

Parkinson's Disease Detection Using Ensemble Techniques and Genetic Algorithm

Najmeh Fayyazifar

Information Technology and Computer Engineering
Department
Shahrood University of Technology
Shahrood, Iran
najme.fayazi@shahroodut.ac.ir

Najmeh Samadiani

Computer Engineering Department
Kosar University of Bojnord
Bojnord, Iran
nsamadiani@kub.ac.ir

Abstract—Parkinson's disease (PD) is a neurological disorder which progress by time. People suffering from PD experience shortage of Dopamine which is a chemical present in brain nerve cells. The symptoms of PD are tremor, rigidity, and slowness of movements and people with PD experience more severity by time progress. Therefore, the automation in early detection of PD is an important issue. In the literature, different classification methods have been proposed. Also, due to the high dimension of extracted features of voice, many feature selection algorithms have been developed. In this paper, we aim to propose a method for early detection of PD from voice recordings. The Genetic algorithm is used to select the optimal set of features which can reduce feature vector dimension from 22 to 6 features. We have achieved 96.55% and 98.28% detection rate by employing AdaBoost and Bagging algorithms for classification process, respectively.

Keywords—Parkinson's disease (PD); Ensemble algorithms; Genetic algorithm; feature selection.

I. INTRODUCTION

Parkinson's disease is a long-term degenerative disorder which affects the motor system. The most obvious symptoms are shaking, difficulty with walking, thinking, talking and drinking. It happens when there is a problem with certain nerve cells in the brain which make Dopamine, a chemical responsible for sending signals to a region of the brain which controls muscles movement. It is estimated that 7 to 10 million people are suffering from PD [1]. The disorder doesn't directly lead to death but symptoms get worse by time progress and the patient experience more server condition. Thus, the early detection and medication can provide a more convenient situation for people with PD. Unfortunately, there is no single definitive diagnostic test currently and detection of PD strongly relies on the information that the patient provides. Therefore, developing an automatic system which can detect the disorder in early stages is very essential.

In the literature, many studies have been conducted to detect the early PD. These investigations have studied the gait or speech signals of patients. The authors of [2] evaluated the performance of disparate image and video processing techniques to detect PD through assessment of gait. The researchers of [3] employed the silhouette method to recognize the gait pattern.

In the speech of people with PD, there are some subtle abnormalities which may not be perceived by listeners but can be beneficial to detect PD by employing acoustical analysis. Recently, many techniques have been developed to analyze the speech in order to classify people suffering from PD and healthy people. In these researches, the speech of people has been recorded and some specific features have been extracted to discriminate people with PD from healthy subjects by employing different techniques [4-7].

Some recent studies have aimed to propose new features while others have employed different feature selection methods to discard the replicated features and reduce the size of feature vector used for classification purpose. A comparative research of various inspired algorithms is represented in [8] which evaluates performance of different evolutionary methods in choosing optimal features needed for diagnosis of PD. Also, some investigations applied various machine learning and classification methods. Reference [9] proposed spirometry features to assess the motor function in Parkinson's disease. The authors of [10] used the non-motor features of sleep Behavior Disorder (RBD) and Cerebrospinal fluid (CSF) measurements to detect early PD. They applied Naïve Bayes, Support Vector Machine (SVM), Boosted Trees and random Forest classifiers.

Chen et.al [11] investigated the performance of extreme learning machine (ELM) and kernel ELM (KELM) for early diagnosis of PD. They also employed four feature selection methods including maximum relevance minimum redundancy (mRMR), Information Gain (IG), relief and t-test. The researchers of [12] employed a non-invasive method to detect early PD. They fed wavelet features to SVM classifier in order to identify Parkinson's gait. Reference [13] proposed a Q-back propagated time delay neural network (Q-BTDNN) classifier that creates a temporal classification model, which performs the classification and prediction task in clinical decision-making system (CDMS).

Comparing to the works discussed in the literature our proposed method has higher classification accuracy while reduces the size of feature vector significantly. We have applied ensemble learning methods in early detection of PD which has not been previously employed.

The rest of this paper is organized as follows. In section II a short overview of the Genetic algorithm is proposed. Section III offers brief background knowledge on ensemble techniques including AdaBoost and Bagging. Experimental results are presented in section IV. Finally, section V concludes the conclusion of our research.

II. GENETIC ALGORITHM

A genetic algorithm is an **explorative procedure** that is mostly able to discover optimal solutions to complex problems [14]. For this purpose, the algorithm holds a set of trial solutions (often called individuals) and forces them to evolve toward an appropriate solution. First, a representation for possible solutions must be developed. Then, the genetic algorithm starts with an initial random population. The fittest population in each generation survives. By using this population, besides the knowledge available in the gene pool, the ability of a new generation to solve the problem is improved. This goal is achieved by employing a four-step process including **evaluation, reproduction, recombination, and mutation**.

Genetic algorithm also provides optimal solution in feature selection problems. Authors in [15] have used Genetic algorithm for selection of optimal features for a classification problem. In this problem, the values of a binary vector which demonstrate the presence of a feature in a reduced feature vector are optimized. Indeed, the number of gens in each chromosome represents the number of features available in primary feature vector. In each instance, if a feature contributes in an optimal and reduces feature vector, 1 is associated to the corresponding gene, otherwise it is associated with 0. In this paper, Genetic algorithm is applied to select the most appropriate and optimal features which provide the highest detection accuracy of PD using the Ensemble classification methods.

III. ENSEMBLE ALGORITHMS

Ensemble algorithms, better to be called “**meta-algorithms**” use **multiple learning algorithms** and combine them into a predictive model for **decreasing the variance** (Bagging), **bias** (Boosting) or **improving the predictive force** (stack alias ensemble). In this paper, AdaBoost and Bagging algorithms are used as classifiers in order to detect PD.

A. AdaBoost Algorithm

Boosting is a general and effective method which produces a highly accurate classifier by combining rough and quite inaccurate rules of thumb. This strategy has roots in the fact that finding many rough rules of thumb can be much easier than finding a single, highly accurate classifier. An algorithm called a “**weak learner**” is defined to find the rules of thumb. The Boosting algorithm repeatedly calls this weak learner. Each time a **different distribution of training data** is fed to this weak learner (AdaBoost). Each call generates a weak classifier which must be combined into a single classifier that is significantly more accurate than any one of the rules.

AdaBoost, short for “Adaptive Boosting”, is an algorithm for constructing a strong classifier as a linear combination of weak algorithms. It is a popular Boosting algorithm for binary

classification. The algorithm trains learners sequentially. For every learner with index t , AdaBoost **computes the weighted classification error**:

$$\epsilon_t = \sum_{n=1}^N d_n^{(t)} I(y_n \neq h_t(x_n)) \quad (1)$$

Where x_n is a vector of predictor values for observation n . y_n is the true class label. h_t is the prediction of the learner (hypothesis) with index t . I is the indicator function. $d_n^{(t)}$ is the weight of observation n at step t .

AdaBoost then **increases weights for observations misclassified by learner t and decreases weights for observations correctly classified by learner t** . The next learner $t+1$ is then trained on the data with updated weights $d_n^{(t+1)}$.

When training is finished, AdaBoost computes prediction for new data using

$$f(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (2)$$

Where $\alpha_t = \frac{1}{2} \log \frac{1-\epsilon_t}{\epsilon_t}$ is the weight of the weak hypotheses in the ensemble.

Training by AdaBoost can be viewed as **stage-wise minimization of the exponential loss**:

$$\sum_{n=1}^N w_n \exp(-y_n f(x_n)) \quad (3)$$

Where $y_n \in \{-1, +1\}$ is the true class label. w_n are normalized observation weights to add up to 1 and $f(x_n) \in (-\infty, +\infty)$ is the predicted classification score [16-17].

B. Bagging algorithm

Bagging (Breiman, 1996) is named after “**bootstrap aggregation**”. Bagging was the first beneficial ensemble learning method and also one of the more conventional methods used for arching [18]. The meta-algorithm is a kind of **model averaging**. Basically, it was developed for classification purposes and is commonly applied to decision tree models; however, it can be used with any kind of model for classification or regression. In this method, **multiple versions of the training set are used by employing bootstrap, i.e. sampling with replacement**. Each of these data sets trains a different model. The outputs of the models are combined by averaging (in regression problems) or voting (in classification cases) to have a single output. Bagging is merely **effective if unstable** (a small change in the training set can lead to a notable change in the model) **nonlinear models** are used.

Bagging operates by **training learners on resampled versions of data**. Resampling commonly is done by **bootstrapping observations which means selecting N out of N observations with a replacement for every new learner**. Furthermore, each tree in the ensemble is able to randomly **select predictors for decision splits**. This random selection is made for all splits and **each tree includes numerous splits**. In this paper, this parameter is set to a square root of the number of predictors [19] and the output of multiple classifiers is combined by voting.

IV. METHODOLOGY

In this paper, early Parkinson's disease is detected using machine learning techniques. Firstly, Genetic algorithm is employed to reduce the size of feature vector and select the optimal set of features. Then two ensemble methods including AdaBoost and Bagging are used to classify input data into two healthy and patient categories.

The Genetic algorithm is an efficient optimization technique which can aims to the selection of optimal features in classification problems. In our research, we have applied the Genetic algorithm to reduce the size of the feature vector. The number of extracted feature in UCI Parkinson database [20] is 22 features for each recording. In our proposed method, every chromosome has 22 genes. The genes are the binary representation of primary features. We initiate the algorithm with 20 random-generated chromosomes. The fitness function used in this paper is the classification rate of AdaBoost as a classifier which diagnoses the PD. In each iteration of this algorithm, four of the best chromosomes, chromosomes with the higher classification rate, are chosen for the production and mutation of the next generation. In the next step, we have employed the Bagging algorithm as a classifier in order to evaluate the performance of this classifier in the detection of PD. In both experiments, the algorithm terminates when a fixed number of iteration has carried out which is 15 and 5 for AdaBoost and Bagging, respectively. The value of parameters is obtained by trial and error.

V. EXPERIMENTAL RESULTS

The database used in this paper was taken from UCI repository. This dataset was made by Max Little of Oxford University, in collaboration with National Centre for Voice and Speech [20]. It is made up of a variety of biomedical voice measurements of 31 people, 23 of whom with PD. It involves 195 recordings and 22 features for each of them are extracted. The main goal of this data is to distinguish healthy and patient people who suffer from PD; therefore, we face a binary classification problem.

We applied the genetic algorithm to the data available in UCI database to reduce the size of feature vector. By employing the AdaBoost as a classifier for detection of PD, the size of feature vector was reduced from 22 to just 7 features. In this case, the recognition rate was 98.28%. It can be seen that we have achieved a high classification rate by only using 7 features.

In the next experiment, we used the Bagging algorithm for classification purpose. By employing the Genetic algorithm, 10 features were selected as an optimal set of features and the classification rate was 96.55%. Then we applied the genetic algorithm on the optimal feature vector gained in the previous experiment in order to test if it is possible to reduce the number of features available in the classification process. In this case, by iterating the Genetic algorithm for 15 times, we again achieved 96.55% classification accuracy with just using 6 features. The results are shown in "Fig1".

As it is demonstrated in "Fig1" we have achieved high classification accuracy for detection of PDs just by employing 6 or 7 features out of 22 features available in UCI database.

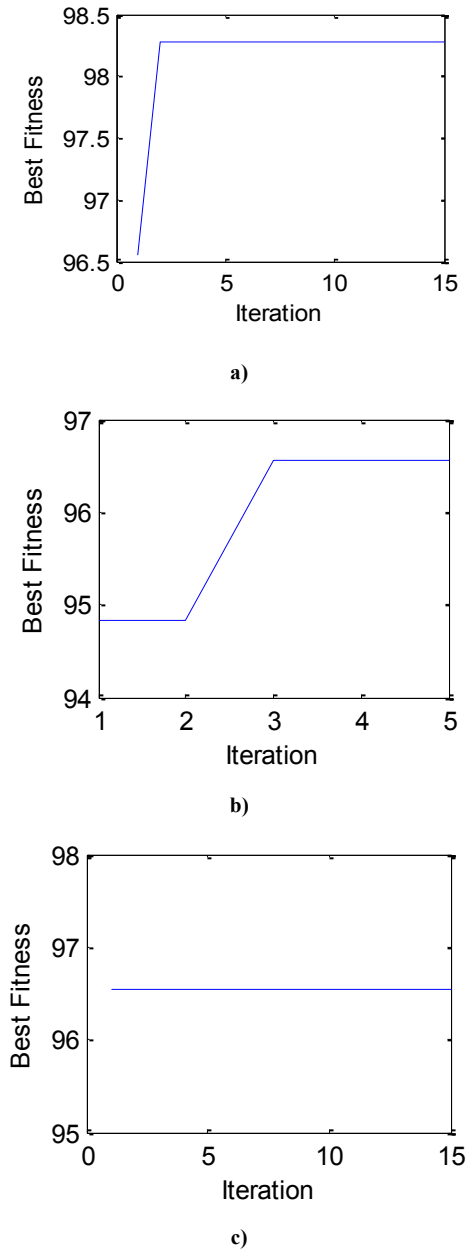


Figure 1. Best achieved detection accuracy after employing Genetic algorithm for a) AdaBoost classifier, b) First evaluation of Bagging algorithm, c) Second evaluation of Bagging algorithm

Table1 compares our proposed method with previously developed methods in the literature.

As Table1 shows, our proposed method is capable of selecting the most appropriate features which have a significant impact on classification accuracy. The fewer features can reduce the computational load. Also, the complexity of the predictive system can be decreased notably. Although our first

proposed method i.e. employing AdaBoost as a classifier, and the method developed in [8] use equal number features for classification process, our method outperforms the algorithm of [8]. To the best of our knowledge, the classification accuracy of our second method i.e. using the Bagging algorithm as a classifier is significantly higher than any other previously developed method.

Table I. Comparison of our proposed methods with other previous methods

Work	Methodology	Number of selected features	Accuracy (%)
[8]	Neural network + Genetic algorithm	8	79.93
[8]	Neural network + Binary Bat algorithm	6	93.60
[10]	Boosted Tree + multimodal features	22	95.08
[11]	KELM classifier + mRMR filter	15	94.19
The first proposed method	AdaBoost + Genetic algorithm	6	96.55
The second proposed method	Bagging + Genetic algorithm	7	98.28

VI. CONCLUSION

Parkinson's disease is a chronic disorder which causes a patient to experience more severity as the time progress. Therefore, early detection of PD has gained considerable attention. In previously proposed methods the voice recording replications have not been commonly addressed in an appropriate way for distinguishing of PD because the dependence nature of the data has usually been ignored. In this paper, we have proposed a method which can fill the gap between the feature selection and replicated data classification. In comparison with other methods, our proposed system can employ fewer features for classification and also achieve higher recognition rate. In addition to the reduction of computational load, this method can reduce the complexity of the predictive system.

REFERENCES

[1] L., Naranjo, C.J., Pérez, J., Martín, Y., Campos-Roca, "A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications", COMPUT METH PROG BIO, vol. 142, pp.147-156, 2017.

[2] H., Lee, L., Guan, I., Lee, "Video analysis of human gait and posture to determine neurological disorders", EURASIP J. Image Video PROC., vol. 1, pp. 380867, 2008.

[3] P., Huang, C., Harris, M.S., Nixon, "Human gait recognition in canonical space using temporal templates", IEEE Proceedings-Vision, Image and Signal Processing, vol. 146, pp. 93-100, 1999.

[4] L., Naranjo, C.J., Pérez, Y., Campos-Roca, J. Martín, "Addressing voice recording replications for Parkinson's disease detection", EXPERT SYST APPL, vol. 46, pp. 286-292, 2016.

[5] C.J., Pérez, L., Naranjo, J., Martín, Y., Campos-Roca, "A latent variable-based Bayesian regression to address recording replications in Parkinson's Disease", In: IEEE Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European, pp. 1447-1451, 2014.

[6] M., Novotný, J., Ruzs, R., Čmejla, E., Růžicka, "Automatic evaluation of articulatory disorders in Parkinson's disease", IEEE TRANS Audio Speech LANG PROC (TASLP), vol. 22, pp. 1366-1378, 2014.

[7] M., Hariharan, K., Polat, R., Sindhu, "A new hybrid intelligent system for accurate detection of Parkinson's disease", COMPUT METH PROG BIO, vol. 113, pp. 904-913, 2014.

[8] P., Shrivastava, A., Shukla, P., Vepakomma, N., Bhansali, K., Verma, "A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease", COMPUT METH PROG BIO, vol. 139, pp. 171-179, 2017.

[9] A., Sadikov, V., Groznik, M., Možina, J., Žabkar, D., Nyholm, M., Memedi, D., Georgiev, "Feasibility of spirometry features for objective assessment of motor function in Parkinson's disease", ARTIF INTELL MED, vol. 13, pp. 267-276, 2017.

[10] R., Prashanth, S.D., Roy, P.K., Mandal, S., Ghosh, "High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning", INT J MED INFORM, vol. 90, pp.13-21, 2016.

[11] H.-L., Chen, G., Wang, C., Ma, Z.-N., Cai, W.-B., Liu, S.-J., Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease", Neurocomputing, vol. 184, pp. 131-144, 2016.

[12] D., Joshi, A., Khajuria, P., Joshi, "An automatic non-invasive method for Parkinson's disease classification", COMPUT METH PROG BIO, vol. 145, pp. 135-145, 2017.

[13] Y.N., Jane, H.K., Nehemiah, K., Arputharaj, "A Q-backpropagated time delay neural network for diagnosing severity of gait disturbances in Parkinson's disease", J BIOMED INF, vol. 60, pp. 169-176, 2016.

[14] D.E., Goldberg, "Genetic algorithms in search, optimization, and machine learning", Reading: Addison-Wesley, 1989.

[15] W., Siedlecki, J., Sklansky, "A note on genetic algorithms for large-scale feature selection", PATTERN RECOGN LETT, vol. 10, pp. 335-347, 1989.

[16] Y., Freund, R.E., Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", In Springer European conference on computational learning theory, pp. 23-37, 1995.

[17] R.E. Schapire, "The boosting approach to machine learning: An overview", In Springer Nonlinear estimation and classification, pp. 149-171., 2003.

[18] L., Breiman, "Bagging predictors", MACH LEARN, vol. 24, pp.123-140, 1996.

[19] G., Liang, X., Zhu, Ch., Zhang, "An empirical study of bagging predictors for different learning algorithms", proceeding of the 25th AAAI conference on artificial intelligence, pp.1802-1803, 2011.

[20] M.A., Little, P.E., McSharry, S.J., Roberts, D.A., Costello, I.M., Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection", BIOMED ENG Online, vol. 6, pp. 23, 2007.