COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 113 (2014) 904–913

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

A new hybrid intelligent system for accurate detection of Parkinson's disease

M. Hariharan[a,*], Kemal Polat[b], R. Sindhu[c]

[a] School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP), 02600, Campus Pauh Putra, Perlis, Malaysia
[b] Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Abant Izzet Baysal University, 14280 Bolu, Turkey
[c] School of Microelectronic Engineering, Universiti Malaysia Perlis (UniMAP), 02600, Campus Pauh Putra, Perlis, Malaysia

ARTICLE INFO

Article history:
Received 25 June 2013
Received in revised form
26 December 2013
Accepted 2 January 2014

Keywords:
Parkinson's disease
Dysphonia features
Feature weighting
Feature selection
Classification

ABSTRACT

Elderly people are commonly affected by Parkinson's disease (PD) which is one of the most common neurodegenerative disorders due to the loss of dopamine-producing brain cells. People with PD's (PWP) may have difficulty in walking, talking or completing other simple tasks. Variety of medications is available to treat PD. Recently, researchers have found that voice signals recorded from the PWP is becoming a useful tool to differentiate them from healthy controls. Several dysphonia features, feature reduction/selection techniques and classification algorithms were proposed by researchers in the literature to detect PD. In this paper, hybrid intelligent system is proposed which includes feature pre-processing using Model-based clustering (Gaussian mixture model), feature reduction/selection using principal component analysis (PCA), linear discriminant analysis (LDA), sequential forward selection (SFS) and sequential backward selection (SBS), and classification using three supervised classifiers such as least-square support vector machine (LS-SVM), probabilistic neural network (PNN) and general regression neural network (GRNN). PD dataset was used from University of California-Irvine (UCI) machine learning database. The strength of the proposed method has been evaluated through several performance measures. The experimental results show that the combination of feature pre-processing, feature reduction/selection methods and classification gives a maximum classification accuracy of 100% for the Parkinson's dataset.

1. Introduction

Parkinson's disease is a progressive neurodegenerative disorder which can be characterized by several indications like tremor, rigidity and slowness of movements. PWP may have difficulty in walking, talking or completing other simple tasks [1–3]. According to the statistics by Parkinson's disease foundation, it is estimated that seven to 10 million people are living with PD worldwide [4]. In next 25 years, the number of PWP is expected to increase due to the raise in proportion of elderly people [1–3]. Age is the most important risk factor for the onset of PD as the incidence of PD increases with age. With the new and effective medications of PD, several improvements are possible to enhance PWP's quality of life. Researchers have proposed various non-invasive methods to

* Corresponding author. Tel.: +60 134622469; fax: +60 49885167.
E-mail addresses: hari@unimap.edu.my, wavelet.hari@gmail.com (M. Hariharan).
0169-2607/$ – see front matter
http://dx.doi.org/10.1016/j.cmpb.2014.01.004

detect the severity of PD using acoustic analysis of voice signal, physiological signals, wearable sensors and gait analysis etc. Among them, detecting PD progression using acoustic analysis of voice signal has drawn significant attention [5–11].

Little and his co-researchers have proposed different non-linear features using non-linear dynamics for the objective analysis of voice signals of PWP and healthy controls [9–11]. These methods could help to improve the existing methods through better representation of voice signals. Little et al. have conducted substantial research using the non-linear and several conventional pitch/amplitude perturbation based features for the discrimination of the voice signals of PWP and healthy controls. The success rate was 99% using the best selected features through support vector machine based classifier. Several studies have been conducted using Little's PD dataset in the past three years and they have achieved the accuracy from 75.03% to 100% using various feature selection, feature pre-processing and classification algorithms [12–21]. For instance, Kemal Polat has applied a feature pre-processing method based on fuzzy *c*-means clustering for the classification of PD's and obtained a maximum accuracy of 96% [22] with *k*-nearest neighbor classifier.

In this paper, a new feature weighting method using Model-based clustering (Gaussian mixture model) was suggested to enrich the discriminative ability of the features (dysphonia features). In order to reduce/select the best dysphonia features, two projection based feature reduction techniques (PCA and LDA) and two step-wise feature subset selection techniques (SBS and SFS) were used. Two validation schemes were used such as 10-fold cross validation and conventional validation to demonstrate the efficacy of the proposed method. LS-SVM, PNN and GRNN were used as classifiers to discriminate the voice signals of PWP and healthy controls. From the experimental results, it can be concluded that the proposed hybrid intelligent system gives excellent classification accuracy of 100% using weighted dysphonia features.

The organization of the paper is as follows: Section 2 presents the dataset description, fundamentals of feature weighting and feature reduction/selection methods. Section 3 describes the review of LS-SVM PNN and GRNN. Experimental results are presented in Section 4 and the results are discussed in Section 5. Section 6 concludes the paper.

2. Materials and methods

To evaluate the effectiveness of the proposed method, PD dataset has been taken from UCI machine learning database, which consists of both conventional and non-linear dysphonia features (22 raw features – RF) [9–11,23]. This dataset was created by Little and his colleagues in collaboration with 10 medical centers in US [9–11,23]. The dataset consists of features extracted from the speech samples of 31 people (23 with Parkinson's disease + 8 normal). An average of six sustained phonations was recorded from each subject, ranging between 1 s and 36 s in length. 48 sustained phonations from 8 normal people and 147 sustained phonations from 23 PD patients were used to extract the RF and prepare the PD dataset. Table 1 shows the list of dysphonia features used in this study.

Table 1 – List of dysphonia features [8–11].

MDVP: Fo(Hz), Fhi(Hz), and Flo(Hz)	Average, maximum and minimum vocal fundamental frequencies (3 features)
MDVP: Jitter (%), Jitter (Abs), RAP, PPQ and DDP	Measures of variation in fundamental frequency (5 features)
MDVP: Shimmer, Shimmer (dB), APQ3, APQ5, APQ and DDA	Measures of variation in amplitude (6 features)
NHR and HNR	Measures of ratio of noise to tonal components in the voice (2 features)
RPDE and D2	Non-linear dynamical complexity measures (2 features)
DFA	Signal fractal scaling exponent (1 feature)
Spread 1, Spread2, and PPE	Non-linear measures of fundamental frequency variation (3 features)

2.1. *Proposed method*

Classification of PWP and healthy controls is a typical pattern classification problem. The proposed hybrid intelligent system include feature pre-processing using Gaussian mixture model based feature weighting, feature reduction/selection using PCA, LDA, SFS and SBS and feature classification using LS-SVM, PNN and GRNN. Fig. 1 illustrates the block diagram of the proposed hybrid intelligent system. In this present work, several experiments were performed using 22 original raw features and weighted features.

2.2. *Feature weighting using Model-based clustering (Gaussian mixture model)*

Feature weighting is one of the pre-processing techniques in any pattern classification problem. These techniques have been applied by the researchers either to discard the irrelevant features or to improve the discrimination ability of the features. The performance of any classifier always depends on the relevant and robust features. Several clustering based algorithms have been proposed for feature weighting. In this work, Model-based clustering (Gaussian mixture model) approach was proposed to improve the robustness of the dysphonia features. Generally, Gaussian mixture model (GMM) was used as classification model in different pattern recognition

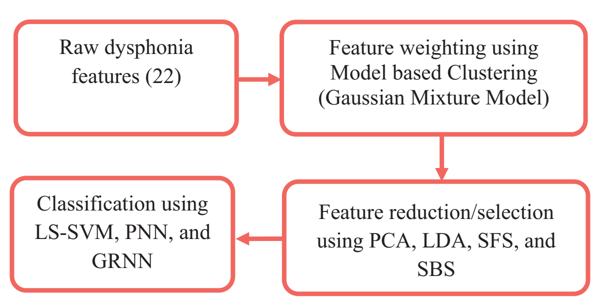


Fig. 1 – Block diagram of the proposed hybrid intelligent system (feature weighting, feature reduction/selection and classification).

applications [24–29]. GMM based feature reduction was proposed in [29] for the diagnosis of vocal fold pathology. Different applications of GMM [24–29] motivate us to suggest GMM based feature weighting. In a Model-based approach, certain models are used for clusters and attempting to optimize the fit between the data and model. Each cluster can be mathematically represented by a Gaussian (parametric) distribution. The entire dataset x is modeled by a weighted sum of M number of mixtures of Gaussian component densities and is given by the equation

$$p(i|x) = \sum_{i=1}^{M} w_i p(x|\mu_i, \Sigma_i) \tag{1}$$

where x is a N-dimensional continuous valued dysphonia features, w_i, $i = 1, 2, 3, \ldots, M$ are the mixture weights, and $p(x|\mu_i, \Sigma_i)$, $i = 1, 2, 3, \ldots, M$ are the component Gaussian densities. Each component density is an N-variate Gaussian function of the form:

$$p(x|i) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} \exp^{-1/2(x-\mu_i)^T \sum_i^{-1}(x-\mu_i)} \tag{2}$$

where μ_i is the mean vector and Σ_i is the covariance matrix. The mixture weights should satisfy the constraint that $\sum_{i=1}^{M} w_i = 1$. By combining mean vectors, covariance matrices and mixture weights Gaussian mixture model is parameterized and they are frequently used for data clustering. Clusters are assigned by selecting the component that maximizes the posterior probability [24–26,30]. Gaussian mixture modeling uses an iterative expectation maximization (EM) algorithm that converges to a local optimum and assigns posterior probabilities to each component density with respect to each observation. Gaussian mixture model based clustering may be more appropriate than k-means clustering and it is sometimes considered as a soft clustering method. The posterior probabilities for each point indicate that each data point has some probability of belonging to each cluster [24–26,30]. The working of clustering based feature weighting is summarized (Fig. 2) as follows: firstly, the component means (cluster centers) of each feature belonging to dataset using GMM based clustering method was found. Next, the ratios of means of features to their centers were calculated. Finally, these ratios were multiplied with each respective feature. Fig. 2 shows the block diagram of the proposed Model-based clustering (Gaussian mixture model) for feature pre-processing/weighting. After applying GMM clustering based feature weighting method, the raw features were known as weighted features (WF).

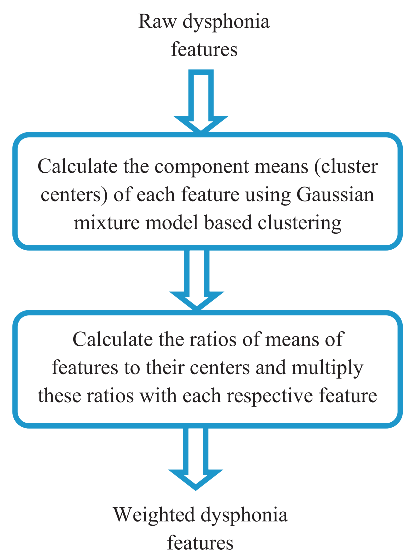


Fig. 2 – The block diagram of proposed GMM clustering based feature weighting method.

2.3. Feature reduction and selection

In order to improve the generalization performance and reduce the complexity of the learning phase of any classifier, feature reduction/selection is usually applied before the classification process. In this work, two projection based feature reduction methods (PCA and LDA) and two stepwise feature subset selection (SFS and SBS) methods were used. The brief description of the feature reduction and selection methods is given as follows:

2.3.1. *Projection based feature reduction methods*

PCA is a well-known linear projection method and also an unsupervised feature transformation method since it performs the vector projection without any knowledge of their class labels. PCA reveals about the hidden information from the original feature space by maximizing the variance of the projected vectors [31–33]. A N-dimensional weighted dysphonia features x_i ($i = 1, 2, 3, \ldots, m$, $N < m$) was projected on the Eigenvectors of its covariance matrix and transformed matrix (v) was obtained as follows

$$v = U^T x_i \tag{3}$$

From the transformed matrix (v), the features of Eigenvalue greater than ‘1’ alone were selected and they form the subset of 5 uncorrelated features.

LDA is a supervised projection method and maximizes the ratio of the between and within class scatters of the feature set as shown in the following equation [31,34,35].

$$Y_{opt} = arg \max_{y} \frac{|Y^T S_b Y|}{|Y^T S_w Y|} = [y_1, y_2, \ldots, y_P] \tag{4}$$

where $\{y_i | 1 \le i \le P\}$ are the LDA subspace base vectors, P is the dimension of the subspaces. Using the transformation matrix Y, the between-class scattering (S_b) was maximized whereas the within-class scattering (S_w) was minimized and hence LDA seeks to reduce dimensionality while preserving as much of the class discriminative power as feasible. In this work, LDA was employed to map the twenty two weighted dysphonia feature space into a one-dimensional feature space (1 feature) based on the criterion given in Eq. (4).

2.3.2. *Stepwise feature subset selection*

Among the 22 weighted dysphonia features, SBS and SFS were applied to find the most informative/discriminative features. Both these methods form the best feature subset by adding or removing each feature based on discrimination ability, which has been found through an objective function. In this work, 1-nearest neighbor leave-one-out classification performance was used as the objective function [36] which has the highest ability to evaluate each feature subset to increase the detection rate of Parkinson's disease.

SBS method starts from the full set of 22 weighted dysphonia features and sequentially eliminates the worst features based on the lowest value of the objective function while retaining the best features based on the highest value of the objective function. The procedure was repeated until the best feature subset was obtained. SFS works in the opposite direction of SBS. SFS method starts from an empty set and sequentially adds the best features based on the highest value of the objective function [37]. Although SFS and SBS are simple search techniques, they do not always achieve the best solution. The main limitation of both these methods is their inability to re-evaluate the usefulness of the features, after it has been added or discarded [37]. Feature selection algorithm was implemented by using MATLAB pattern recognition toolbox [38].

3. Classification of normal and PWP

In this work, three supervised classifiers were used for the classification of healthy controls and PWP. The fundamentals of the suggested classifiers were reviewed in the following subsections.

3.1. *Support vector machine*

SVM was originally proposed for solving two-class classification problems and extended for multi-class classification problems too. It is a promising method for solving nonlinear classification problems, function estimation and pattern recognition tasks [39–46]. First SVM maps raw/weighted dysphonia features into a high dimensional feature space using a certain kernel function. Using quadratic optimization, an optimal hyper plane was found and the margin of separation between the classes is maximized. RBF kernel function was used in our work. Best values for RBF kernel such as regularization parameter (γ, *gam*) and σ^2 (sig2, squared bandwidth) were found through experiments to obtain higher accuracy. In this work, LS-SVMLab toolbox [39] was used to perform the classification of healthy controls and PWP.

3.2. *Probabilistic neural network and general regression neural network*

PNN and GRNN architectures were proposed by Donald F. Specht to perform classification and general linear regression respectively [47]. These networks have several advantages than multilayer perceptron (MLP) network such as much faster learning, more accurate than MLP and relatively insensitive to outliers [31,48–50]. Both PNN and GRNN comprises of four layers which includes input layer, pattern layer, summation layer and output layer. The output variable is categorical for PNN and continuous for GRNN. The accuracy of PNN and GRNN highly depends on suitable smoothing parameter or spread factor (η) [47,31,48–54]. Appropriate η value was found between 0.01 and 0.1 through experimental investigations.

4. Results

GMM based clustering method was proposed to enhance the discriminatory power of the raw dysphonia features. In this work, two projection based feature reduction techniques (PCA and LDA) and two step-wise feature searching techniques (SBS and SFS) were used to reduce/select the best weighted dysphonia features. Three supervised classifiers and two validation schemes (conventional and 10-fold cross validation) were employed to gauge the strength of the proposed method. In 10-fold cross validation, the raw/weighted dysphonia feature set was divided randomly into 10 sets and training and testing were repeated for 10 times. Overall accuracy was obtained from the average of the 10 iterations. In the conventional validation method, 50% of the data (74 abnormal + 24 normal) was used for training and remaining 50% (73 abnormal + 24 normal) for testing. To gauge the classifiers' performance, five different performance measures namely precision, sensitivity, specificity, overall accuracy and area under curve (AUC) from receiver operating characteristic curve (ROC) were considered. These performance measures were calculated from true positive (TP, the classifier classified as pathology when pathological samples are present), true negative (TN, the classifier classified as normal when normal samples are present), false positive (FP, the classifier classified as pathological when normal samples are present), and false negative (FN, the classifier classified as normal when pathological samples are present). The five performance measures were calculated using the Eqs. (5)–(9).

$$\text{Precision(PR)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{Sensitivity(SE)} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{6}$$

$$\text{Specificity(SP)} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \tag{7}$$

$$\text{Overall accuracy(ACC)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{8}$$

$$\text{Area under curve(AUC)} = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}}\right) \tag{9}$$

All the simulations were conducted in MATLAB 2011 with Intel Core-i7, 2.20 GHz CPU and 4 GB RAM.

4.1. *Analysis of features*

Fig. 3(a) and (b) shows the class distribution of raw and weighted features. The class distribution of raw and weighted PCA features was shown in Fig. 4(a) and (b). Fig. 5(a) and (b) depicts the class distribution of the raw and weighted LDA feature. From Figs. 3(a), 4(a) and 5(a) it can be inferred that there is a higher degree of overlap between two classes (raw, raw PCA and raw LDA). After applying GMM based feature weighting, this degree of overlap was minimized and therefore

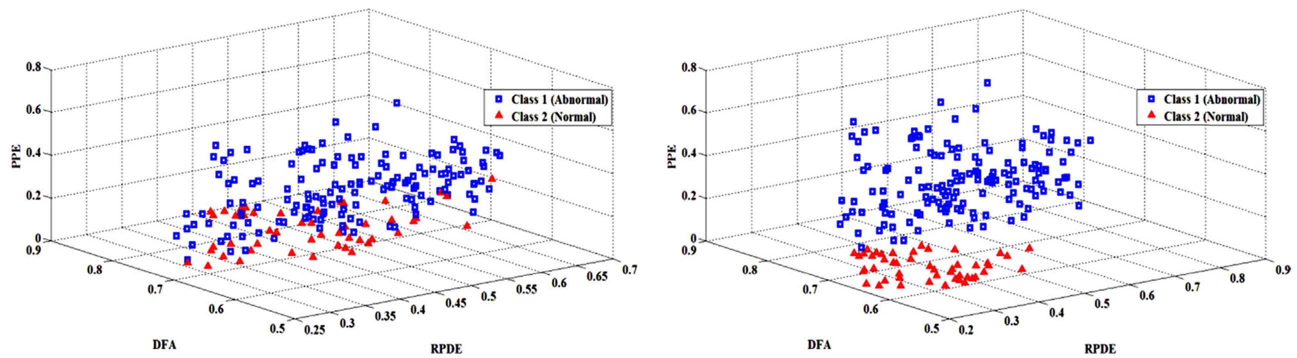
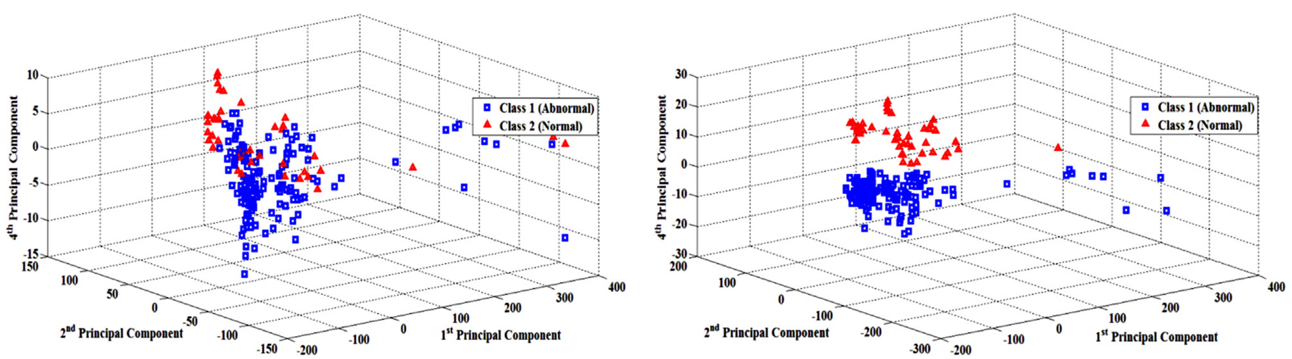


Fig. 3 – (a) Class distribution of raw features according to RPDE, DFA & PPE. (b) Class distribution of weighted features according to Feature RPDE, DFA & PPE.

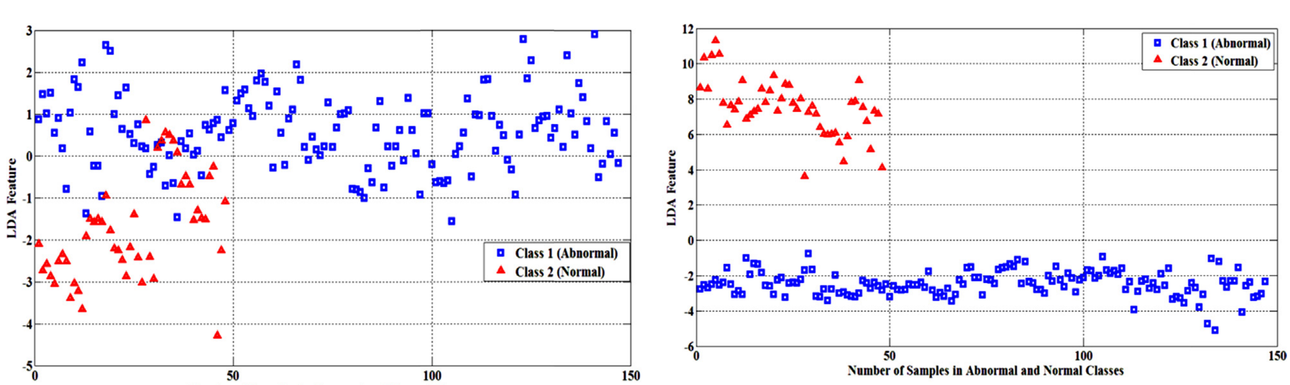


(a) (b)

Fig. 4 – (a) Class distribution of raw PCA features. (b) Class distribution of weighted PCA features.

the discrimination ability of features was greatly enhanced (Figs. 3(b), 4(b) and 5(b)). The discriminating power of the raw and weighted features was also analyzed using F-score technique and one-way of analysis of variance (ANOVA). Generally, F-score is used to measure the discriminatory power of two sets of real numbers (raw and weighted features) and it has been depicted in Fig. 6. Weighted dysphonia features have the larger F-scores than raw dysphonia features, as they are more discriminative. From the ANOVA test, F-value was recorded and shown in Fig. 7. F-value is the ratio of between group mean square and within group mean square. From Fig. 7, it can be found that the ANOVA–F-values of weighted dysphonia features were larger compared to raw dysphonia features.

The classification results using 10-fold cross validation and conventional validation were tabulated in Table 2. Average classification accuracy of above 95% was attained for the original twenty-two raw features, which is closer to the results published in the literature. After applying model based clustering as feature weighting method, the average classification accuracy was improved to 100% for all the three supervised classifiers used in this work. 5 uncorrelated features from PCA and 1 feature having good class separabilty from LDA were obtained. After applying feature reduction (PCA and LDA), no



(a) (b)

Fig. 5 – (a) Class distribution of raw LDA feature. (b) Class distribution of weighted LDA feature.

Table 2 – Classification results of raw and weighted features for Parkinson's dataset.

Classifier	Feature reduction/ selection	10-fold cross validation						Conventional validation					
		PR	SE	SP	ACC	AUC	Time(s)	PR	SE	SP	ACC	AUC	Time(s)
LS-SVM	22 RF	97.48 ± 0.56	96.44 ± 0.80	92.05 ± 1.63	95.38 ± 0.76	0.94 ± 0.06	0.0043 ± 0.0007	94.93 ± 3.36	93.67 ± 3.18	84.90 ± 8.29	91.24 ± 2.39	0.89 ± 0.04	0.0029 ± 0.0005
	22 WF	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	0.0038 ± 0.0006	100.00 ± 0.00	99.86 ± 0.43	100.00 ± 0.00	99.90 ± 0.33	1.00 ± 0.00	0.0027 ± 0.0004
	WF + PCA	100.00 ± 0.00	99.87 ± 0.42	100.00 ± 0.00	99.90 ± 0.32	1.00 ± 0.00	0.0042 ± 0.0006	100.00 ± 0.00	99.20 ± 1.29	100.00 ± 0.00	99.38 ± 1.00	1.00 ± 0.01	0.0030 ± 0.0009
	WF + LDA	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	0.0034 ± 0.0004	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	0.0025 ± 0.0004
	WF + SBS	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	0.0035 ± 0.0004	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	0.0027 ± 0.0007
	WF + SFS	97.82 ± 0.54	100.00 ± 0.00	93.77 ± 1.42	98.36 ± 0.40	1.00 ± 0.00	0.0035 ± 0.0005	98.08 ± 1.47	100.00 ± 0.00	94.64 ± 4.06	98.56 ± 1.11	0.97 ± 0.02	0.0027 ± 0.0008
PNN	22 RF	95.99 ± 0.88	97.99 ± 0.87	88.47 ± 2.29	95.49 ± 1.02	0.93 ± 0.06	1.0304 ± 0.0101	94.71 ± 3.36	95.55 ± 3.18	84.92 ± 8.29	92.61 ± 2.39	0.90 ± 0.04	1.2269 ± 0.0005
	22 WF	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	0.9817 ± 0.0188	100.00 ± 0.00	99.92 ± 0.17	100.00 ± 0.00	99.94 ± 0.13	1.00 ± 0.00	1.2343 ± 0.0080
	WF + PCA	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	1.0053 ± 0.0304	99.89 ± 0.11	99.38 ± 0.34	99.68 ± 0.32	99.43 ± 0.28	1.00 ± 0.00	1.2230 ± 0.0041
	WF + LDA	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	0.9955 ± 0.0243	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	1.2190 ± 0.0050
	WF + SBS	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	0.9704 ± 0.0055	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	1.2234 ± 0.0097
	WF + SFS	98.84 ± 1.02	100.00 ± 0.00	96.66 ± 2.93	99.13 ± 0.77	1.00 ± 0.00	0.9930 ± 0.0255	98.90 ± 0.94	99.95 ± 0.07	96.91 ± 2.64	99.13 ± 0.67	0.98 ± 0.01	1.2222 ± 0.0050
GRNN	22 RF	95.85 ± 0.93	98.13 ± 1.01	88.20 ± 2.19	95.49 ± 0.79	0.89 ± 0.08	1.0047 ± 0.0034	94.71 ± 0.61	95.57 ± 1.64	84.91 ± 1.23	92.63 ± 1.19	0.90 ± 0.01	1.3010 ± 0.0311
	22 WF	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	0.9561 ± 0.0038	100.00 ± 0.00	99.35 ± 0.57	100.00 ± 0.00	99.49 ± 0.45	1.00 ± 0.00	1.3076 ± 0.0118
	WF + PCA	100.00 ± 0.00	99.87 ± 0.42	100.00 ± 0.00	99.90 ± 0.32	1.00 ± 0.00	0.9590 ± 0.0047	99.95 ± 0.10	98.95 ± 0.37	99.83 ± 0.29	99.14 ± 0.29	0.99 ± 0.00	1.2754 ± 0.0274
	WF + LDA	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	0.9700 ± 0.0206	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	1.2895 ± 0.0286
	WF + SBS	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	0.9573 ± 0.0031	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	1.00 ± 0.00	1.3003 ± 0.0307
	WF + SFS	98.84 ± 1.02	100.00 ± 0.00	96.66 ± 2.93	99.13 ± 0.77	1.00 ± 0.00	0.9568 ± 0.0072	99.77 ± 0.27	99.88 ± 0.12	99.34 ± 0.75	99.73 ± 0.26	1.00 ± 0.00	1.2918 ± 0.0281

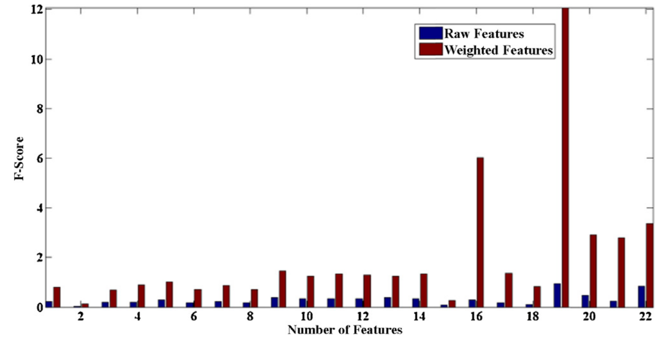


Fig. 6 – Comparison of *F*-score between raw and weighted features.

significant changes were inferred from the performance of the classifiers. Four best weighted dysphonia features such as spread 1, spread 2, D2 and PPE were found from SBS. From SFS, Jitter (%), Jitter (Abs), RAP and APQ were obtained as best weighted dysphonia features. The performance of the classifiers using selected weighted dysphonia features was slightly lower than using all the weighted dysphonia features. However, the classification accuracy was better when compared to those in the literature. Though, the performance of the classifiers during conventional validation was also same, there is higher variations in the classification during some experiments. From the experimental results, we can conclude that the proposed method offers 100% accuracy in detecting PWP using the weighted dysphonia features. To the best of our knowledge, there is no previous work on GMM based feature weighting in the detection of PWP.

5. Discussions

The aim of this paper is to investigate GMM based clustering as feature weighting method to enhance the discriminatory power of the raw dysphonia features. The weighted dysphonia features were fed to three supervised classifiers and the classification accuracy was improved upto 100%. Table 3 presents the comparison of our results with some of the significant works available in the literature for PD dataset. Classification of healthy controls and PWP using traditional and non-linear measures along with kernel support vector

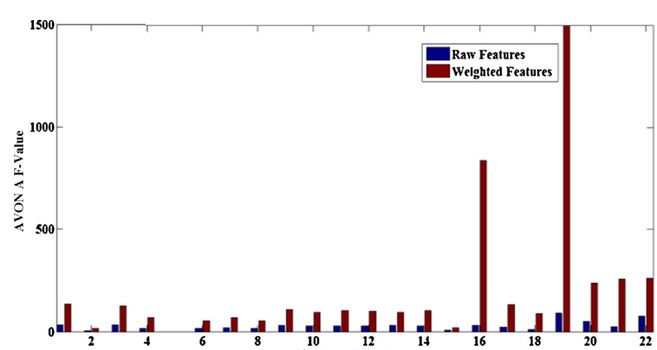


Fig. 7 – Comparison of ANOVA-*F*-value between raw and weighted features.

Table 3 – Comparison of our proposed method with the previous works in the literature.

Method	Accuracy (%)	Method	Accuracy (%)
Pre-selection filter + exhaustive search + SVM [9]	91.4 (bootstrap with 50 replicates)	Particle swarm optimization + OPF [21]	73.53 (hold-out)
Dirichlet process mixtures [55]	87.7 (5-fold CV)	Harmony search + OPF [21]	84.01 (hold-out)
ANN [56]	92.9 (hold-out)	Gravitational search algorithm + OPF [21]	84.01 (hold-out)
Mutual information based feature selection + SVM [57]	92.75 (bootstrap with 50 replicates)	Parallel NN [58]	91.20 (hold-out)
Improved mRVMs [59]	89.47 (10-fold CV)	PCA-FKNN [12]	96.07 (average 10-fold CV)
GP-EM [60]	93.1 (10-fold CV)	RF ensemble of IBk (a *k*-nearest neighbor variant) algorithm [19]	97
CFS-RF [20]	87.1 (10-fold CV)	Haar wavelets as projection filter, linear logistic regression [18]	100 (test data)
Fuzzy-based non-linear transformation + SVM [15]	93.47 (hold-out)	Multinomial logistic regression classifier with Haar wavelets transformation as projection filter [17]	100 (test data)
Fuzzy entropy measures + Similarity classifier [16]	85.03 (hold-out)	PSO-FKNN	97.47 (10-fold CV)
Our proposed method			
GMM based feature weighting + LDA + LS-SVM or PNN or GRNN			100 (average 10-fold CV)
GMM based feature weighting + SBS + LS-SVM or PNN or GRNN			100 (50% test data)

machine was carried out. Through an exhaustive search of all possible combinations of these measures [9], they have obtained 91.4% accuracy after finding four best measures (HNR, RPDE, DFA and PPE). In [55], researchers have introduced a new non-linear classification model based on Dirichlet process mixtures to model the non-linear relationship between the response variable and covariates. They have achieved a classification accuracy of 87.7% using all the raw dysphonia features and also have reported that the computation cost was substantially higher compared to other non-linear methods used in their work. Four independent classification models such as DMNeural, neural network, regression and decision tree were utilized and a best classification accuracy of 92.9% was reported using 65% of the data for training and rest of the data for testing [56]. Mutual information measure with permutation test, maximum-relevance-minimum-redundancy (mRmR) and a support vector machine were used to classify healthy controls and PWP [57]. They have found four best features (spread 1, MDVP: Fo (Hz), Shimmer: APQ3, and D2) and obtained a best accuracy of 92.75% with bootstrap resampling validations. Multiclass relevance vector machines (mRVMs) were developed and tested on PD dataset and an average classification accuracy of 89.47% using 10-fold cross validation was reported [59]. An automated genetic programming (GP) and expectation maximization algorithm (EM) based detector was proposed and an average detection accuracy of 93.12% was obtained [60]. Classifier ensemble with rotation forest was constructed and correlation based feature selection (CFS) algorithm was used to improve the classification accuracy of healthy controls and PWP [20]. They have obtained a maximum accuracy of 87.1% from 10-fold cross validation.

Using fuzzy-based non-linear data transformation method and support vector machine, raw dysphonia features were classified [15]. Their system yielded a maximum classification accuracy of 93.47%. An average classification accuracy of 85.03% was achieved using only two features (spread 1 and spread 2) through Fuzzy entropy measure based feature selection and similarity classifier [16]. Feature selection using particle swarm optimization (PSO), harmonic search (HS) and gravitational search algorithm (GSA) and an optimal forest classifier (OPF) were applied to detect PWP [21]. From the original 22 features 14, 10 and 8 best raw dysphonia features were found using PSO, HS and GSA based feature selection respectively. 73.53%, 84.01% and 84.01% of respective accuracies were achieved using PSO-OPF, HS-OPF, and GSA-OPF.

A parallel neural network approach was developed for prediction of PWP with less decision error [58], in which a maximum classification accuracy of 91.20% was achieved using 10 raw dysphonia features and the selection method of those 10 measures was not reported. Fuzzy *k*-nearest neighbor approach with PCA was implemented to enhance the classification of healthy controls and PWP. They have reported a maximum classification accuracy of 96.07%. In [19], SVM based feature selection with rotation forest ensemble classifiers was proposed and obtained a maximum classification accuracy of 97%. Haar wavelet transformation based feature projection filter and regression based classifier were proposed [17,18] to increase the prediction of PWP and reported a maximum classification accuracy of 100%. PSO was employed to select best raw dysphonia features and parameter optimization as well [61] through which a maximum detection accuracy of 97.47% was achieved.

From Table 3, it was observed that the different feature reduction/transformation and feature subset selection, classification models were proposed. Researchers have reported different combinations of raw dysphonia features for improving the detection of PD. Most of the studies in the literature exemplifies that non-linear features yielded the maximum accuracy when compared to the traditional pitch and amplitude based perturbation features. Our study also confirms that we have obtained the maximum average classification accuracy of 100% using the four best non-linear measures (spread 1, spread 2, D2 and PPE) found by SBS. Four best weighted dysphonia features (Jitter (%), Jitter (Abs), RAP and APQ) were found by SFS which reduces the classifiers' performance by atleast 1% or 2%.

6. Conclusions

In any pattern recognition applications, feature pre-processing and feature reduction/selection are the essential steps. The selection of highly useful features from a dataset will eventually help to reduce the complexity of the learning phase and thereby improves the generalization ability of the classifiers. In this work, an integrated approach was proposed to improve the accuracy of detection of PWP. Model-based clustering was proposed as feature weighting method to improve the robustness and discriminative ability of the raw dysphonia features. Two projection based feature reduction techniques (5 features from PCA and 1 feature from LDA) and two step-wise feature subset selection methods (4 features: spread 1, spread 2, D2 and PPE from SBS; 4 features: Jitter (%), Jitter (Abs), RAP and APQ from SFS) were also employed to reduce the number of features or to find the best feature subset from 22 weighted features. Three supervised classifiers (LS-SVM, PNN and GRNN) were applied to test the effectiveness of the proposed weighted features. Different performance measures like precision, sensitivity, specificity, overall classification accuracy and AUC were used to gauge the strength of the proposed method. The experimental results unfold that best weighted dysphonia features found by SBS provided maximum classification accuracy when compared to SFS. The proposed integration of feature weighting method, feature reduction/selection method and classifiers gives a very promising classification accuracy of 100% which is closer to the results published in the literature. From the simulation results, we can also conclude that the proposed method may be instrumental to the physicians in detecting PWP accurately. In the future, the proposed method will be applied to other medical datasets to enhance the discriminatory power of the clinical features.

Acknowledgments

This work is supported by FRGS Research Grant, Ministry of Higher Education, Malaysia. Grant Number: 9003-00297.

REFERENCES

[1] A.E. Lang, A.M. Lozano, Parkinson's disease, New England Journal of Medicine 339 (1998) 1044–1053.

[2] A. Elbaz, J.H. Bower, D.M. Maraganore, S.K. McDonnell, B.J. Peterson, J.E. Ahlskog, D.J. Schaid, W.A. Rocca, Risk tables for parkinsonism and Parkinson's disease, Journal of Clinical Epidemiology 55 (2002) 25–31.

[3] S.K. Van Den Eeden, C.M. Tanner, A.L. Bernstein, R.D. Fross, A. Leimpeter, D.A. Bloch, L.M. Nelson, Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity, American Journal of Epidemiology 157 (2003) 1015–1022.

[4] Statistics on Parkinson's, 2013, http://www.pdf.org/en/parkinson_statistics (accessed on January, 2013).

[5] D.G. Hanson, B.R. Gerratt, P.H. Ward, Cinegraphic observations of laryngeal function in Parkinson's disease, The Laryngoscope 94 (1984) 348–353.

[6] M.E. Smith, L.O. Ramig, C. Dromey, K.S. Perez, R. Samandari, Intensive voice treatment in Parkinson disease: laryngostroboscopic findings, Journal of Voice 9 (1995) 453–459.

[7] L. Ramig, S. Sapir, S. Countryman, A. Pawlas, C. O'Brien, M. Hoehn, L. Thompson, Intensive voice treatment (LSVT®) for patients with Parkinson's disease: a 2 year follow up, Journal of Neurology, Neurosurgery & Psychiatry 71 (2001) 493–498.

[8] L. Cnockaert, J. Schoentgen, P. Auzou, C. Ozsancak, L. Defebvre, F. Grenez, Low-frequency vocal modulations in vowels produced by Parkinsonian subjects, Speech Communication 50 (2008) 288–300.

[9] M.A. Little, P.E. McSharry, E.J. Hunter, J. Spielman, L.O. Ramig, Suitability of dysphonia measurements for telemonitoring of Parkinson's disease, IEEE Transactions on Biomedical Engineering 56 (2009) 1015–1022.

[10] A. Tsanas, M.A. Little, P.E. McSharry, L.O. Ramig, Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests, IEEE Transactions on Biomedical Engineering 57 (2010) 884–893.

[11] A. Tsanas, M.A. Little, P.E. McSharry, L.O. Ramig, Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity, Journal of the Royal Society Interface 8 (2011) 842–855.

[12] H.L. Chen, X.G. Yu, X. Xu, X. Sun, G. Wang, S.J. Wang, An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach, Expert Systems with Applications 40 (2012) 263–271.

[13] Y. Delican, L. Ozyilmaz, T. Yildirim, Evolutionary algorithms based RBF neural networks for Parkinson's disease diagnosis, in: 2011 7th International Conference on Electrical and Electronics Engineering (ELECO), 2011, pp. II-311–II-315.

[14] B.K. Kihel, M. Benyettou, Parkinson's disease recognition using artificial immune system, Journal of Software Engineering and Applications 4 (2011) 391–395.

[15] D.-C. Li, C.-W. Liu, S.C. Hu, A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets, Artificial Intelligence in Medicine 52 (2011) 45–52.

[16] P. Luukka, Feature selection using fuzzy entropy measures with similarity classifier, Expert Systems with Applications 38 (2011) 4600–4607.

[17] I. Mandal, N. Sairam, Accurate telemonitoring of Parkinson's disease diagnosis using robust inference system, International Journal of Medical Informatics 82 (2012) 359–377.

[18] I. Mandal, N. Sairam, New machine-learning algorithms for prediction of Parkinson's disease, International Journal of Systems Science 45 (2014) 647–666.

[19] A. Ozcift, SVM feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease, Journal of Medical Systems 36 (2012) 2141–2147.

[20] A. Ozcift, A. Gulten, Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms, Computer Methods and Programs in Biomedicine 104 (2011) 443–451.

[21] A.A. Spadoto, R.C. Guido, F.L. Carnevali, A.F. Pagnin, A.X. Falcao, J.P. Papa, Improving Parkinson's disease identification through evolutionary-based feature selection, in: 2011

Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 7857–7860.

[22] K. Polat, Classification of Parkinson's disease using feature weighting method on the basis of fuzzy c-means clustering, International Journal of Systems Science 43 (2012) 597–609.

[23] A. Tsanas, M. Little, P. McSharry, J. Spielman, L. Ramig, Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease, IEEE Transactions on Biomedical Engineering 59 (2012) 1264–1271.

[24] D.A. Reynolds, R.C. Rose, Robust text-independent speaker identification using Gaussian mixture speaker models, IEEE Transactions on Speech and Audio Processing 3 (1995) 72–83.

[25] D.A. Reynolds, T.F. Quatieri, R.B. Dunn, Speaker verification using adapted Gaussian mixture models, Digital Signal Processing 10 (2000) 19–41.

[26] Y. Wang, J. Yang, Y. Zhou, Color-texture segmentation using JSEG based on Gaussian mixture modeling, Journal of Systems Engineering and Electronics 17 (2006) 24–29.

[27] U.R. Acharya, S.V. Sree, M. Muthu Rama Krishnan, N. Krishnananda, S. Ranjan, P. Umesh, J.S. Suri, Automated classification of patients with coronary artery disease using grayscale features from left ventricle echocardiographic images, Computer Methods and Programs in Biomedicine 112 (2013) 624–632.

[28] S. İçer, Automatic segmentation of corpus collasum using Gaussian mixture modeling and Fuzzy C means methods, Computer Methods and Programs in Biomedicine 112 (2013) 38–46.

[29] V. Majidnezhad, I. Kheidorov, A novel GMM-based feature reduction for vocal fold pathology diagnosis, Research Journal of Applied Sciences 5 (2013) 2245–2254.

[30] H. Zeng, Y.M. Cheung, A new feature selection method for Gaussian mixture clustering, Pattern Recognition 42 (2009) 243–250.

[31] M. Hariharan, C. Fook, R. Sindhu, B. Ilias, S. Yaacob, A comparative study of wavelet families for classification of wrist motions, Computers & Electrical Engineering 38 (2012) 1798–1807.

[32] A. Quinquis, Digital Signal Processing Using MATLAB, John Wiley &Sons, New Jersy, USA, 2008.

[33] L.I. Smith, A Tutorial on Principal Components Analysis, vol. 51, Cornell University, USA, 2002, pp. 1–26.

[34] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 711–720.

[35] F. Tang, H. Tao, Fast linear discriminant analysis using binary bases, Pattern Recognition Letters 28 (2007) 2209–2218.

[36] M.K. Arjmandi, M. Pooyan, M. Mikaili, M. Vali, A. Moqarehzadeh, Identification of voice disorders using long-time features and support vector machine with different feature reduction methods, Journal of Voice 25 (2011) e275–e289.

[37] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley-Interscience, New York, 2001.

[38] R. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. De Ridder, D. Tax, S. Verzakov, PRTools4.1, A Matlab Toolbox for Pattern Recognition, Delft University of Technology, 2007.

[39] K. De Brabanter, P. Karsmakers, F. Ojeda, C. Alzate, J. De Brabanter, K. Pelckmans, B. De Moor, J. Vandewalle, J. Suykens, LS-SVMlab Toolbox User's Guide, 2010.

[40] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, Least Squares Support Vector Machines, World Scientific, Singapore, 2003.

[41] Y. Li, P.P. Wen, Clustering technique-based least square support vector machine for EEG signal classification, Computer Methods and Programs in Biomedicine 104 (2011) 358–372.

[42] J.S. Sartakhti, M.H. Zangooei, K. Mozafari, Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA), Computer Methods and Programs in Biomedicine 108 (2012) 570–579.

[43] Z. Song, Z. Ji, J.-G. Ma, B. Sputh, U.R. Acharya, O. Faust, A systematic approach to embedded biomedical decision making, Computer Methods and Programs in Biomedicine 108 (2012) 656–664.

[44] T. Sun, J. Wang, X. Li, P. Lv, F. Liu, Y. Luo, Q. Gao, H. Zhu, X. Guo, Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set, Computer Methods and Programs in Biomedicine 111 (2013) 519–524.

[45] H.U.Ä. Ÿuz, A hybrid system based on information gain and principal component analysis for the classification of transcranial Doppler signals, Computer Methods and Programs in Biomedicine 107 (2012) 598–609.

[46] A. Ben-Hur, D. Horn, H.T. Siegelmann, V. Vapnik, A support vector clustering method, in: Proceedings of the 15th International Conference on Pattern Recognition, 2000, 2000, pp. 724–727.

[47] D.F. Specht, Probabilistic neural networks, IEEE Transactions on Neural networks 3 (1990) 109–118.

[48] M. Hariharan, J. Saraswathy, R. Sindhu, W. Khairunizam, S. Yaacob, Infant cry classification to identify asphyxia using time–frequency analysis and radial basis neural networks, Expert Systems with Applications 39 (2012) 9515–9523.

[49] M. Hariharan, R. Sindhu, S. Yaacob, Normal and hypoacoustic infant cry signal classification using time–frequency analysis and general regression neural network, Computer Methods and Programs in Biomedicine 108 (2012) 559–569.

[50] D.F. Specht, A general regression neural network, IEEE Transactions on Neural Networks 2 (1991) 568–576.

[51] M. Hariharan, L.S. Chee, S. Yaacob, Analysis of infant cry through weighted linear prediction cepstral coefficients and probabilistic neural network, Journal of Medical Systems 36 (2012) 1309–1315.

[52] M. Hariharan, M. Paulraj, S. Yaacob, Detection of vocal fold paralysis and oedema using time-domain features and probabilistic neural network, International Journal of Biomedical Engineering and Technology 6 (2011) 46–57.

[53] M. Hariharan, S. Yaacob, S.A. Awang, Pathological infant cry analysis using wavelet packet transform and probabilistic neural network, Expert Systems with Applications 38 (2011) 15377–15382.

[54] M. Hariharan, Y. Sazali, Time-domain features and probabilistic neural network for the detection of vocal fold pathology, Malaysian Journal of Computer Science 23 (2010) 60–67.

[55] B. Shahbaba, R. Neal, Nonlinear models using Dirichlet process mixtures, The Journal of Machine Learning Research 10 (2009) 1829–1850.

[56] R. Das, A comparison of multiple classification methods for diagnosis of Parkinson disease, Expert Systems with Applications 37 (2010) 1568–1572.

[57] C.O. Sakar, O. Kursun, Telediagnosis of Parkinson's disease using measurements of dysphonia, Journal of Medical Systems 34 (2010) 591–599.

[58] F. Åström, R. Koker, A parallel neural network approach to prediction of Parkinson's disease, Expert Systems with Applications 38 (2011) 12470–12474.

[59] I. Psorakis, T. Damoulas, M.A. Girolami, Multiclass relevance vector machines: sparsity and accuracy, IEEE Transactions on Neural Networks 21 (2010) 1588–1598.
[60] P.-F. Guo, P. Bhattacharya, N. Kharma, Advances in detecting Parkinson's disease, in: Medical Biometrics, Springer, Berlin Heidelberg, 2010, pp. 306–314.
[61] W.-L. Zuo, Z.-Y. Wang, T. Liu, H.-L. Chen, Effective detection of Parkinson's disease using an adaptive fuzzy *k*-nearest neighbor approach, Biomedical Signal Processing and Control 8 (2013) 364–373.