2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)
July 7-11, 2015. Busan, Korea

Pattern Recognition Methods for Multi Stage Classification of Parkinson's Disease Utilizing Voice Features

Wahyu Caesarendra*, Farika T Putri, Mochammad Ariyanto dan Joga D Setiawan

***Abstract*— A number of papers has presented a pattern recognition method for Parkinson's Disease (PD) detection. However, the literatures only able to classify subjects as either healthy of suffering from PD. This paper presents a pattern recognition method for multi stage classification of PD utilizing voice features. 22 features are obtained from University of California-Irvine (UCI) data repository. These features are extracted using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). It is found that PCA is better than LDA in terms of extracting significant features. Some classifiers such as Support Vector Machine (SVM), Adaptive Boosting (AdaBoost), K-Nearest Neighbor (KNN) and Adaptive Resonance Theory-Kohonen Neural Network (ART-KNN) are then used and compared. These methods are applied in multi stage classification. The classification results show that SVM has better testing accuracy than the other methods.**

I. Introduction

Recently, diseases which related to neurodegenerative are increasing. Parkinson's disease (PD) is one of the most common neurodegenerative disease associated with the evolving growth of the elderly population. PD is a neurological disorder related to dopamine deficiency in part of brain called substantia nigra, with four main symptoms such as slow movement (bradykinesia), muscle stiffness (rigidity), shaking (tremor), balance or walking problem and voice impairment [1].

In Indonesia, PD is one of the top ten most common illnesses in Rumah Sakit Ciptomangunkusumo (RSCM) [2]. Most of low income people with PD in Indonesia cannot afford the medical cost for PD detection such as MRI or EEG. Thus, most individuals with PD in Indonesia allow the disease to progress without any appropriate treatment or medication. This case resulted in worsening of symptoms that can affect the social life of patients and their families. This paper aimed to develop a diagnostic method for PD. Researchers have developed diagnostic method for PD by using a number of pattern recognition methods. These are only able to classify subjects as either healthy or suffering from PD [3, 4, 5]. Pattern recognition method for PD stage classification is important, since it help neurologist to provide appropriate treatment and medicine for patients.

Two scales are used for measuring the progression of PD: the Unified Parkinson's Disease Rating Scale (UPDRS) and the Hoehn and Yahr (H&Y) scale. UPDRS is used to assess the progression of the most important symptoms of PD such as bradykinesia, rigidity and tremor. The H&Y scale, focuses on the postural instability of individuals with PD [2]. This paper present a pattern recognition method for stage classification based on the H&Y scale.

To provide a low cost PD detection, this study presents PD detection based on pattern recognition method. Once the measurement and acquisition of raw voice data e.g. by using a microphone, pattern recognition method usually employs three main computational steps: (1) feature calculation, (2) feature extraction or reduction, and (3) classification. Feature calculation usually employed a number of voice features related to frequency component of the voice signal. It noted that not all features can distinguish the different between voice signal acquired from healthy people and from patient with PD (PWP). Therefore, to extract the significant features from a number of calculated features, some methods such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are used. Once the significant features obtained, some classifiers such as Support Vector Machine (SVM), Adaptive Boosting (AdaBoost), *K*-Nearest Neighbour (*K*NN) and Adaptive Resonance Theory-kohonen neural network (ART-KNN) are employed. SVM has been used in previous study [2].

II. Methods

A. KNN

The *K*NN classifies data for closest example of training in the features used. The method's aim is to define nearest neighbor of an unknown test pattern to determine its true class. The *K*NN method also known as instance-based learning [6]. There are traditional nearest neighbor rules [7]:

1. Define the k-NN from N training vectors, out of class label
2. Define vectors number k_i which belong to class c_i, i = 1, 2, …, l from samples k. Where $\Sigma\, k_i = k$.
3. Define x to the class c_i with the maximum number k_i of samples.

Lets assign training data set $T = \{(x_1, y_1), \ldots, (x_l, y_l)\}$ for set of vector examples $\mathbf{x}_i \in X \in R^n$ and hidden states $y_i \in Y = \{1, \ldots, c\}$. A ball centered in vector x which lie k example vectors $\mathbf{x}_i$, $i \in \{1, \ldots, l\}$, i.e., $|\{\mathbf{x}_i : \mathbf{x}_i \in R^n(\mathbf{x})\}| = k$ define as $R^n(\mathbf{x}) = \{\mathbf{x}' \| \mathbf{x} - \mathbf{x}'\| \leq r^2\}$. The k-nearest neighbor classification rule q: $X \rightarrow Y$ is defined as [7]:

$$q(\mathbf{x}) = \arg\max_{y \in Y} \; v(\mathbf{x}, y) \tag{1}$$

*W. Caesarendra, Farika T Putri, M. Ariyanto and Joga D Setiawan are with the Mechanical Engineering Department, Faculty of Engineering, Diponegoro University, Semarang 50275, Indonesia (corresponding author e-mail: w.caesarendra@gmail.com).

978-1-4673-9107-8/15/$31.00 ©2015 IEEE

where $v(\mathbf{x}, y)$ is number of prototype vectors $\mathbf{x}_i$ with hidden state $y_i = y$ which lie in the ball $\mathbf{x}_i \in R^n(\mathbf{x})$.

B. AdaBoost

Freund and Schapire introduced AdaBoost algorithm in 1995 which solved earlier boosting algorithm difficulties [8]. AdaBoost is a method which uses a set of learner of training on weighted training set. Let define a training set input $(x_i,y_i),\ldots,(x_m,y_m)$. Where x_i belong to domain X and y_i belong to data set Y. AdaBoost algorithm theory is distribution weight over training set maintaining. AdaBoost given a base learning algorithm in a round series $t = 1,\ldots T$. $D_t(i)$ is a notation for weight of training sample distribution on example i at round t [8].

The base learning algorithm is used to find a weak hypothesis $h_t : X \rightarrow \{-1, +1\}$. To determine whether weak hypothesis is good or not is by calculating the error as (2)

$$\in t = \Pr_{i \sim Dt}[h_t(x_i) \neq y_i] = \sum_{i : h_t(x_i) \neq y_i} D_t(i) \tag{2}$$

- Distribution D_t to train weak learner
- Weak hypothesis $h_t : X \rightarrow \{-1, +1\}$ with eror

$$\in t = \Pr_{i \sim Dt}[h_t(x_i) \neq y_i] \tag{3}$$

- Choose $\alpha_t = \frac{1}{2}\ln \frac{1 \in_t}{\in_t}$
- Update

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} \text{ if } h_t(x_i) = y_i \\ e^{\alpha_t} \text{ if } h_t(x_i) \neq y_i \end{cases} \tag{4}$$

$$= \frac{D_t(i)\exp(-\alpha_i y_i h_t(x_i))}{Z_t}$$

where Z_t is a normalization factor. The output final hypothesis is as follows

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{5}$$

AdaBoost choose the parameter α_t which measures the importance of h_t, if α_t become larger then ϵ_t become smaller.

C. ART-KNN

Application of artificial neural networks is used widely for classification. One type of neural network is ART-KNN. ART-KNN is combination of Adaptive Resonance Theory (ART) and Kohonen Neural Network (KNN). Evaluation of Euclidean distances weights between input vector X and each neuron of layer are done as similarity. The smallest weight becomes the winning neuron as in (6) [6]

$$\| B_J - X \| < \| B_j - X \| \; ; (j, J = 1, 2, \ldots, n; j \neq J) \tag{6}$$

where B_J is the weight of the winning neuron and B_j is the weight of j^{th} neuron in the layer. Input vector X reinserted in comparison layer, the similarity define as (7) [6]

$$S = \frac{\| B_J \| - \| B_J - X \|}{\| B_J \|} \tag{7}$$

The similarity S become smaller if the Euclidean distances between weight of neuron (B_J) and input vector (X) is getting larger. The criteria evaluation of similarity is defined using ρ parameter. The J^{th} cluster is define as similar to X when the similarity S is larger than ρ. The weight of J^{th} cluster is calculated as (8) to make the weight more accurate in cluster corresponding [6]

$$B_J = (n * B_{J0} + X)/(n + 1) \tag{8}$$

where B_{J0} is the origin weight, B_J elevated weight and n is changed time. But when the J^{th} is define much different to X the networks need a new neuron which weight define as (9) [6]

$$B_{n+1} = X \tag{9}$$

III. MATERIALS

One of the earliest indicators of the onset of the illness may be vocal impairment [9]. In this paper, voice data from the University of California-Irvine (UCI) data repository [10] are used. 22 voice features or measurement methods are collected from each individual volunteer. The volunteers include 8 healthy controls and 23 patients with PD (PWP). The details of the features measured can be found in [3] and summarised in Table 1.

TABLE I. LIST OF MEASUREMENT METHODS APPLIED TO ACOUSTIC SIGNALS FROM INDIVIDUAL VOLUNTEER

Feature	Description
MDVP: Fo (Hz)	Average vocal fundamental frequency [11].
MDVP: Fhi (Hz)	Maximum vocal fundamental frequency [11].
MDVP: Flo (Hz)	Minimum vocal fundamental frequency [11].
MDVP: Jitter (%)	Kay Pentax MDVP jitter as a percentage [11].
MDVP: Jitter (Abs)	Kay Pentax MDVP absolute jitter in microseconds [11].
MDVP: RAP	Kay Pentax MDVP Relative Amplitude Perturbation [11].
MDVP: PPQ	Kay Pentax MDVP five-point Period Perturbation Quotient [11].
MDVP: Shimmer	Kay Pentax MDVP local shimmer [11].
MDVP: Shimmer (dB)	Kay Pentax MDVP local shimmer in decibels [11].
MDVP: APQ	Kay Pentax MDVP 11-point Amplitude Perturbation Quotient [11].
Jitter: DDP	Average absolute difference of differences between cycles, divided by the average period [11].
Shimmer: APQ3	Three point Amplitude Perturbation Quotient [11].
Shimmer: APQ5	Five point Amplitude Perturbation Quotient [11].
Shimmer: DDA	Average absolute difference between consecutive differences between the amplitudes of consecutive periods [11].

Feature	Description
NHR	Noise-to-Harmonics Ratio [11].
HNR	Harmonics-to-Noise Ratio [11].
RPDE	Recurrence Period Density Entropy [12].
DFA	Detrended Fluctuation Analysis [12].
Spread1	Non-linear measure of fundamental frequency [13].
Spread2	Non-linear measure of fundamental frequency [13].
D2	Correlation dimension [14].
PPE	Pitch Period Entropy [3].

MDVP stands for (Kay Pentax) Multidimensional voice program. See [3] for detailed descriptions of the algorithm used to calculated these features.

Multi class classification (healthy controls and H&Y stages)

To date, pattern recognition method has been applied in PD data to classify two classes that are healthy and PWP. A list of 31 volunteers with sex, age information and H&Y stage of PD is presented in Table 2. Further, the data listed in Table 2 is classified into four classes: (1) healthy, (2) PWP with H&Y stage 1, (3) PWP with H&Y stage2, and (4) PWP with H&Y stage 3. These classifications are based on the measurements in [3]. The detail of the data used for multi class classification is presented in Table 3. These data are then separated for training and testing process in the classification step as shown in Section IV.

TABLE II. LIST OF VOLUNTEERS WITH SUBJECT CODE, SEX, AGE AND H&Y STAGE OF PD

PWP				Healthy control		
Subject code	***Sex***	***Age***	***H&Y Stage***	***Subject code***	***Sex***	***Age***
S01	M	78	3	S07	F	48
S02	M	60	2	S10	F	46
S04	M	70	2.5	S13	M	61
S05	F	72	3	S17	F	64
S06	F	63	2.5	S42	F	66
S08	F	48	2	S43	M	62
S16	M	62	2.5	S49	M	69
S18	M	61	2.5	S50	F	66
S19	M	73	1			
S20	M	70	3			
S21	F	81	1.5			
S22	M	60	1.5			
S24	M	73	2.5			
S25	M	74	3			
S26	F	53	2			
S27	M	72	2.5			
S32	M	50	1			
S33	M	68	2			
S34	F	79	2.5			
S35	F	85				
S37	M	76	1			
S39	M	64	2			
S44	M	67	1.5			

TABLE III. MULTI CLASS CLASSIFICATION DATA

Class of Subject	Subject Code
Healthy	S07, S10, S13, S17, S42, S43, S49, S50
H&Y Stage 1	S19, S32, S37
H&Y Stage 2	S02, S08, S26, S33, S39
H&Y Stage 3	S01, S05, S20, S25

"H&Y" refers to the Hoehn and Yahr PD stage, where higher values indicate greater level of disability

IV. RESULTS AND DISCUSSION

It has been investigated from previous study [2] that within 22 features some features can distinguish the different between healthy control and PWP and the others are unable. These 22 features were then extracted into 3 significant using PCA and LDA. In PCA, the significant features are principal component 1 (PC1), principal component 2 (PC2) and principal component 3 (PC3). In LDA, the significant features are linear discriminant 1 (LD1), linear discriminant 2 (LD2) and linear discriminant 3 (LD3). The results of PCA and LDA for four class classification are presented in Figs. 1(a) and (b).

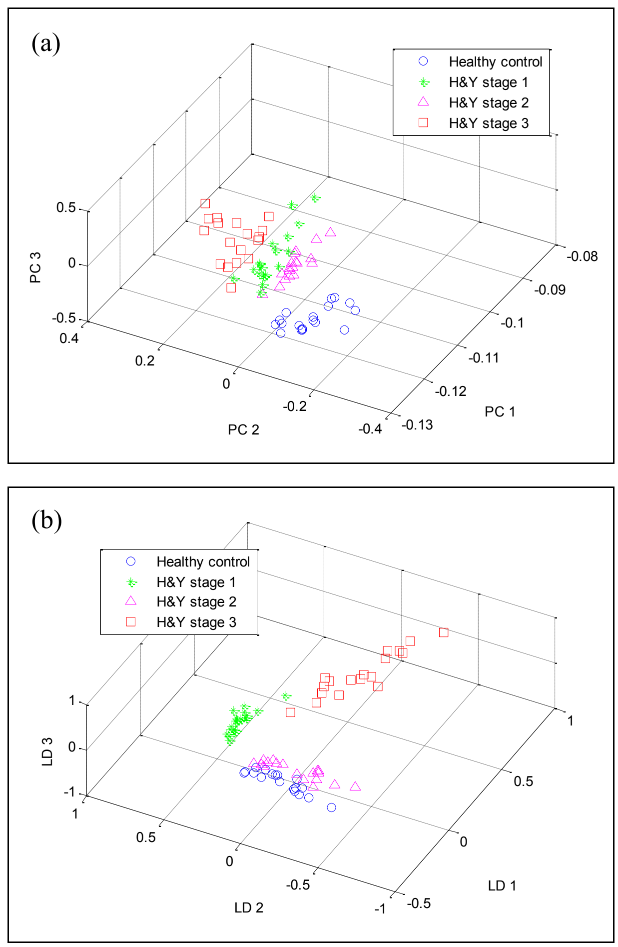

Figure 1. Features extraction results: (a) PCA; (b) LDA.

In classification, four classifiers are employed. They are SVM [15], KNN [15], ART-KNN [15], and AdaBoost [8]. These methods are selected based on the literature review presented in [16]. SVM has been studied previously for PD stage classification [2]. In this paper, H&Y stage 1, stage 2 and stage 3 were selected as the input data. The multi class classification data is presented in Table 3. The data in Table 3 is then divided into training and testing categories. It can be seen from Table 3 that the PD data sets for each stage is limited especially for H&Y stage 1. Therefore, In H&Y stage 1, the data sets S19, S32 and S37 were used for training and repeat the data set S19 for testing data. For H&Y stage 2 the allocations are: S02, S08 and S26 for training and S33 for testing. In H&Y stage 3, data sets S01, S05, S20 are used for training and data set S25 is used for testing. For healthy data, S07, S10, S13 and S17 were used for training and the remaining datasets were used for testing. As the number of training data is different with the number of testing data, the plotted data between training and testing will be different as well. The classification result for each classifier is presented as follows:

A. SVM

In SVM, pair of PC2 vs PC3 and LD1 vs LD3 were selected for input data of classification. These feature pairs were selected based on the distance between each feature. Self-minimum optimization SVM (SMO SVM) with Radial Basis Function (RBF) as a kernel function is used to train and test these pairs. The training and testing results for PCA and LDA are presented in Figs. 2 and 3, respectively.

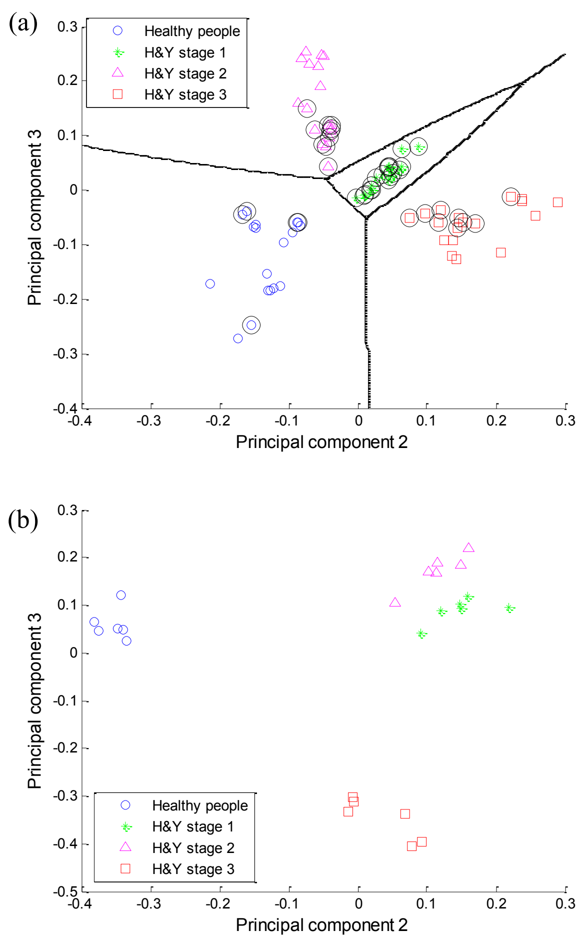

Figure 2. SVM classification for PCA: (a) Training; (b) Testing.

It can be seen from Figs. 2(a) and 3(a) that each class can be separated in different class in SVM training. The trained SVM is then used to classify the testing data and the result is presented in Figs. 2(b) and 3(b). The error is summarized in Table 4. The black circle marks indicates an area inside petagon hyperplane of SVM. It noted that this SVM result is obtained from previous study [2].

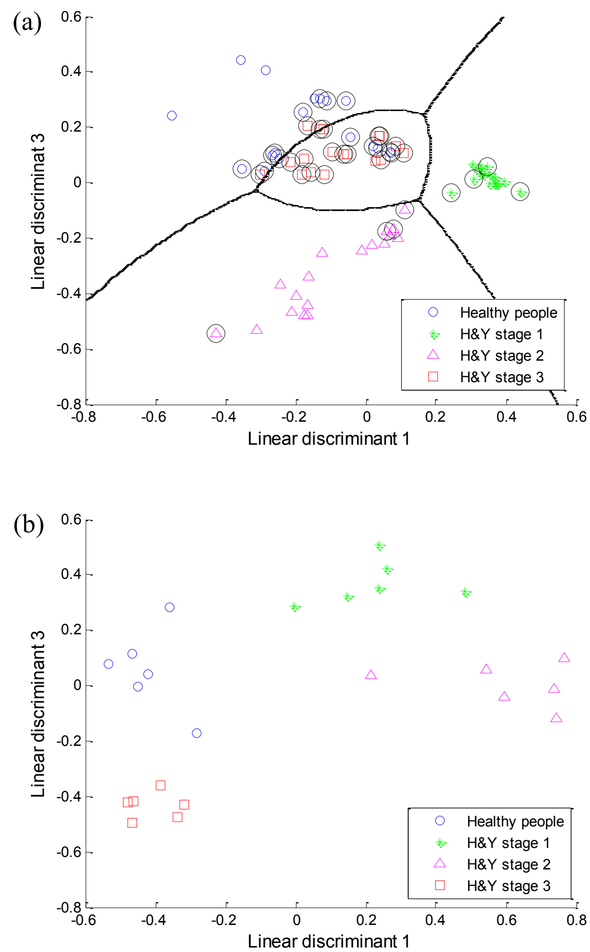

Figure 3. SVM classification for LDA: (a) Training; (b) Testing.

B. KNN

Similar training and testing data used in the SVM were also used for KNN classifier. The classification result of KNN is different to SVM and is presented in Fig. 4. The *x*-axis indicates the number of feature and *y*-axis is the class number. Where class 1 is healthy control, class 2 is H&Y stage 1, class 3 is H&Y stage 2 and class 4 is H&Y stage 3. The total number of features plotted in Fig. 4 is 24, where it is consist of 6 features for each class. Six features are obtained from 6 different voice data acquisition for each subject. It is shown in Fig. 4(a) that KNN can identify accurately for class 1 and 2. However, for class 3 to 4 most of features are classified at incorrect classes. In constras, for LDA, better classification is shown in class 1 and the remaining classes were misclassified. The classification result is presented in Table IV. The number of *k* was investigated from 3 to 15. It is suggest that the optimize k is 5.

The red circle indicates the training features and the blue star represents the testing features. If all testing features are fit into training features, the testing accuracy is 100%. The accuracy formula is shown in (10).

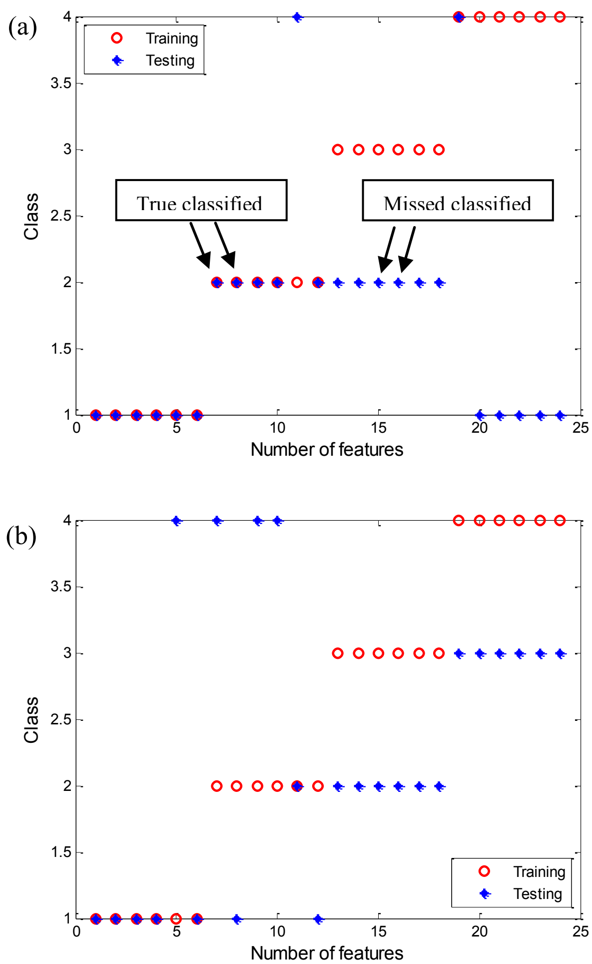

Figure 4. KNN classification: (a) PCA features; (b) LDA features.

C. AdaBoost

Similar to KNN, the 6 features which are extracted from PCA and LDA were used for AdaBoost classification. In this paper, multi class AdaBoost is employed. The classification result is presented in Fig. 5. It can be seen that some features are also missed classification to the other classes. The summarised of AdaBoost classification result is presented in Table IV. It is shown in Table IV that the classification accuracy of AdaBoost is better than KNN for both PCA and LDA features. In contrast, the classification accuracy of AdaBoost is lower than SVM for PCA feature; however the classification accuracy of LDA feature of AdaBoost is better than SVM. The accuracy is calculated by the following equation:

$$Accuracy = \frac{N_t \quad N_f}{N} \text{ x } 100\% \tag{10}$$

where N_t is number of true classification, N_f is number of false classification and N is total number of features.

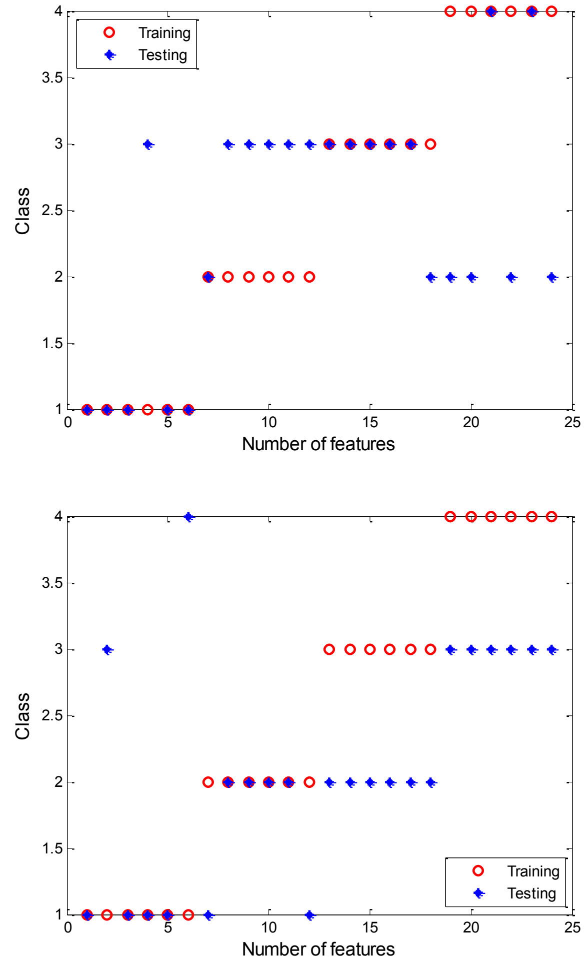

Figure 5. AdaBoost classification: (a) PCA features; (b) LDA features.

D. ART-KNN

Similar to KNN and AdaBoost, ART-KNN is applied to compare the classification accuracy of SVM. The classification result is presented in Fig. 6 and the summary of four classifiers is presented in Table IV.

The result of ART-KNN seems worsen than KNN and AdaBoost. It is shown that most of features are classified in incorrect classes.

TABLE IV. MULTI CLASS CLASSIFICATION RESULT

Classifier	Classification accuracy of PCA feature extraction (%)	Classification accuracy of LDA feature extraction (%)
SVM	79.17	29.17
KNN	50	25
AdaBoost	54.17	33.33
ART-KNN	25	29.17

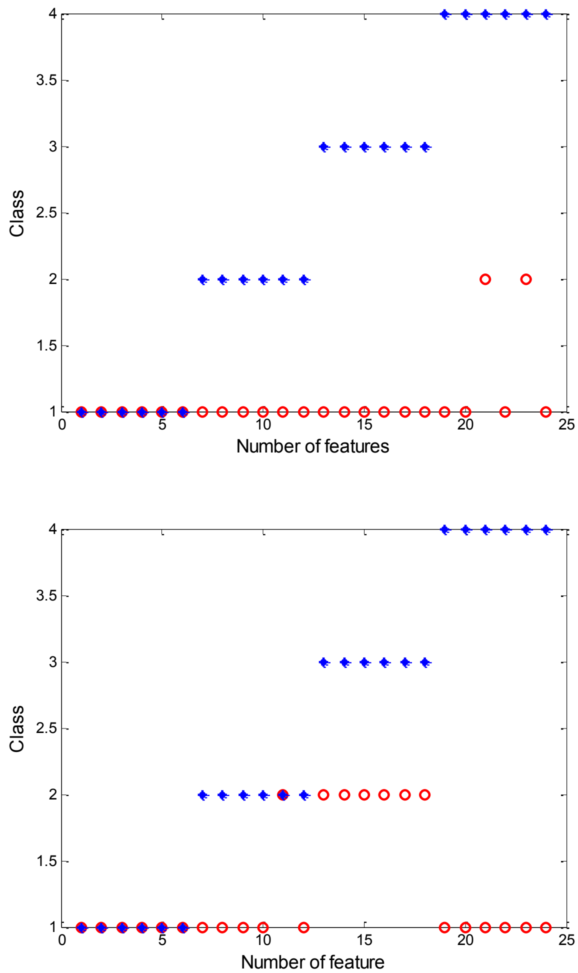

Figure 6. KNN classification: (a) PCA; (b) LDA.

V. Conclusion

The pattern recognition methods for multi stage classification of PD have been presented. The method consists of feature extraction and classification steps. The classification accuracy of SVM, AdaBoost and KNN of PCA features were higher than LDA features. This indicated that PCA can extract the 22 voice features better than LDA. In additions, SVM has better accuracy than KNN, AdaBoost and ART-KNN for PCA feature classification.

References

[1] M. Shahbakhti and D. Taherifar, "Linear and non linear speech features for detection of Parkinson's disease," *Biomedical Engineering International Conference*, 978-1-4799-1466-1, 2013.

[2] W. Caesarendra, M. Ariyanto, J.D Setiawan, M. Arozi and C.R. Chang, "A Pattern Recognition Method for Stage Classification of Parkinson's Disease Utilizing Voice Features," *In: Proceeding of Second International Conference on Biomedical Engineering (IECBES), Miri, Malaysia, December 2014.*

[3] 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, April 2009.

[4] 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, pp. 904–913, 2014.

[5] R. Das, "A comparison of multiple classification methods for diagnosis of Parkinson disease," *Expert Syst. Appl.*, vol. 37, pp. 1568–1572, 2010.

[6] R. G. Ramani and G. Sivagami, "Parkinson Disease classification using data mining algorithm," *International Journal of Computer Applications*, vol. 32, no. 9, 2011.

[7] B.S. Yang, A. Widodo, Introduction of Intelligent Machine Fault Diagnosis and Prognosis, Nova Science Publishers, NY, 2009.

[8] Y. Freund and R. E. Schapire, "A short introduction to boosting," Journal of Japanese Society for Artificial Intelligent, vol. 14, 1999.

[9] J.R. Duffy, Motor Speed Disorders: Substrates, Differential Diagnosis, and Management, 2nd ed. St. Louis, MO: Elsevier, 2005

[10] https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring (accessed on July 2014)

[11] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," *Glot Int.*, vol. 5, pp. 341–345, 2001.

[12] 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, p. 23, 2007.

[13] I. Rustempasic, M. Can, "Diagnosis of Parkinson's disease using fuzzy C-means clustering and pattern recognition," *Southeast Eur. J. Soft Comput.*, vol. 2, p. 421-49, year 2013.

[14] H. Kantz and T. Schreiber, Nonlinear Time Series Analysis, New ed. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[15] B.S. Yang, A. Widodo, Introduction of Intelligent Machine Fault Diagnosis and Prognosis, Nova Science Publishers, NY, 2009.

[16] X. Wu et al., "Top ten algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, p. 14:1-37, year 2008.