based on $k = 1$). KNN is a lazy algorithm with a simple model that is not inclined to refine itself, and it is time-consuming to compute all distances between two nearby nodes in the training set.

B. RANDOM FORESTS

RF is an ensemble learning method that uses multiple decision trees that are randomly built to solve classification or regression problems. Every decision tree in the random forest considers a sub-set by taking random samples from the original dataset, and every sub-tree is built of randomly selected features. Each decision tree learns by training to obtain results independently and gets the final production by taking the vote. The majority of the decision trees conduct kinds of situations that do not perform properly but can be used as a basis for other trees to work better. Compared to a single decision tree model, random forests show better performance in data sets for addressing the shortcoming of overfitting. They can handle highly multi-dimensional, discrete or continuous data without feature selection, and they have good noise immunity.

C. LINEAR REGRESSION

Linear regression is a kind of regression analysis that finds a quantitative linear connection between variables in the training data using least squares regression. The values' dependent variable depends on mapping values of the independent variable and the regression function fitted from the training set. The least square and gradient descent are two common methods used in linear regressions to estimate the fitting of the regression by computing the loss between the estimated value and the ground truth. By contrast, logistic regression is a kind of generalized linear regression model that is nonlinear and has more than one independent variable. It can handle nonlinear relations and binary classification problems, which a linear regression makes mistakes with.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i| \tag{2}$$

D. M5P REGRESSION TREE

The M5P Regression Tree is an accurate model with first-order segmented multiple linear regressions. It divides the sample space of the training set on the principle of the standard deviation reduction (SDR) that differs from the information gain in the decision tree model. The splitting of the data sample stops when the sample size is smaller than a certain value or when the standard deviation reduction of the samples of the node meets a threshold value. Then, a corresponding regression model is built for each partition of samples based on data sample features while combining the pruning and smoothing of trees. The M5P regression tree utilizes the pruning method from leaf nodes to the root to avoid over-learning for the tree. It connects the branches and nodes with the regression tree according to regression linear equations and finds optimal pruning positions where the attributes that maximize the expected error reduction are selected for separating at that node. The smoothing of the trees fits the linear regression models of every couple adjacent leaf nodes by multivariate linear fitting equations, which corrects the discontinuity between neighboring leaves. For an experiment, the expected error reduction is calculated with the following formula:

$$\Delta error = stdev(S) - \sum_i (\frac{|S_i|}{|S|} stdev(S)) \tag{3}$$

where S means the set of samples traversed to the node, stdev(S) denotes the standard deviation, and S_i is the subset of S that results from splitting the node in accordance with the selected attribute. The progression of splitting into new nodes is terminated when there are only a few instances (four or fewer) to process further or when the variation in the output values of the instances that pass the node is very slight. Once the tree has been created, a linear model is designed at each node that is a regression equation.

E. DMLP CLASSIFICATION ALGORITHM

MLP is a multiple feedforward artificial neural network that maps input vectors to output vectors. The MLP can be defined as a directed graph with multiple node layers, where the input layer is on the bottom, the output layer is on the top, and the others in the middle are the hidden layers. Every node in the upper level has connections with all the nodes in the lower level; this is called a fully connected network. One or more hidden layers are allowed. Each node represents a neuron (or processing unit) with a nonlinear activation function except for input layer nodes. A supervised learning methodology referred to as a backpropagation algorithm (BP) is often utilized to train MLP. MLP is the extension of a single-layer perceptron that corrects the weakness that single-layer perceptrons cannot solve nonlinear data. It can learn non-linearly separable decisions, in contrast to single-layer perceptrons. Figure 5 depicts three-layer MLP:

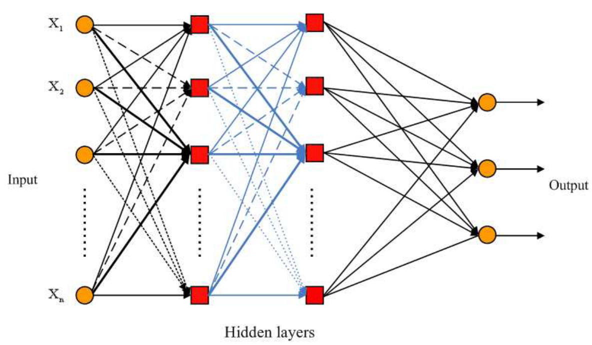


FIGURE 5. Three-layer MLP.

DMLP [47] refers to the structure described from this point onwards, in which five or ten hidden layers are adopted; by contrast, the maximum number of layers traditionally employed by simple MLP is two. Furthermore, the sigmoid and tanh activation functions are traditionally employed in MLP because they provide good performance in smaller and medium-sized networks. For the DMLP model developed for cloudUPDRS, ReLU or softplus are preferred instead because the latter activation functions can address the vanishing gradient problem that affects deep networks. The function enables them to obtain sparse representations by hard-limiting the input of negative hidden nodes to zero.

Connections that cross over a couple of layers are named shortcuts. Most DMLPs have a connection architecture, where all nodes of one layer are connected to the whole nodes of the next layer without shortcuts. The nodes of layers are listed as follows:

- *Succ(i)* is the set of all nodes j when the connection $i \rightarrow j$ exits
- *Pre(i)* is the set of all nodes j when the connection $j \rightarrow i$ exits

There are assignments of nonnegative real numbers (weights) to all the connections. w_{ji} denotes the weight of the connection $i \rightarrow j$. All hidden and output nodes have a bias weight and some variable net_i (''network input''). The bias node is connected to other nodes with trainable weights. w_{i0} represents the bias weight of neuron i. All nodes have some variable a_i (''activation''/''output''). The MLP algorithm can be described as follows:

Algorithm 1 Multi-Layer Perceptrons

Require:
pattern $\vec{x}$, DMLP, count all nodes in topological sequence;
Ensure:
compute output of DMLP;
for all initial nodes i **do**
  compute activation of input nodes: $a_i \leftarrow x_i$;
for all the hidden and output nodes i in topological sequence **do**
  $net_i \leftarrow w_{i0} + \text{âĹ}'_{\text{jâĹĹ}Pred(i)} w_{ij} a_i j$;
  $a_i \leftarrow f\text{ı}ognet_i$;
for all output nodes i **do**
  gather a_i in output vector y;
return $\vec{y}$

V. EXPERIMENTAL ANALYSIS AND PERFORMANCE EVALUATION

A. PERFORMANCE METRICS

The metrics that are widely known to be used to assess the performance of classification algorithms are accuracy, sensitivity, specificity, and MCC. It is acknowledged that accuracy is the ratio of accurately classified examples to the whole examples and is a good measure for evaluating a model's performance.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{7}$$

where TP means the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. Sensitivity and specificity are statistical measures of a test's ability to accurately classify a person as with or without PD, respectively. The Matthews Correlation Coefficient (MCC) has a range between the predicted and observed binary classifications and takes a value of $+1$ when all of the prediction values are correct, -1 for a completely incorrect binary classifier, and 0 when the classification is worse than a random estimation.

B. CROSS-VALIDATION (CV)

In cross-validation, we leave aside a subset of the data provided. Instead, we use the remaining data, i.e., the training set, in the algorithm to build a model. The test set, which is the part of the dataset that was not applied in training, measures the accuracy with which the model classifies the test set instances, aiming to assess the model's capability. In K-fold cross-validation (KFCV), we arbitrarily split the dataset into K subsets of equal size, use one subset as a test set, and apply the remaining $K - 1$ subsets for training. To ensure that each subset is used as a test set exactly once, the cross-validation must be performed n times. The detail of KFCV is that the data set is randomly allocated into K approximately equally sized subsets. For each part k, we leave out part k, fit the model to the other $K - 1$ sets (combined), and subsequently compute predictions for the remaining kth part. For each part $k = 1, 2, \ldots K$, the same is done successively, and ultimately, the results are combined. The advantage of KFCV is that all samples in the dataset are eventually used both to train and to test. Setting $K = n$ yields n-fold or leave-one out cross-validation (LOOCV). LOOCV is the degenerate case of KFCV, where N is chosen as the total number of examples.

C. RESULTS AND DISCUSSIONS

As predictors, we use the time since recruitment and the four most relevant biomedical voice features according to [1]: HNR, RPDE, DFA and PPE. The Personalized Telemonitoring scenario is an adaptive system for estimating the UPDRS score. In temporal train-test splits, the first 2/3 of the data is used as training data, and the remaining 1/3 of the data is used as test data. A separate model is built for each patient using only those instances corresponding to that patient, resulting in personalized models. We average the estimations belonging to the same-day final estimation of the UPDRS score for a particular day. We estimate both the motor and total UPDRS scores. The primary evaluation metric is the mean absolute error (MAE). We also use the normalized mean absolute error NMAE and root mean squared error (RMSE) and observe similar trends. The accelerometer data is recorded by phone, at a 50 Hz sampling frequency, for an individual.

As shown in Figure 6, we find that the Net-5 and Net-10 methods predict visibility with almost the same absolute error, with M5P being slightly higher. Thus, we use DMLP to recognize visibility over different degrees precisely, and the classifications are independent. Our approach in adopting the DMLP classifier in Figure 6 produces the best performance across all speech features. Note in particular that the magnitude of the standard deviation is very small, implying that the initialization bias has been avoided. When the difference between the errors from the estimation and the MAE of DMLP is larger than the standard deviations, as computed from the error values observed during CV, the smartphone is breaking down. All the performance metrics are given for validation sets in Table 2. With summarized feature vectors, the performance improves significantly. All methods present the positive MCC. The experimental results for the smartphone dataset are computed with the same classifiers and depicted in Table 3. Further, in this case, the DMLP approach also surpasses the other classifiers across all features, often by a significant margin.

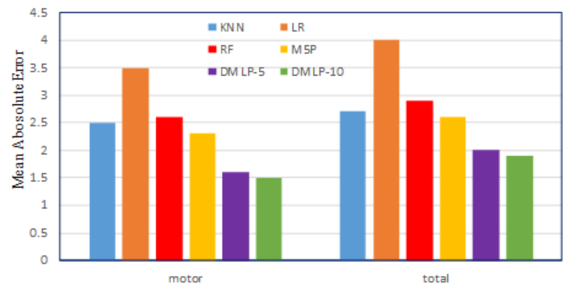


FIGURE 6. The MAE of predicting PD severity using classification algorithms.

TABLE 2. Results for validation set using LOOCV.

	Accuracy	Specificity	Sensitivity	MCC
KNN	72.5%	75%	70%	0.45
Random Forests	65%	65%	60%	0.25
Linear Regression	77.5%	70%	85%	0.55
M5P	75%	70%	72.5%	0.45
DMLP-5	76%	71%	77.5%	0.35
DMLP-10	80%	73%	72.5%	0.40

TABLE 3. Classification for 3 actions: sit, stand, walk.

	Precision Average (Cross-Validation 10 folds)	Precision Average (Percentage Split 30)
KNN	89.2	90.1
Random Forests	90.9	89.5
Linear Regression	85.4	86.8
M5P	92.3	93.1
DMLP-5	94.3	95.1
DMLP-10	97.9	97.3

Of the examined models, artificial neural networks seem to perform best. While an automated estimation of the UPDRS score with personalized apps running on smartphones or tablets is a promising research direction, further research is required to achieve a sufficient accuracy for real-world applications, especially if only a small amount of training data is available for each patient or if the voice data is captured under daily-life conditions (instead of well-controlled laboratory settings). Additional data sources (based on the user's interaction with the smartphone or tablet, such as her typing patterns or performance in logical games) could improve the estimation of the UPDRS score.

Classifications for 3 actions (sit, stand, walk) are shown in Table 3. The experimental results shows a maximum of 97.9% accuracy in action classification with 3 actions with the classification algorithm. Because we could not collect our own PD data with ground truth, the model created with control subjects was unsuitable for accurate classification. The main limitations of this experiment were the small sample size and the lack of real PD patients available to create the model. Some of the smartphone data also lacked long, continuous data spanning from morning to evening, which makes meaningful comparisons difficult.

VI. CONCLUDED REMARKS AND FUTURE WORK

In smart healthcare, mobile smartphone presents a promising opportunity for the provision of practical and cost effective diagnosis at population scale as the expert clinical treatment fail to collect the persistent activities and signs. In the chronic and progressive disease, monitoring the activities of daily life is of great importance. Yet, to completely reach their potential such apps should provide a seamless user experience. In this paper, the classifiers, including KNN, RF, Linear Regression, M5P, and DMLP are applied to accelerometer data continuously captured from smartphones. In our experiments both datasets perform reliably and show promising results. Of the experimental models, it is demonstrated that DMLP performs the best in both datasets.

Since recording our own PD patient data with ground truth was not achieved, it would be useful to do so and have a more direct comparison with the knowledge of actions. With more data, even if they were not recorded by this study, the classification would also be more accurate. Furthermore, longer-scale data spanning months would be useful for observing not just small cycles in actions but also the progression of symptoms. In addition to more data, future endeavors could consider security in the implementation of this research, such as limiting access to only patient and medical professionals. We hope this may improve the precision of treatments and interventions and ultimately to advance smart health.

REFERENCES

[1] 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.

[2] I. You, K.-K. R. Choo, C.-L. Ho, I. You, and K.-K. R. Choo, "A smartphone-based wearable sensors for monitoring real-time physiological data," *Comput. Elect. Eng.*, vol. 65, pp. 376–392, Jan. 2018.

[3] C. J. D'Orazio and K.-K. R. Choo, "Circumventing IOS security mechanisms for APT forensic investigations: A security taxonomy for cloud apps," *Future Gener. Comput. Syst.*, vol. 79, pp. 247–261, Feb. 2018.

[4] W. Meng, W. Li, L.-F. Kwok, and K.-K. R. Choo, "Towards enhancing click-draw based graphical passwords using multi-touch behaviours on smartphones," *Comput. Secur.*, vol. 65, pp. 213–229, Mar. 2017.

[5] C. J. D'Orazio and K.-K. R. Choo, "A technique to circumvent SSL/TLS validations on iOS devices," *Future Gener. Comput. Syst.*, vol. 74, pp. 366–374, Sep. 2017.

[6] C. J. D'Orazio, K.-K. R. Choo, and L. T. Yang, "Data exfiltration from Internet of Things devices: IOS devices as case studies," *IEEE Internet Things J.*, vol. 4, no. 2, pp. 524–535, Apr. 2017.

[7] Q. Do, B. Martini, and K.-K. R. Choo, "Is the data on your wearable device secure? An Android wear smartwatch case study," *Softw., Pract. Exper.*, vol. 47, no. 3, pp. 391–403, 2017.

[8] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in Parkinson's disease," *J. Neurosci. Methods*, vol. 167, no. 2, pp. 340–348, 2008.

[9] S. R. Hundza *et al.*, "Accurate and reliable gait cycle detection in Parkinson's disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 127–137, Jan. 2014.

[10] S. Wan, Y. Zhang, and J. Chen, "On the construction of data aggregation tree with maximizing lifetime in large-scale wireless sensor networks," *IEEE Sensors J.*, vol. 16, no. 20, pp. 7433–7440, Oct. 2016.

[11] K.-C. Lan and W.-Y. Shih, "Early diagnosis of Parkinson's disease using a smartphone," in *Proc. FNC/MobiSPC*, 2014, pp. 305–312.

[12] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2016.

[13] K. Wang, Y. Shao, L. Shu, C. Zhu, and Y. Zhang, "Mobile big data fault-tolerant processing for eHealth networks," *IEEE Netw.*, vol. 30, no. 1, pp. 36–42, Jan./Feb. 2016.

[14] K. Wang, Y. Shao, L. Shu, G. Han, and C. Zhu, "LDPA: A local data processing architecture in ambient assisted living communications," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 56–63, Jan. 2015.

[15] Y. Shao, K. Wang, L. Shu, S. Deng, and D.-J. Deng, "Heuristic optimization for reliable data congestion analytics in crowdsourced eHealth networks," *IEEE Access*, vol. 4, pp. 9174–9183, 2016.

[16] L. Xiao, X. Wan, C. Dai, X. Du, X. Chen, and M. Guizani. (2018). "Security in mobile edge caching with reinforcement learning." [Online]. Available: https://arxiv.org/abs/1801.05915

[17] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMed. Eng. OnLine*, vol. 6, no. 1, p. 23, 2007.

[18] B. E. Sakar *et al.*, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 4, pp. 828–834, Jul. 2013.

[19] 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.

[20] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, May 2012.

[21] E. Naydenova, A. Tsanas, C. Casals-Pascual, and M. De Vos, "Smart diagnostic algorithms for automated detection of childhood pneumonia in resource-constrained settings," in *Proc. IEEE Global Humanitarian Technol. Conf. (GHTC)*, Oct. 2015, pp. 377–384.

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

[23] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary ant lion approaches for feature selection," *Neurocomputing*, vol. 213, pp. 54–65, Nov. 2016.

[24] Y. Zhao and C. Yin, "The expected discounted penalty function under a renewal risk model with stochastic income," *Appl. Math. Comput.*, vol. 218, no. 10, pp. 6144–6154, 2012.

[25] M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent system for accurate detection of Parkinson's disease," *Comput. Methods Programs Biomed.*, vol. 113, no. 3, pp. 904–913, 2014.

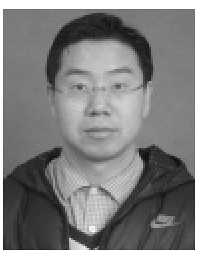
[26] J. Jiang, L. Liu, and Y. Wu, "Symmetric positive solutions to singular system with multi-point coupled boundary conditions," *Appl. Math. Comput.*, vol. 220, pp. 536–548, Sep. 2013.

[27] W.-L. Zuo, Z.-Y. Wang, T. Liu, and H.-L. Chen, "Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach," *Biomed. Signal Process. Control*, vol. 8, no. 4, pp. 364–373, 2013.

[28] F. Meng and J. Shao, "Some new Volterra–Fredholm type dynamic integral inequalities on time scales," *Appl. Math. Comput.*, vol. 223, pp. 444–451, Oct. 2013.

[29] Z. Cai, J. Gu, and H.-L. Chen, "A new hybrid intelligent framework for predicting Parkinson's disease," *IEEE Access*, vol. 5, p. 17188–17200, 2017.

[30] P. Li and G. Ren, "Some classes of equations of discrete type with harmonic singular operator and convolution," *Appl. Math. Comput.*, vol. 284, pp. 185–194, Jul. 2016.
[31] C. G. Goetz *et al.*, "Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Process, format, and clinimetric testing plan," *Movement Disorders*, vol. 22, no. 1, pp. 41–47, 2007.
[32] S. Arora, V. Venkataraman, S. Donohue, K. M. Biglan, E. R. Dorsey, and M. A. Little, "High accuracy discrimination of Parkinson's disease participants from healthy controls using smartphones," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 3641–3644.
[33] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754–767, Jan. 2016.
[34] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. ESANN*, 2013, pp. 437–442.
[35] L. Palmerini, L. Rocchi, S. Mellone, F. Valzania, and L. Chiari, "Feature selection for accelerometer-based posture analysis in Parkinson's disease," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 3, pp. 481–490, May 2011.
[36] M. Alhussein, "Monitoring Parkinson's disease in smart cities," *IEEE Access*, vol. 5, pp. 19835–19841, 2017.
[37] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi, "Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system," *IEEE Access*, vol. 5, pp. 22199–22208, 2017.
[38] H.-C. Chang, Y.-L. Hsu, S.-C. Yang, J.-C. Lin, and Z.-H. Wu, "A wearable inertial measurement system with complementary filter for gait analysis of patients with stroke or Parkinson's disease," *IEEE Access*, vol. 4, pp. 8442–8453, 2016.
[39] A. Murad and J.-Y. Pyun, "Deep recurrent neural networks for human activity recognition," *Sensors*, vol. 17, no. 11, p. 2556, 2017.
[40] B. Andò *et al.*, "A wearable device to support the pull test for postural instability assessment in Parkinson's disease," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 218–228, Jan. 2018.
[41] J. Camps *et al.*, "Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit," *Knowl.-Based Syst.*, vol. 139, pp. 119–131, Jan. 2018.
[42] C. L. Pulliam, D. A. Heldman, E. B. Brokaw, T. O. Mera, Z. K. Mari, and M. A. Burack, "Continuous assessment of levodopa response in Parkinson's disease using wearable motion sensors," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 159–164, Jan. 2018.
[43] L. Moro-Velázquez, J. A. Gómez-García, J. I. Godino-Llorente, J. Villalba, J. R. Orozco-Arroyave, and N. Dehak, "Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson's disease," *Appl. Soft Comput.*, vol. 62, pp. 649–666, Jan. 2018.
[44] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explorations Newslett.*, vol. 12, no. 2, pp. 74–82, Dec. 2010.
[45] S. A. Mostafa *et al.*, "Evaluating the performance of three classification methods in diagnosis of Parkinson's disease," in *Proc. Int. Conf. Soft Comput. Data Mining*. Cham, Switzerland: Springer, 2018.
[46] M. Lichman. (2013). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml
[47] C. Stamate *et al.*, "Deep learning Parkinson's from smartphone data," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2017, pp. 31–40.



SHAOHUA WAN received the Ph.D. degree from the School of Computer, Wuhan University, in 2010. From 2016 to 2017, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, Technical University of Munich, Germany. Since 2015, he has been a Post-Doctoral Researcher with the State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology. He is currently an Associate Professor and a Master Advisor with the School of Information and Safety Engineering, Zhongnan University of Economics and Law. His main research interests include massive data computing for Internet of Things and edge computing.

YAN LIANG is currently pursuing the master's degree with the School of Information and Safety Engineering, Zhongnan University of Economics and Law, China. Her main research interests are edge computing.

YIN ZHANG (SM'16) is currently an Associate Professor with the School of Information and Safety Engineering, Zhongnan University of Economics and Law (ZUEL), China. He is a Wenlan Distinguished Scholar with ZUEL and a Chutian Distinguished Scholar in China. He has authored over 80 prestigious conference and journal papers, including eight ESI Highly Cited Papers. His research interests include intelligent service computing, big data, and social network. He is the Vice-Chair of the IEEE Computer Society Big Data STC. He serves as an Editor or an Associate Editor for the IEEE Access, the IEEE Sensors Journal, and the *Journal of Information Processing Systems*. He is a Guest Editor of the *Future Generation Computer Systems*, the IEEE IoT Journal, the *Mobile Networks and Applications*, the *Sensors*, the *Multimedia Tools and Applications*, the *Wireless Communications and Mobile Computing*, the *Electronic Markets*, the *Journal of Medical Systems*, and the *New Review of Hypermedia and Multimedia*. He served as the Track Chair for the IEEE CSCN 2017 and the TPC Co-Chair for CloudComp 2015 and TRIDENTCOM 2017.

MOHSEN GUIZANI received the bachelor's degree (Hons.) and the master's degree in electrical engineering, and the master's and doctorate degrees in computer engineering from Syracuse University, Syracuse, NY, USA, in 1984, 1986, 1987, and 1990, respectively. He was the Associate Vice President of graduate studies with Qatar University; the Chair of the Computer Science Department, Western Michigan University; and the Chair of the Computer Science Department, University of West Florida. He held academic positions with the University of Missouri–Kansas City, University of Colorado-Boulder, Syracuse University, and Kuwait University. He is currently a Professor and the Electrical and Computer Engineering Department Chair with the University of Idaho. He has authored nine books and over 400 publications in refereed journals and conferences. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He also served as a member, the Chair, and the General Chair of a number of international conferences. He received the Best Research Award from three institutions. He currently serves on the editorial boards for several international technical journals and the Founder and the Editor-in-Chief for *Wireless Communications and Mobile Computing Journal* (Wiley). He was a Guest Editor of a number of special issues in IEEE journals and magazines. He was selected as the Best Teaching Assistant for two consecutive years at Syracuse University. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the TAOS Technical Committee. He served as the IEEE Computer Society Distinguished Speaker from 2003 to 2005.