A Cascade Ensemble Learning Model for Parkinson’s Disease Diagnosis Using Handwritten Sensor Signals

Shoujiang Xu¹², Zhaohui Zhu¹ and Zhigeng Pan*¹

¹ Virtual Reality and Intelligent Systems Research Institute, Hangzhou Normal University, 2318 Yuhangtang Road, Hangzhou, Zhejiang, China
² School of Information Engineering, Jiangsu Food and Pharmaceutical Science College, 4 Meicheng Road, Huaian, Jiangsu, China
Email: zgpan@hznu.edu.cn

Abstract. Computer-aided diagnosis of Parkinson’s disease (PD) has gained lots of attention recently, which is mainly performed with machine learning methods using PD patients’ clinical manifestations, such as freezing gait, distorted writing, and abnormal speech. This paper presents a new methodology to differentiate PD patients from healthy controls (HC) based on two datasets regarding handwritten sensor signals. And a novel cascade ensemble learning method which is composed of two random forest (RF) classifiers and two extremely random trees (ExtraTrees) classifiers in each layer is proposed. Augmented features are generated by the four classifiers in a layer, which will be concatenated with the initial input data and fed to the next cascade layer. Finally, the classification result will be obtained from the final layer. To improve the classification performance, we employed principal component analysis (PCA) technique to reduce the dimensionality of sampled signals before they are fed to the ensemble model. Experimental results show that the proposed framework achieved reasonable classification performances with 81.17% accuracy for PD diagnosis.

1. Introduction

PD is a common movement disorder in middle-aged and elderly people. It was originally described by a doctor called James Parkinson in 1817. PD is a chronic and neurodegenerative disease, because of the lack of neurotransmitter dopamine in the brain [1]. PD patients have some motor symptoms such as essential tremor, freezing gait, bradykinesia, and postural instability, as well as some non-motor symptoms such as sleep, speech, urinary and olfactory disorders [2]. Younger trend of patients with PD emerged in recent years. However, PD patients are difficult to detect and easily overlooked in the early stage, which leads to miss the optimal treatment time. And because of the long duration and high disability rate of PD, there is no effective cure currently, and the economic burden of patients, families and society is extremely heavy [3].

As the artificial intelligence develops, machine learning methods has been widely applied in many domains. Among them, the research and application of machine learning methods in the medical fields are particularly prominent [4-6]. For example, intelligent-aided diagnosis can assist doctors’ diagnosis, improves the accuracy of doctors’ diagnosis, and saves people's time for medical treatment. In recent years, with the development of artificial intelligence technique, some achievements have been made in the auxiliary diagnosis of PD and intelligent disease severity rating [7]. The development of machine learning technology provides intelligent decision-making for the computer-aided classification and diagnosis of PD, avoiding the dilemma of diagnosis without professional clinicians. Machine learning
models can help detect PD patients in the early stage, facilitate early and timely intervention to prevent the development of the disease. At present, many typical manifestations of PD in clinical are utilized to develop machine learning models for PD detection, such as speech disorder [8], clinical questionnaires [9], and freezing gait [10].

In this paper, we designed a novel framework based on the idea of PD patients’ micro-graphing (distorted and smaller writing) caused by motor symptoms. First, sensor signals of two specific writing exams of PD patients using a smart pen are pre-processed and PCA is employed to reduce dimensionality of the sample set. More importantly, an innovative cascade ensemble learning method fusing RF and ExtraTrees classifiers is designed to train the PD classification model. Compared with support vector machines (SVM) and RF classifiers, the proposed model has achieved more promising performances.

2. Related Work

Intelligent diagnosis models for PD detection using data-driven approach have attracted more attentions recently. The data used for PD diagnosis are mainly consisted of different signals recorded by different instruments and will be pre-processed and fed to the classifiers for PD detection. The classifiers used are mainly divided into conventional machine learning approaches and deep learning methods. For instance, Braga [11] proposed a methodology to early detect PD using speech datasets which are in uncontrolled background conditions based on RF and SVM classifiers which belong to conventional methods. Parisi [12] designed a hybrid feature-driven machine learning method fusing multi-layer perception and Lagrangian SVM to facilitate the prognostic assessment of PD patients. Multi-layer perception was employed for feature selection and the selected features were considered as input to the Lagrangian SVM for classification. Parisi [13] invented a novel spectral data representation strategy based on freezing of gait signals of previous and current windows, then the data representation was utilized to train a deep learning model to detect freezing of gait episodes in PD patients. The approach presented achieved more promising performances than previous methods. In Ref. [14], Hakan employed two Convolutional Neural Networks (CNN) trained with different combination of various speech feature sets to classify PD patients. In Ref. [15], a two-channel deep learning model fusing a CNN model for spatial features and a long short-term memory model for temporal patterns was developed to automatically rate the PD severity from gait sensor signals. Above all, various clinical manifestations of PD have been utilized to distinguish PD patients from healthy individuals.

Furthermore, many studies regarding handwritten dynamics were conducted for PD detection using machine learning methods. First, a dataset HandPD [16] concerning different handwriting exams was built by Pereira et al. The PD patients and HC performed six different exams using a smart pen during the test and six different kinds of sensor signals were recorded by the pen. In Ref. [17], Pereira coped with automatic PD recognition by means of a CNN model. Features were extracted respectively from the third and fourth exams of HandPD dataset for classification. Results showed that the CNNs were able to learn features from the signals successfully and achieve outperforming results. Also, Pereira [18] utilized the image processing techniques to automatically separate the template and the drawings of the third and fourth exams which are drawing spirals and meanders in forms, respectively. Then Optimum-Path Forest, Naïve Bayes and SVM were employed for PD classification. In Ref. [19], the sensor signals were transformed to image domain by means of recurrence plots, and the generated images were fed to CNNs for classification. Compared to previous works [18, 19] on the meander and spiral dataset of HandPD, significant performance improvement has been achieved with a mean accuracy of over 87%. The above three literatures all have made some contributions to PD classification based on handwritten dynamics dataset named HandPD which is an unbalanced dataset containing 74 PD patients and 18 healthy individuals. Also, Refs. [17, 19] utilized hold-out strategy to perform data partitioning and obtained relative high accuracy results. Instead, Ref. [18] adopted cross validation which can generate more objective and convincing results than hold-out strategy to split the dataset and evaluate the classification results, with around 67% accuracy obtained. Recently, new
version of handwritten dynamics dataset has been released with 31 PD patients and 35 healthy individuals and this dataset named as NewHandPD is a balanced one. To evaluate the PD classification performances scientifically, we will employ stratified k-fold cross validation strategy to split the data and compute the average performances using all splits.

3. Dataset
NewHandPD dataset was collected at Botucatu Medical School, São Paulo State University - Brazil and contains 66 individuals in total. Each person was requested to perform 12 handwriting exams using a smart pen, being 4 of them related to drawing spirals and 4 related to drawing meanders. The rest 4 movements are drawing a circle in the air, drawing a circle on the paper, left-handed diadochokinesis and right-handed diadochokinesis, respectively. This paper mainly focuses on the spiral and meander exams to identify PD patients. The smart pen was equipped with 6 different sensors to capture a 6-channel signal when an exam was performed by an individual, as shown figure 1. Figure 1a represents a spiral drawn by a healthy individual and figure 1b is the corresponding sensor signal. Figure 1c represents a meander drawn by a PD patient and figure 1d is the corresponding sensor signal. It is inferred that, for the spiral and meander exams, 264 6-channel sensor signals for each exam were collected in the NewHandPD dataset.

![Figure 1. Illustration of handwriting images and the corresponding sensor signals.](image)

4. Methodology
In this paper, we developed a new methodology to classify PD using handwritten dynamics dataset and flowchart of the proposed framework is shown in figure 2. We utilize the sensor signals of spiral
and meander exams to recognize PD patients, respectively. First data pre-processing and dimensionality reduction are performed and then the generated features are fed to the proposed novel cascade ensemble learning model for PD classification.

![Figure 2. Overview of the proposed framework for PD prediction.](image)

5. Data Preprocessing

Due to the different duration of time taken by different persons when drawing a meander or spiral, we need to subsample each channel of sensor signals with a same length N for further processing. Consequently, a sensor signal's dimension is (6, N). To obtain more convincing results, we empirically set N = 3000 when the two exams were employed to classify PD.

5.1. Dimensionality Reduction Using PCA

To extract the handwriting features more concisely, PCA technique is employed to reduce the dimensionality of the two sample sets and extract effective components to improve the classification efficiency. Suppose that the number of signals in spiral dataset is m, and we define the ith sensor signal $x_i$ of sample set $X$ as $[x_i^1, x_i^2, x_i^3, ..., x_i^d]$, where $1 \leq i \leq m$ and $d = 6N$. The concrete steps of PCA is as follows:

Step 1: Obtain the mean value of $x_i$ according to equation (1).

$$ E = \frac{1}{m} \sum_{i=1}^{m} x_i $$

Step 2: Calculate the covariance matrix $C$ of the sample set based on equation (2).

$$ C = \frac{1}{m} \sum_{i=1}^{m} (x_i - E)(x_i - E)^T $$

Step 3: Calculate the eigenvalues $\lambda$ and eigenvectors $Q$ based on equation (3) to 5.

$$ C = Q \cdot \sum_{i=1}^{d} Q $$

$$ \sum_{i=1}^{d} = diag(\lambda_1, \lambda_2, \ldots, \lambda_d) \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0 $$

$$ Q = [q_1, q_2, \ldots, q_d] $$

Step 4: Calculate the cumulative variance contribution rate $\theta$ of the first $k$-row principal elements according to equation (6).

$$ \theta = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i} $$

Step 5: Suppose the first-$k$ rows of matrix $Q$ as $Q_k$, then transformed set $Y$ with $k$-dimension is calculated on equation (6). In our experiments, we empirically set $k = 200$ for the meander and spiral datasets, and two $\theta$ values are obtained (0.99 < $\theta$ < 1).
AICS 2020
Journal of Physics: Conference Series 1631 (2020) 012168 doi:10.1088/1742-6596/1631/1/012168

\[ Y = X \cdot Q_k \] (7)

5.2. The Cascade Ensemble Learning Model
Deep forest is a novel ensemble learning method invented by Zhou [20] and has achieved promising results in many domains [21, 22]. As an alternative deep learning method, it has the characteristic of few parameters and easy implementation. In this paper, we incorporate the core idea of deep forest composed of RF and completely-random tree classifiers and design a novel cascade ensemble learning model with RF [23] and ExtraTrees classifiers [24] to automatically differentiate PD patients from HC, as shown in figure 3. In the proposed model, each classifier of current layer will generate 2 augmented features and then total 8 augmented features in a layer concatenated with the input data will be fed into the next layer. If the classification performances are not significantly improved after W layers, the training process will terminate automatically. In this paper, to gain better performances, we set W=3.

![Figure 3. The proposed cascade ensemble learning model.](image)

Random forest and ExtraTrees classifiers fused in the model are both based on randomized decision trees. More concisely, the main core idea behind the two basic classifiers is as shown in figure 4. For a specific classifier, the final prediction is generated by all the decision trees each of which will generate a class vector representing the probabilities of classification. The red path represents a class vector’s generation for a given instance through a decision tree. The concrete steps of RF in our paper are as follows: (1) Generating T sample sets. Randomly select M samples as a sample set using bootstrap sampling strategy for constructing a decision tree. (2) Construct T Classification and Regression Trees (CART). Randomly select m candidate features from 6N features (m <<6N) and one feature with best Gini value will be chose to split when generating a decision tree without pruning. (3) Obtain final prediction using T CARTs and majority voting strategy. ExtraTrees is also tree-based ensemble learning methods and its two main differences with RF are that choosing cut-points for split is fully random and the decision trees of ExtraTrees are generated using the whole learning sample.

Each classifier of the current layer in the proposed model will produce a class vector as an augmented feature for a given instance. To reduce the risk of overfitting, 5-fold cross validation is employed to generate class vectors during the training stage, which means each instance of the sample set will be considered as training data for 4 times and 4 class vectors will be generated for a given instance. Then augmented features for an instance will be generated by averaging all the 4 class vectors. Finally, 8 augmented features will be produced for an instance in a layer, and we can obtain the classification accuracy of the current layer in the cascade ensemble learning model using the strategy of final layer in figure 3. The augmented features concatenated with the input will be fed to
next layer for further learning, meanwhile, classification accuracy of each layer will be calculated till the termination condition is satisfied.

![Randomized decision trees](image)

**Figure 4.** The core idea of randomized decision trees.

5.3. **Stratified 5-Fold Cross Validation**

To evaluate the proposed framework for PD detection more objectively and accurately, we employ stratified 5-fold cross validation to generate the training and test dataset. The sample set after dimensionality reduction will be split into 5 subsets each of which has the same proportion of PD patients and healthy individuals. More importantly, the persons in a pair of training and test datasets of a split are mutually exclusive. Also, each instance in the sample set will be considered as training data for 4 times. The classification performances of each pair of training and test sets will be averaged to estimate the performances of the proposed framework. What needs to be emphasized is that the 5-fold cross validation used in Section 4.3 plays a different role with this stratified 5-fold cross validation. The cross validation in Section 4.3 is designed for augmented features generation and avoiding the risk of overfitting, instead the cross validation of this section is designed to evaluate the performances of the proposed method more accurately and convincingly.

5.4. **Evaluation Metrics**

According to the medical diagnosis in clinical, the positive result represents the individual with diseases and negative result the healthy one. Table 1 shows the relation between the predicted class and the true class. The metrics for evaluating PD classification performances used in this paper are accuracy (ACC), specificity (SPE), sensitivity (SEN) and F1-score, as shown in equation (8).

| Table 1. Relation between predicted and true class. |
|-----------------------------------------------|
| **Predicted class** | **True class** | **Negative (HC)** |
| Positive | True positive (TP) | False positive (FP) |
| Negative | False negative (FN) | True negative (TN) |

\[
\begin{align*}
    ACC &= \frac{TN + TP}{TN + FP + FN + TP} \\
    SPE &= \frac{TN}{TN + FP} \\
    SEN &= \frac{TP}{TP + FN} \\
    F1\text{-score} &= \frac{2TP}{2TP + FP + FN}
\end{align*}
\] (8)
6. Experimental Results

We conducted the experiments on windows 10 platform with hardware configuration listed as follows: CPU is Inter® Core™ i7-6700 3.40GHZ and RAM is 16GB. The programming language is python3.5 and machine learning library we used is Scikit-learn. For comparisons, we chose two representative classifiers with PCA to classify PD patients from healthy individuals and the classifiers are SVM with PCA (SVM_pca) and RF with PCA (RF_pca) respectively. For RF_pca and our proposed model, the number T of decision trees were both equal to 500 which is recommended in paper [20] and the other hyper-parameters are default values of Scikit-learn library. For SVM classifiers, we found linear SVM can achieve better results than the other kernel functions. Then we used grid-search strategy to obtain the optimal penalty coefficient C which is limited to [1, 1e1, 1e2, 1e3, 1e4]. Besides, we select the first 200 principal elements when reducing the dimensionality using PCA. On the meander and spiral datasets, we performed 20 consecutive runs for each classifier and average performance comparisons are shown in tables 2 and 3.

Table 2. Performance comparisons of different classifiers on meander dataset.

| Classifiers | ACC   | SPE   | SEN   | F1-score |
|-------------|-------|-------|-------|----------|
| SVM_pca     | 70.86%| 67.71%| 74.45%| 70.23%   |
| RF_pca      | 77.49%| 79.96%| 74.66%| 75.24%   |
| Our model   | 77.38%| 76.31%| 80.66%| 78.05%   |

From table 2, we found RF_pca and our proposed ensemble learning model achieved more promising results than SVM_pca on meander dataset when T=500. The two classifiers’ performances ACC, SPE, F1-Score exceed around 10% than SVM_pca, and RF_pca and our model achieved comparable results on meander dataset. Particularly, our model obtained the best performances of F1-score and sensitivity, and RF_pca achieved the optimal results in terms of accuracy and specificity. From table 3, our model generated the optimal performances regarding ACC, SEN, and F1-score on spiral dataset. The SVM_pca’s performances are also the lowest among the three classifiers. It is inferred that the classification performances of spiral dataset are slightly better than those of meander dataset.

Table 3. Performance comparisons of different classifiers on spiral dataset.

| Classifiers | ACC   | SPE   | SEN   | F1-score |
|-------------|-------|-------|-------|----------|
| SVM_pca     | 76.69%| 69.14%| 85.15%| 77.38%   |
| RF_pca      | 79.54%| 76.64%| 82.90%| 79.10%   |
| Our model   | 80.99%| 76.38%| 88.98%| 79.86%   |

To further verify the effectiveness of our proposed ensemble learning model, we have performed some PD classifications using RF_pca and the proposed model when the number T of decision trees is equal to 1000, and the results are shown in table 4. We found that the two comprehensive performance metrics of our model are both better than RF_pca on the two datasets. The accuracy we obtained on meander dataset is 78.18% and F1-score 78.85%. The accuracy we obtained on spiral dataset is 81.17% and F1-score 80.51%. We can infer two conclusions about our proposed framework for PD detection using the handwriting sensor signals of drawing meanders and spirals. (1) The results of spiral dataset are better than those of meander dataset when using the same classifier. (2) the performances of the two tree-based models when T=1000 is better than the corresponding performances when T=500.

During the experiments, we conducted all the above PD classifications using stratified 5-fold cross validation. We can conclude that the proposed model with PCA on spiral dataset can achieve the optimal classification performances for PD detection in our study.
In this paper, we proposed a new framework based on machine learning methods and PCA technique to assist recognizing PD patients using handwriting sensor signals. Compared to conventional classifiers, our proposed classifier with PCA achieved the best performances on the meander and spiral dataset. The model has the characteristic of few hyper-parameters to finetune and easy training. Also, we chose stratified 5-fold cross validation which is more objective and convincing than previous hold-out strategy [17, 19] in the process of data partitioning and performance evaluation.

With the outstanding development of deep learning methods in many recognition domains, in the future works, we will try to employ CNNs with stratified 5-fold cross validations to distinguish PD patients from healthy individuals on these two datasets and better evaluate the classification performances.

Also, the handwritten dynamics test for PD patients contains 6 different exams, and we only evaluate the classification performances on the meander and spiral exam datasets separately in this paper. The more scientific way of PD recognition is to model a specific classifier for predicting the probability on each exam of handwritten dynamics and make a final prediction for an individual using majority voting strategy.

Furthermore, the current computer-aided method of diagnosing PD using clinical manifestations only employ a single dysfunction like speech disorder, clinical questionnaire, and handwriting test to learn the potential patterns of PD patient. In fact, the physicians always recognize PD patients from different manifestations and make the comprehensive prediction in clinical. Naturally, the multi-modal information composed of many PD’s clinical manifestations should be combined to detect the PD patients.

8. Conclusions

In this paper we proposed a novel framework fusing PCA technique and a new cascade ensemble learning model to prove the feasibility of diagnosing PD using handwriting sensor signals and machine learning methods. Two basic classifiers based on randomized decision trees are fused in our proposed model, which is different from the original deep forest composed of two random forest classifiers and two completely-random tree classifiers. Reasonable experimental results proved that there exist the potential patterns which can be utilized to assist diagnosing PD patients behind the handwritten dynamics test.

Acknowledgement

This study was funded by National Key R&D Program of China (grant number 2018YFB1004901).

References

1. Lees A J, Hardy J and Revesz T 2009 Parkinson’s disease The Lancet 73 2055-2066.
2. Jankovic J 2008 Parkinson’s disease: clinical features and diagnosis J. Neurol. Neurosurg. Ps. 79 368-376.
3. Zhao Y J, Tan L C, Lau P N, Au W L, Li S C and Luo N 2008 Factors affecting health-related quality of life amongst asian patients with parkinson’s disease Eur. J. Neurol. 15 737-742.
4. Vo D M, Nguyen N and Lee S 2019 Classification of breast cancer histology images using incremental boosting convolution networks Inform. Sciences 482 123-138.
5. Jin L and Dong J 2018 Intelligent health vessel ABC-DE: an electrocardiogram cloud computing service IEEE T. Cloud Comput. 1-1.

Table 4. Performance comparisons of classifiers on meander and spiral dataset when T=1000.

| Classifiers  | Meander dataset | Spiral dataset |
|-------------|-----------------|----------------|
|             | ACC  | F1-score | ACC  | F1-score |
| RF_pca      | 78.05% | 75.28%   | 79.81% | 78.96%   |
| Our model   | 78.18% | 78.85%   | 81.17% | 80.51%   |

8
[6] Jo T, Nho K and Saykin A J 2019 Deep learning in alzheimer’s disease: diagnostic classification and prognostic prediction using neuroimaging data Front Aging Neurosci. 11 220.

[7] Pereira C R, Pereira D R, Weber S A, Hook C, Albuquerque V H C D and Papa J P 2019 A survey on computer-assisted Parkinson’s disease diagnosis Artif. Intell. Med. 95 48-63.

[8] Tsanas A, Little M A, Mcsharry P E, Spielman J and Raming L O 2012 Novel speech signal processing algorithms for high-accuracy classification of Parkinson’s disease IEE T. Bio-Med. Eng. 59 1264-1271.

[9] Prashanth R and Roy S D 2018 Early detection of Parkinson’s disease through patient questionnaire and predictive modelling Int. J. Mea. Inform. 119 75-87.

[10] Brognara L, Palumbo P, Grimm B and Palmerini L 2019 Assessing gait in Parkinson’s disease using wearable motion sensors: A systematic review Diseases 7 18.

[11] Braga D, Madureira A, Coelho L and Ajith R 2019 Automatic detection of Parkinson’s disease based on acoustic analysis of speech Eng. Appl. Artif. Intel. 77 148-158.

[12] Parisi L, Narrendar R and Marianne L M 2018 Feature-driven machine learning to improve early diagnosis of Parkinson’s disease Expert Sys. Appl. 110 182-190.

[13] Camps J, Samà A, Martín M, Rodríguez-Martín D, Pérez-López C, Arostegui J M M, Cabestany J, Català A, Alcaine S, Mestre B, et al. 2018 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. 139 119-131.

[14] Hakan G 2019 Deep learning-based Parkinson’s disease classification using vocal feature sets IEEE Access 7 115540-115551.

[15] Zhao A, Qi L, Li J, Dong J and Yu H 2018 A hybrid spatio-temporal model for detection and severity rating of Parkinson’s disease from gait data Neurocomputing 315 1-8.

[16] http://wwwp.fc.unesp.br/~papa/pub/datasets/Handpd/.

[17] Pereira C R, Weber S A T, Hook C, Rosa G H and Papa J P 2016 Deep learning-aided Parkinson’s disease diagnosis from handwritten dynamics 29th SIBGRAPI Conf. on Graphics, Patterns and Images (Sao Paulo, Brazil) pp 340-346.

[18] Pereira C R, Pereira D R, Silva F A D, Masieiro J P, Weber S A T, Hook C and Papa J P 2016 A new computer vision-based approach to aid the diagnosis of Parkinson’s disease Comput. Meth. Prog. Bio. 136 79-88.

[19] Afonso L C S, Rosa G H, Pereira C R, Weber S A T, Hook C, Albuquerque V H C D and Papa J P 2019 A recurrence plot-based approach for Parkinson’s disease identification Future Gener. Comp. Sys. 94 282-292.

[20] Zhou Z and Feng J 2018 Deep forest arXiv: Learning.

[21] Xu S, Tang Q, Jin L and Pan Z 2019 A cascade ensemble learning model for human activity recognition with smartphones Sensors 19 2307.

[22] Utkin L V and Ryabinin M A 2018 A siamese deep forest Knowl-Based Syst. 139 13-22.

[23] Breiman L 2001 Random forests Mach. Learn. 45 5-32.

[24] Geurts P, Ernst D and Wehenkel L 2006 Extremely randomized trees Mach. Learn. 63 3-42.