International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Two-stage Whale Optimization Method for Classification of Parkinson's Disease Voice Recordings

Saban OZTURK[1], Yavuz UNAL[2]

Submitted: 02/02/2020 Accepted: 04/05/2020

Abstract: The definitive treatment of Parkinson's disease, which causes movement disorders and has been increasing in recent years, is still not available today. However, effective studies are being conducted to improve the quality of life of the patients. In this study, a method for the efficient classification of data from Parkinson's disease and normal individuals is proposed. Since the dataset used in the proposed study consists of replicated samples, independence-based classifiers cannot be used for this dataset. When the distribution of the features in the dataset is examined, the success of the conventional classifiers is very low due to the fact that the distribution centers between the clusters are very close to each other. Based on the basic idea that increasing the distance of the cluster centers from each other will increase the success, dimensionality techniques such as PCA, ICA, Relieff, and RICA have been used. When the desired success is not achieved, a bond theory was established using a two-stage Whale optimization algorithm. Accordingly, the features of the three samples taken from an individual are closed to each other in the feature space, the total samples belonging to the same class are drawn to one side of the feature space and the feature space of the other class is positioned farthest from the center point. Thus, three different samples belonging to the same individual will be classified with the same label. In addition, since the class difference will be high, conventional classifiers such as SVM, k-NN, and LDA can work successfully. The proposed method is compared with other techniques and as a result, it is seen that the representation ability in the feature space is stronger than other related methods.

Keywords: *Parkinson disease, Machine Learning, Classification, Optimization*

1. Introduction

Parkinson's disease (PD), which causes movement disorders, is the second most common neurodegenerative disease worldwide. Given that the world's population is getting older, it is clear that the diagnosis and treatment of this disease will gain importance. This disease, which needs to be controlled because of its continuous progress, is more common in men than in women [1]. It is associated with the destruction of dopamine cells, which play a significant role in the control of body movements. Although the cause is not yet known, family, stress, environmental factors are thought to be influential. As the symptoms of the disease vary from person to person, the progression of symptoms also varies. In some patients, the findings begin slowly and progress slowly, while in others, progression may be severe. Although PD presents various symptoms, some of these symptoms include tremor or shaking, slowed movement, impaired posture, rigid muscles, voice change, etc. [2].

One of the first impaired functions seen as early symptoms in patients with Parkinson's is speech and voice. Impairment of respiratory and joint functions also affects voice and speech [3]. In the early stages of the disease, these disorders may not be perceived by humans. When the recorded speech signals are analyzed, these disturbances can be noticed [4]. There is no standard test for the diagnosis of Parkinson's disease. Therefore, voice or speech disorders can be used for early diagnosis in patients [5]. Thanks to the contribution of expert systems proposed for the analysis of medical data, recording, and analysis of audio signals has become possible. Automated analysis of medical data not only simplifies the work of specialists but also reduces the waiting time of patients. This two-way positive contribution has attracted the attention of artificial intelligence researchers.

Some studies in the literature diagnose Parkinson's disease from voice analysis. Naranjo et al. [4] developed a clinical expert system for the diagnosis of Parkinson's disease [4]. This expert system extracts features from audio recordings. They developed a new Bayesian approach and used it in this study. They tried the assessment of this expert system on voice recording replication data from 80 people and successfully diagnosed Parkinson's disease from these data. Naranjo et al. [5] conducted another study on voice recording replication data from the 80 subjects, as mentioned above (some with PD and some with healthy). In this study, they diagnosed PD with a two-stage variable selection and classification method and compared them with the previous study. They contributed to the scientific literature by obtaining a lower computational time than a single classification approach. Harel et al. [6] performed a retrospective analysis of the speech on two well-known individuals and two matched controls to determine whether some acoustic measurements were sensitive markers of early pathophysiological changes or treatment response in PD. Tsanas et al. [7] have demonstrated the clinically beneficial properties of mean PD progression by speech processing algorithms. Then, they used a feature selection algorithm and used the linear and nonlinear regression techniques, including classical least squares and non-parametric classification and regression

[1] *Electric and Electronic Engineering /Technology Faculty, Amasya University, Turkey*
ORCID ID : 0000-0003-2371-8173
[2] *2Computer Engineering/Technology Faculty, Amasya University, Turkey*
ORCID ID : 0000-0002-3007-679X
** Corresponding Author Email: saban.ozturk@amasya.edu.tr*

trees, to map the selected subset of features statistically to the unified Parkinson's disease rating scale (UPDRS). Little et al. [8] obtained 195 voice recordings from 31 people, 23 of whom had Parkinson's disease, and they analyzed the disease status of the individuals by kernel SVM. Timothy J. Wroge et al. [9] analyzed voice data obtained from healthy subjects and people with Parkinson's disease with ANN and other known machine learning algorithms and successfully diagnosed PD. Harihanan et al. [10] performed a weighted clustering model of raw dysphonia data (22 traits) by applying feature selection/reduction with PCA, LDA, SFS, and SBS, and classification with LS-SVM, PNN, and GRNN to diagnose a hybrid Parkinson's disease. Eskidere et al. [11] studied Parkinson's diagnosis using the algorithms SVM, LS-SVM, MLPNN, GRNN on the Parkinson's telemonitoring dataset and obtained the best performance with LS-SVM. Von Orozco-Arroyave et al. [12] analyzed speech signals for the diagnosis of Parkinson's disease using nonlinear dynamics. Novotny et al. [13] have demonstrated a fully automated approach to evaluate articulator disorders in PD. Unlike previous studies that focused primarily on the evaluation of dysphonic patterns, this is the first study to investigate the automatic characterization of acoustic aspects of articulatory dysfunction in PD.

Due to the increasing hardware power, artificial intelligence methods have become widespread recently and have been used in almost every field [35-38]. The purpose of these methods is to make the closest selection to human intelligence during any selection process. But when we go further, the expectation from these systems increases. As a result of increasing expectations, computer programs are expected to make the most appropriate choice. Recently, the use of optimization techniques has met this expectation. Tan et al. [14] developed a decision support system for the diagnosis of skin cancer. They have been used in PSO based feature selection. The feature selection data is classified with deep learning. In a study by Güraksin et al. [15] developed a system for age detection from bone x-ray images of children aged 0-6 years. Image processing techniques have been used to extract features from x-ray images, and these features are given as an introduction to the classifier. They used a new SVM classifier using PSO. Khomri et al. [16] performed vessel segmentation from retinal image images. For this purpose, they applied morphological pre-processing to the eye images and then segmented the retinal images with the help of an artificial bee colony algorithm. They compared this new method with other methods. Subanya and Rajalaxmi [17] classify cardiovascular disease with the help of an artificial bee colony algorithm. They used artificial bee colony detection in feature selection and classified them in SVM by comparing them with various kernel functions. Li and Wang [18] used the ant colony algorithm to detect ship from SAR radar images. They used ant colony algorithms during edge detection and thus successfully detected ships from SAR images. Weiguo has used an artificial immune system algorithm to achieve high weather forecast accuracy [19]. Today, the whale optimization algorithm is a remarkable optimization technique that is successful in the analysis of medical signals. The whale optimization algorithm is a meta-heuristic optimization algorithm. It is an optimization method that mimics the hunting behavior of humpback whales. It was inspired by the bubble hunting strategy used by humpback whales [20].

In this study, the Parkinson Dataset with replicated acoustic features from the University of California-Irvine (UCI) machine learning database is used [21]. The data cannot be classified with independence-based classifiers because they are obtained by duplicating from the same subject in triplicate groups. This makes it difficult to classify people with Parkinson's disease correctly [4].

In this study, we propose a method for the effective classification of people with Parkinson's disease and healthy individuals. Data obtained from the dataset cannot achieve the desired level of classification performance with independence-based classifiers. For a satisfactory classification performance, firstly, the class distributions of the features in the data set are examined. Since class distributions are irregular, independence-based classifiers have failed in this dataset. Dimensionality techniques such as PCA, ICA, Relief, and RICA are used to increase the classification success. Firstly, the effect of these methods on class distribution is examined. When the obtained graphs are interpreted, it is seen that the features of the classes cannot be represented separately in the feature space. To keep the in-class difference less and increase the inter-class difference, a bond theory is formed by using the Whale optimization algorithm. The proposed method has been compared with other techniques, and as a result it has been found that feature representation ability is better in feature space.

This study is organized as follows: Section 2 introduces the dataset, describes the techniques of size reduction and the method of the study. In Section 3, the classification algorithms used are defined, and the obtained analyzes are given. Conclusions and recommendations are given in Section 4.

2. Material and Methods

The main objective of this study is to classify the voice data of PD with a satisfactory success rate. For this purpose, conventional classifiers which are very common in the literature are selected. However, the fact that the dataset does not have any feature differences between the different classes, and the intra-class distinction does not make this process possible. To understand the problem more clearly and to find a solution, the distribution graph of the raw features is examined, and effective solutions are emphasized. Firstly, feature reduction and space replacement methods in the literature have been tried. It has been found that these algorithms which produce successful results for many applications are not suitable for this dataset. This is because three samples taken from each person do not show consistency in this dataset. Because of this negative situation, the current algorithms have limited capacity to represent the features separately. As a solution, three samples taken from an individual should be approximated within themselves and gain linearity, and the characteristics of individuals belonging to the same class should be drawn to the discrete corners of the feature space. In this case, the features of the other class will shift to a farther point in the feature space. In order to carry out the mentioned tasks, a weight is given to each feature, and these weights are updated with the whale optimization algorithm. Also, the problem-specific target function has been determined and updated. The study is shown in Figure 1 in general terms.

2.1. Dataset

The proposed method is used to classify data from Parkinson's Dataset with replicated acoustic features from the UCI machine learning repository [21]. The data in this dataset are obtained from 80 people over 50 years of age. Of these, 40 are healthy and 22 (55%) of them are male and 18 (45) are female. 40 people have Parkinson's disease. Of the 40 patients with Parkinson's disease, 27 (67.5%) are male and 13 (32.5%) are female.

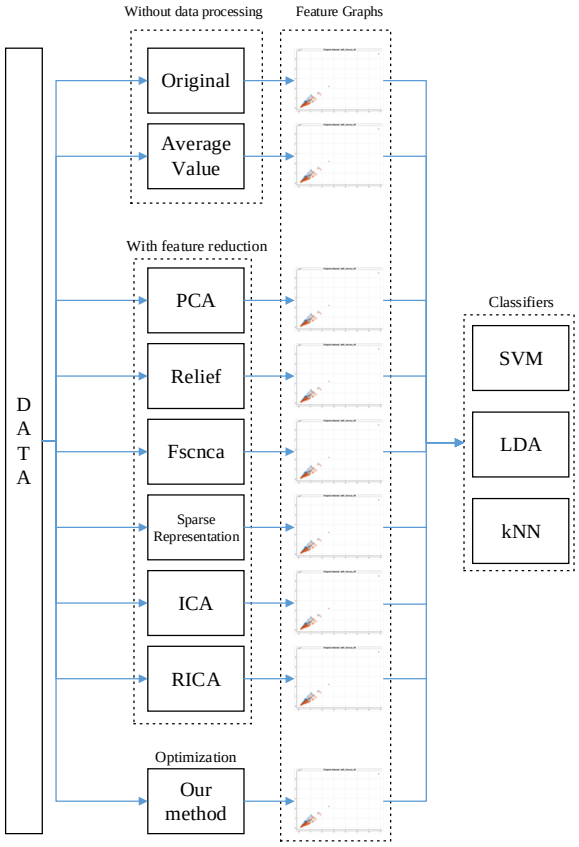


Fig. 1. Overview of this study

This dataset contains 44 acoustic features. These features are divided into 9 groups as shown in Table 1. Group 6, 7, 8, and 9 consist of a single feature, and 5 groups consist of more than one feature. These groups include pitch local perturbation measures, amplitude local perturbation measures, harmonic-to-noise ratio measures, mel frequency cepstral coefficient-based spectral measures, Derivatives of Mel frequency cepstral coefficients, Recurrence period density entropy, Detrended-to-noise excitation ratio.

2.2. Traditional Techniques for Dimensionality Reduction

It is known that as the data size increases, the performance of traditional algorithms used in applications decreases [22]. In this regard, dimension reduction techniques are often applied to simplify the data model, either as part of the data analysis or as a preliminary stage. This generally includes defining a suitable low-dimensional representation for the high-dimensional original dataset. When working with reduced dimensional data, both the computational load is significantly reduced, and more accurate, and easily interpretable results are produced by methods such as classification or clustering. Thanks to size reduction;

- Defining a reduced set of variables can be very useful for information discovery.
- For many learning algorithms, processing time can be reduced by reducing the number of variables.
- Negative effects on accuracy can be eliminated by eliminating noisy or irrelevant features that may have the same effect on classification as predictive variables [23].

In this study, PCA, Relief, Sparse Autoencoder, FSCNCA, ICA and RICA dimensional reduction methods are used to evaluate the performance of the proposed method.

Table 1. Description of the features of the Parkinson Dataset with replicated acoustic features Data Set

Group Number	Name	Features
1	Pitch local perturbation measures	Relative jitter, absolute jitter, relative average perturbation (RAP), pitch pertur- bation quotient (PPQ)
2	Amplitude perturbation measures	Local shimmer, shimmer in dB, 3-point amplitude perturbation quotient (APQ3), 5- point amplitude perturbation quotient (APQ5), and 11-point amplitude perturbation quotient (APQ11).
3	Harmonic-to-noise ratio measures	Harmonic-to-noise ratio in the frequency band 0–50 0 Hz (HNR05), in 0–150 0 Hz (HNR15), in 0–2500 Hz (HNR25), in 0–3500 Hz (HNR35), and in 0–3800 Hz (HNR38).
4	Mel frequency cepstral coefficient-based spectral measures	MFCC0, MFCC1, . . . , MFCC12
5	Derivatives of Mel frequency cepstral coefficients	Delta0…Delta12
6	Recurrence period density entropy	RPDE
7	Detrended fluctuation analysis	DFA
8	Pitch period entropy	PPE
9	Glottal-to-noise excitation ratio	GNE

ICA (Independent Component Analysis); It, which is one of the blind source separation methods, is the determination of the hidden factors underlying random variables, measurements or signals by statistical methods and calculation techniques. It assumes that the subcomponents are non-Gaussian signals and are statistically independent. The data are considered linear combinations of some unknown hidden variables [24].

Let suppose $S = [S_1(t), S_2(t), \dots, S_N(t)]$ are independent signals. Linear combinations of them are calculated as $Y(t) = AS(t)$. The purpose of ICA is to obtain the linear transformation *S(t)* from *Y(t)*. PCA(Principal Component Analysis); It is a multivariate statistical method that provides recognition, classification, size reduction and interpretation. This approach tries to find the most robust pattern in the data. Therefore, it can also be used as a pattern-finding technique [25]. Often the diversity of the data can be captured by a small set of sizes selected from the entire set of sizes. Since the noise in the data is weaker than the patterns, these noises can be removed as a result of size reduction [26]. Let's define the sequence *x=(x1, x2,…, xp)* for *n* observation examples, each with *p* variables. The first principal component, Equation 1, is used to calculate.

$$z_1 \equiv a_1^T x = \sum_{i=1}^{p} a_{i1} x_i \quad (1)$$

where $a_1 = (a_{11}, a_{21}, \dots, a_{p1})$. In this way, Equation 2 is used to calculate the *k*th principal component.

$$z_k \equiv a_k^T x = \sum_{i=1}^{p} a_{i1} x_i \quad (2)$$

where vector $a_k = (a_{1k}, a_{2k}, \dots, a_{pk})$.

Relieff; The Relieff method, proposed by Kira and Rendell [27], finds the value of features by attempting to uncover the interdependencies between them. The method is used for feature selection in binary classification problems. It works by weighting the closest samples in the classes to which the feature is located and not belonging [28].

Sparse Auto-encoder; the automatic encoder is used as a self-supervised learning model because it generates its own tags while training the data. AE aims to learn a representation for the purpose of dimension reduction for a data set. The weights are changed until he sees the same input as the output after the input data encryption-decryption. When the target is reached, the number of nodes in the hidden layer and the input data are represented [29].

FSCNCA; It is one of the feature selection algorithms used for classification. It performs feature selection using neighbourhood component analysis (NCA), which is a non-parametric technique to select features to maximize predictive accuracy of regression and classification algorithms. NCA uses distance metrics to calculate the weights of features. The advantage of it is that it generates positive weights for each feature. Detailed information on the feature selection calculation using the FSCNCA method is available in [30].

RICA; the RICA algorithm, called Reconstruction ICA, is developed to eliminate the deficiencies in the ICA algorithm. The orthonormality constraint has been removed in RICA. Instead, a soft penalty called reconstruction costs is used to prevent degeneration [31].

$$\min_{w} \frac{\lambda}{m} \sum_{i=1}^{m} \left\| W^T W x^{(i)} - x^{(i)} \right\|_2^2 + \sum_{i=1}^{m} \sum_{j=1}^{k} g(W_j x^{(i)}), \quad \text{W.W}^T = 1 \quad (3)$$

2.3. Proposed Method with Whale Optimization

The order of recording the data used in this study, and the relationship of the data with each other is unusual. The distribution of the characteristics representing the data has been intertwined in the interclass order, whereas in the in-class order, they are far from each other. Besides, this pattern is not in this format for all features. In fact, not even an easy-to-follow pattern has been achieved. Therefore, independence-based classifiers and traditional classification approaches do not work for this dataset. Converging the cluster centers belonging to the same class and removing the cluster centers belonging to different classes from each other will increase the success of the classifiers. PCA, ICA, ReliefF, RICA, FSCNCA, and sparse representation techniques are used for this purpose. However, the desired feature representation could not be realized. One of the most important reasons for this is taking three samples for each individual. Under normal circumstances, this process seems to be entirely innocent, but it has broken many balances in this dataset. Because the same characteristics for three samples belonging to the same individual are quite different from each other. When the desired success is not achieved, a bond theory is established using the Whale optimization algorithm. The proposed method has three main objectives;

- Positioning the features of three measurements taken from an individual close to each other in the feature space and accepting them as a single measurement.
- To ensure that the total samples belonging to the same class are drawn to one side of the space and remain close to the class center.
- Positioning other class features farthest from the center point.

Thus, three different samples belonging to the same individual will be classified with the same label. In addition, since the class difference will be high, classical classifiers such as SVM, k-NN, and LDA will be able to work successfully. An error function is created for the whale optimization algorithm to perform all the mentioned tasks. Accordingly, a two-stage training process is carried out. In the first step, a merge is performed for three samples taken from an individual. Figure 2 shows the distribution results of three measurements belonging to the same individual.

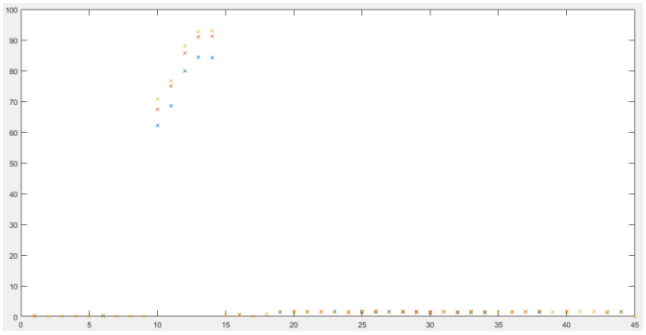


Fig. 2. The distribution results of three measurements belonging to the same individual

If we add the data of another individual to understand the complexity, we get distribution as in Figure 3.

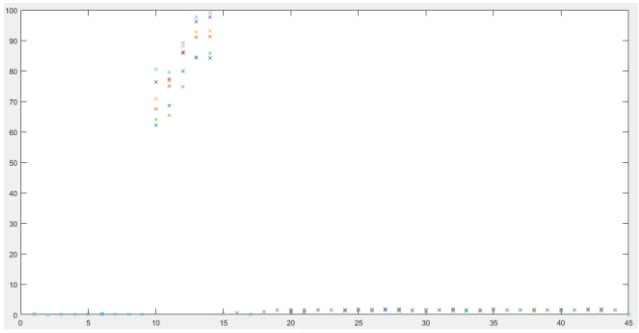


Fig. 3. The distribution results of three measurements belonging to two different people

In order to reduce this complexity, firstly, three measurements taken from a single individual should be successfully expressed on a single axis. Since this operation is contrary to the binary classification process, our study has two stages. In the first step, each individual is expressed with a single feature curve with whale optimization. The target function used for this is shown in Equation 4.

$$f(w_1) = w_1 \left\| (x_1 - x_2) + (x_1 - x_3) \right\|_2^2 \quad (4)$$

where *w* represents updating parameter, *x* represents data from individuals, while subscripts represent the number of samples. After the first update process, the difference between three samples of each individual is minimized. When examined in three different curves, it can be seen that they belong to the same individual. But the difference between classes still has not changed. To change this difference, a second update is made with the whale optimization algorithm. In this process, the difference between the classes will be increased, and the intra-class difference will be reduced.

Equation 5 is used for this.

$$E_1 = \left\| f_{out} - N_{label} \right\|^2, \min \left\| f_{out} < 0.5 \right\|, \max \left\| f_{out} > 0.51 \right\| \quad (5)$$

The proposed optimization framework is shown in Figure 4.

3. Experiments and Experimental Results

In the proposed study, firstly, the results of the original data obtained with the traditional classifiers are examined. Figure 5(a) shows the class distributions of the original features. From the results obtained, the original features do not have a meaningful distribution, and the standard deviation is quite high. As such, these results do not allow reliable findings. As predicted, the success of the classification remains low while the data is original. The feature graph for the two classes examined is visibly problematic. In order to overcome these problems as mentioned earlier, various feature reduction and feature space modification methods have been emphasized. However, firstly, it is tried to combine three samples belonging to an individual by considering the simple solutions to complex problems. Three samples are averaged for this process. Figure 5(b) shows the average class distributions. However, the desired success values have not been achieved. In the second step, the PCA and FSCNCA methods, commonly known as feature reduction methods, are applied to the dataset. 25 features are used primarily in the application of these methods. As such, it can be considered as a feature reduction technique. Figure 5(c) shows 25 distributions for PCA, and Figure 5(e) shows 25 distributions for FSCNCA. Although the 25 characteristic PCA distribution performs better than the original results, the 25 characteristic FSCNCA distribution does not have a reasonable distribution. Since the reduction of features with PCA facilitates the feature space, the reduction of features has been addressed. However, success is reduced when the features are further reduced. Starting from this, 40 features are used, which are almost equal to the number of original features. Figure 5(d) shows 40 distributions for PCA, and Figure 5(f) shows 40 distributions for FSCNCA. In this case, it is observed that the feature distributions approach the original distribution. As a result of the obtained distributions, it is understood that these methods cannot be used.

Relieff, rica, ica, and sparse representation methods, which are used in the literature for the effective positioning of features, are examined in the second stage of our study. Accordingly, 25 and 40 are used as two number of features for each algorithm. In Figure 6(a), the class distributions of the features that are reduced to 25 features by Relieff are given, while in Figure 6(b), the class distributions of the features that are reduced to 40 features by Relieff are given. Class distribution of 25 features by reducing features with RICA is shown in Figure 6(c), and the class distribution of 40 features is shown in Figure 6(d). Class distribution when the number of features is reduced to 25 by sparse representation is shown in Figure 6(e), and class distribution when the number of features is reduced to 40 is shown in Figure 6 (f). The Relieff algorithm makes the features more dispersed in space but fails to provide in-class proximity. The results obtained with Relieff are not understandable. Although Rica algorithm successfully identifies the class centers, the in-class distribution has been quite vast. Therefore, the characteristics of different classes are mixed and do not have the desired characteristic for the classifier. When the results obtained by the Ica algorithm are examined, the centers of the two classes are brought closer to each other while the intra-class difference increases. In theory, this is almost the opposite of the desired situation. This reduces the success of the classification according to the original data and also increases the standard deviation. Experiments with Sparse representation have produced more meaningful results. Although these results do not have a significant impact on the success of the classification, they provide us with an idea of the optimization method proposed by us. Because features represented by sparse representation are very close to the weighted representation of the original features. The proposed algorithm is based on this idea.

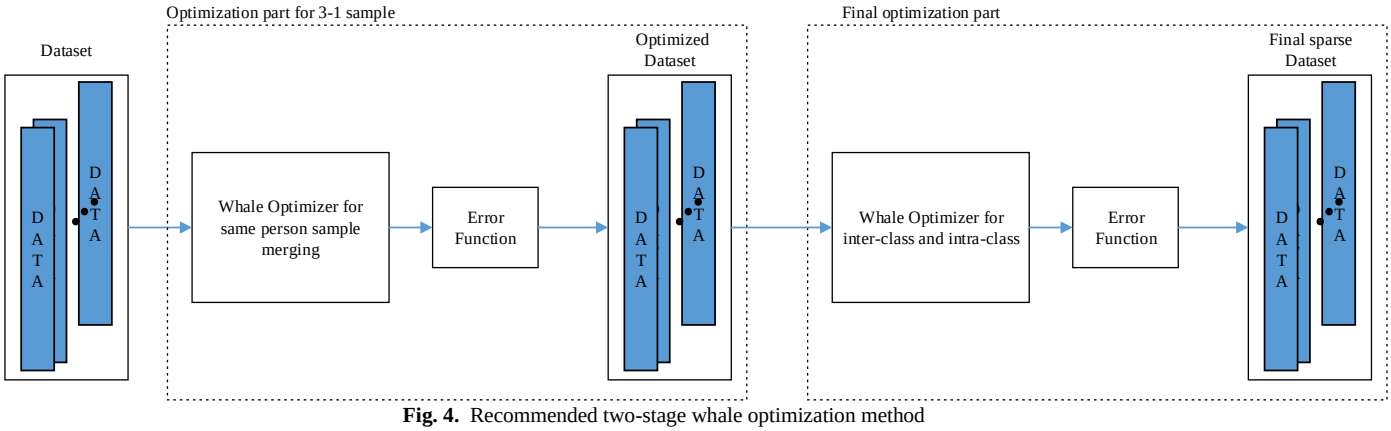


Fig. 4. Recommended two-stage whale optimization method

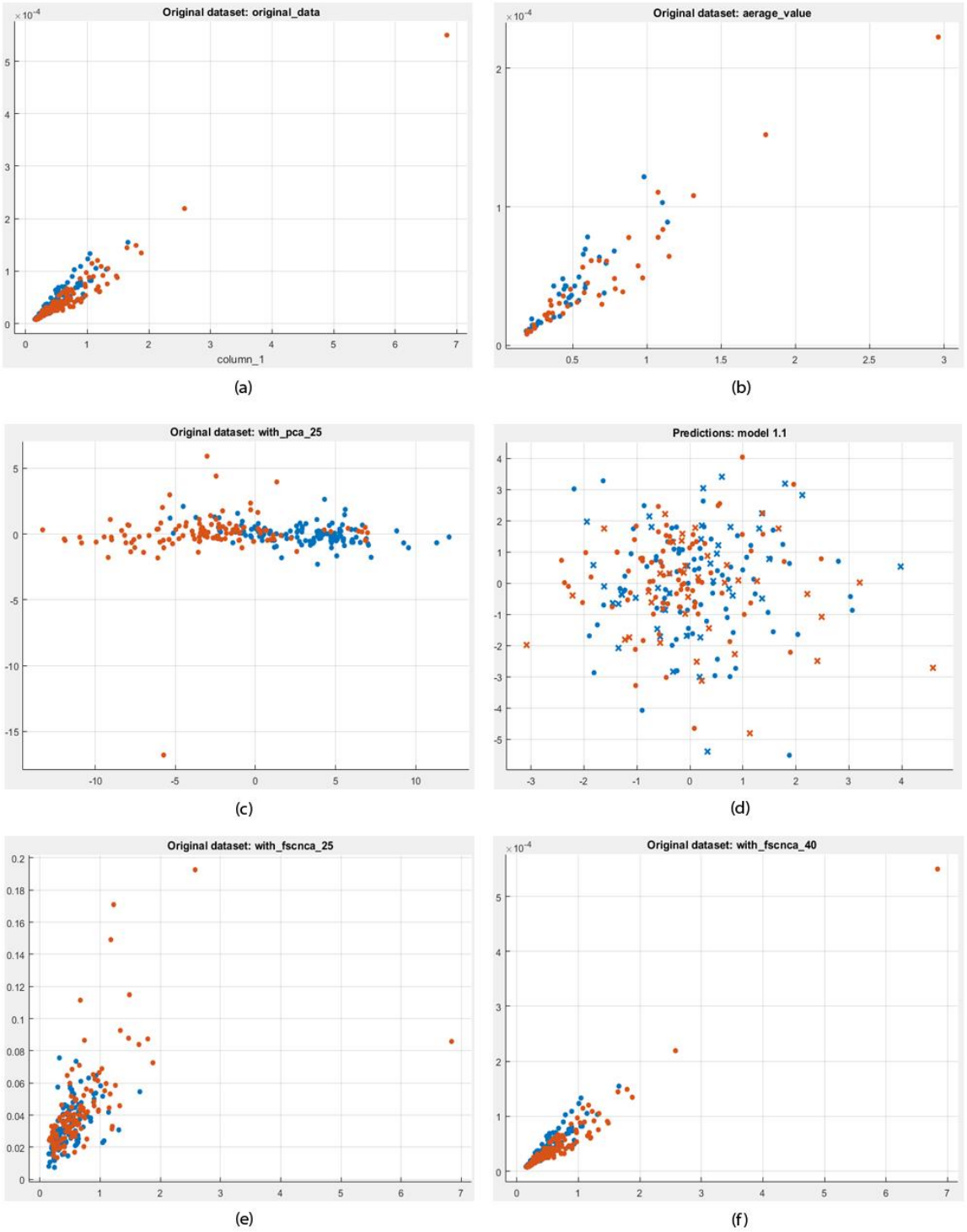


Fig. 5. (a) Class distribution of raw features, (b) Class distribution of average features, (c) 25 features with PCA, (d) 40 features with PCA, (e) 25 features with FSCNCA, (f) 40 features with FSCNCA

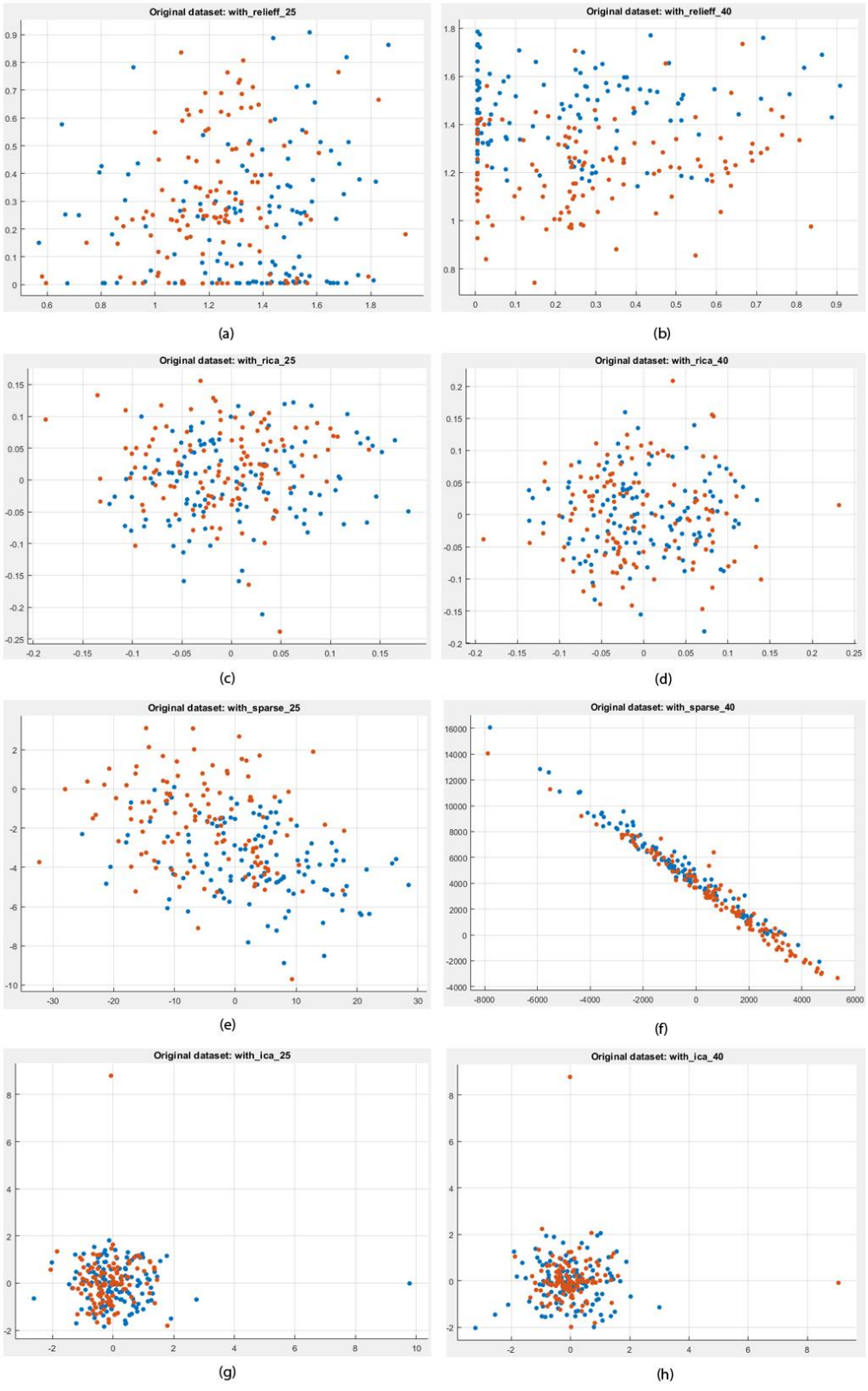


Fig. 6. (a) 25 features with Relieff, (b) 40 features with Relieff, (c) 25 features with RICA, (d) 40 features with RICA, (e) 25 features with Sparse representation, (f) 40 features with Sparse representation, (g) 25 features with ICA, (h) 40 features with ICA

The methods used successfully in the literature and producing impressive solutions for many problems do not work in the dataset. Therefore, a Whale optimization algorithm based approach has been proposed. The results of the proposed method are shown in Figure 7. When Figure 7 is examined, it is understood that solutions that have not been fully realized by other methods are realized. First, a clean distribution is obtained for the patient class represented by red, and the healthy class represented by blue. The centers of the two classes are assigned to distant points to facilitate the distinction between classes. Besides, keeping the in-class difference low improves class success and reduces the standard deviation.

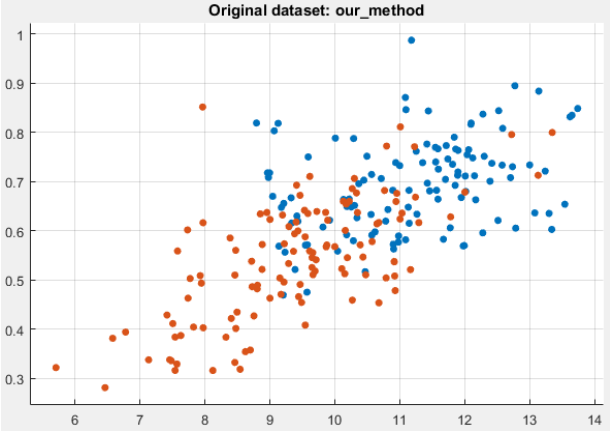


Fig. 7. Distribution of features with the Whale optimization algorithm

4. Performance Evaluation

Examining the distribution of features in graphs helps to develop methods to make the problem easier for classifiers. However, the performance of the proposed methods is measured according to the classification results. Accordingly, the most commonly used SVM, LDA, and kNN algorithms are used in the literature to measure the success of the proposed method. Although these classifiers serve the same purpose, the approaches they follow in this way are different from each other. Therefore, these three classifiers have been chosen. SVM is a method that creates an n-dimensional hyperplane by dividing data into two categories. SVM models are closely related to artificial neural networks, and SVM, which uses the sigmoid kernel function, has a two-layer, feedforward neural network [32]. LDA (Linear Discriminant Analysis) is a method to reduce the size by maximizing the linear separability of groups belonging to different classes in the data. The variance within each group is minimum, and the average of the groups keeps each other at maximum level [33]. kNN is one of the controlled learning algorithms that solves the classification problem. In the kNN method; calculating the similarities of the data to be classified to the normal behavior data in the learning set, with the average of the data thought to be the closest k, according to the threshold value is assigned to classes [34].

When the number of features is reduced to 25 and 40 features, respectively, the results are given in Table 2. According to this, 76.8% classification is achieved with SVM, 75.8% classification with LDA, and 73.8% classification with K-NN without feature reduction. Although the success rates are very close to each other, the standard deviation is quite high. Therefore, it is difficult to obtain consistent results. On the other hand, when the results of all other feature reduction methods are examined, it is seen that success does not increase. However, significant reductions in standard deviation have occurred. Especially the results obtained with sparse representation are very reliable, but the desired level of success has not been achieved. When Table 2 is examined in this respect, it is clear that feature reduction methods have failed due to the nature of the dataset used. In contrast, they have made significant improvements in standard deviation.

In terms of SVM classifier with 25 features; The Relieff algorithm increases the accuracy of the classification slightly, and the success of the classification has decreased in other feature reduction methods. For 40 features; Relieff, Fsnca, and Sparse representation algorithms have improved the accuracy of classification in feature reduction, while the success of classification has decreased in other feature reduction methods.

From the perspective of the LDA classifier with 25 features; PCA and Relieff algorithms have shown that the accuracy of classification increases and classification success decreases with other feature reduction methods. For 40 features; the relieff algorithm improves the classification accuracy by reducing the features, while the success of the classification decreases in other feature reduction methods.

From the point of view of the kNN classifier for 25 and 40 features; it is seen that classification accuracy does not increase with the feature reduction methods used.

Table 2. Classification results of reduced data with SVM, LDA and KNN

Feature Reduction Algorithms	**Classifier**					
	SVM		**LDA**		**K-NN**	
Raw Features (44)	76.8 (±8.1)		75.8 (±7.7)		73.8 (±7.9)	
	25 features	**40 Features**	**25 Features**	**40 Features**	**25 Features**	**40 Features**
PCA	74.9 (± 5.2)	73.2 (± 5.2)	76.1 (± 5.1)	74.0 (± 5.1)	64.3 (± 5.2)	69.9 (± 5.2)
ReliefF	78.1 (± 2.9)	78.0 (± 2.7)	76.2 (± 3.1)	75.9 (± 3.6)	72.6 (± 4)	73.1 (± 3.8)
Fscnca	76.6 (± 4.7)	77.0 (± 4.3)	75.8 (±4.1)	74.2 (± 4.6)	71.2 (±5.4)	69.9 (± 4.9)
Sparse	71.8(± 2.9)	77.6 (± 2.7)	71.6 (± 2.9)	75.6 (± 2.6)	61.7(± 4)	73.8 (± 2.6)
ICA	64.2 (± 7.6)	67.7 (± 6.5)	65.8 (± 7.1)	67.8 (± 7.4)	62.1 (± 7.2)	63.9 (± 7.2)
RICA	72.0 (± 4.3)	70.3 (± 4.1)	75.8 (± 4.3)	64.6 (± 4.6)	58.8 (± 4.7)	60.7 (± 4.5)

Comparative results of other studies on the same data set with the proposed classification accuracy are given in Table 3. Accordingly, Naranjo et al. achieved 86.2% classification accuracy in 2017. Again, Naranjo et al. achieved 75.2% classification accuracy in 2016. The classification accuracy obtained in our study is 86.5%. Two studies by Naranjo et al. are based on producing a solution suitable for the replicated data structure. In the proposed study, works are carried out on the distribution of features of the dataset and how to apply this distribution to classifiers is studied. This is the main reason why the proposed study is successful.

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

Method	Accuracy %
Naranjo et al. (2017) [5]	86.2
Naranjo et al. (2016) [4]	75.2
Our Method	86.58

5. Conclusion

In this study, an effective method for the automatic detection of Parkinson's disease from voice signals, which adversely affects human health, is proposed. For this purpose, the classification of PD signals by conventional classifiers is first attempted. As a result of the results obtained, the distribution of features is examined, and it is concluded that the form of the features should be changed. Following the failure of the most successful feature reduction methods in the literature, an optimization approach has been proposed to ensure the distribution of features automatically. The proposed method consists of two stages. In the first step, replicated signals are drawn to a single axis with the whale optimization algorithm. With this process, we obtain the same classification result from different measurements taken from the same individual. In the second stage, another whale optimization is performed to discrete the center of each class in the updated feature map. The target functions are different in these two whale algorithms. As a result, the success rate of the results obtained by the proposed method is higher than other algorithms. Besides, the standard deviation is lower than other state-of-the-art algorithms.

References

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

[2] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease", IEEE transactions on biomedical engineering, vol. 59(5), 2012.

[3] J.R. Duffy, Motor Speech Disorders: Substrates, Differential Diagnosis, and Management, Elsevier, 2005.

[4] Naranjo, L., Pérez, C. J., Campos-Roca, Y., & Martín, J. (2016). Addressing voice recording replications for Parkinson's disease detection. Expert Systems with Applications, 46, 286-292.

[5] Naranjo, L., Pérez, C. J., Martín, J., & Campos-Roca, Y. (2017). A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications. Computer methods and programs in biomedicine, 142, 147-156.

[6] Harel,B., Cannizzaro, M., &Snyder, P.J. (2004). Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. Brain and Cognition,56,24–29.

[7] Tsanas,A., Little, M.A., Mc Sharry, P.E., &Ramig, L.O.(2010). Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. IEEE Transactions Biomedical Engineering, 57,884–893.

[8] Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., &Ramig, L.O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson'sd isease.I EEETrans-actions on Biomedical Engineering, 56,1015–1022.

[9] T. J. Wroge ; Y. Özkanca ; Cenk Demiroglu ; Dong Si ; David C. Atkins ; Reza Hosseini Ghomi, Parkinson's Disease Diagnosis Using Machine Learning and Voice, 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), December, 2018.

[10] Hariharan M, Polat K, Sindhu R. A new hybrid intelligent system for accurate detection of Parkinson's disease, Computer Methods and Programs in Biomedicine, 113, 904-914, 2014.

[11] Eskidere,O., Ertaç,F.,&Hanilçi,C.(2012). A comparison of regression methods for remote tracking of Parkinson's disease progression. Expert Systems with Applications, 39, 5523–5528.

[12] Orozco-Arroyave J.R., Arias-Londoño J.D., Vargas-Bonilla J.F., Nöth E. (2013) Analysis of Speech from People with Parkinson's Disease through Nonlinear Dynamics. In: Drugman T., Dutoit T. (eds) Advances in Nonlinear Speech Processing. NOLISP 2013. Lecture Notes in Computer Science, vol 7911. Springer, Berlin, Heidelberg

[13] M. Novotny , J. Rusz , R. Cmejla , E. Ruzicka ,Automatic evaluation of articulatory disorders in Parkinson's disease, IEEE/ACM Trans. Audio Speech Lang. Process 22 (9) (2014) 1366–1378 .

[14] Tan, T.Y., Zhang L., Lim C.P., Intelligent skin cancer diagnosis using improved particle swarm optimization and deep learning models, Applied Soft Computing, Volume 84, November 2019, 105725.

[15] G.E. Güraksın, H. Haklı, H. Uğuz, "Support vector machines classification based on particle swarm optimization for bone age determination", Applied Soft Computing, 24, 597-602, 2014.

[16] Kohomri, B., Christodoulidis. A., Djerou., L., Babahenini. M. C., Cheriet. F., Retinal blood vessel segmentation using the elite-guided multi-objective artificial bee colony algorithm, IET Image Processing, Vol.12, Num.12, pp.2163-2171, 2018.

[17] Subanya, B., Rajalaxmi R.R., Feature Selection using Artificial Bee Colony for Cardiovascular Disease Classification, International Conference on Electronics and Communication System (lCECS - 2014), Coimbatore, india, 2014.

[18] Li, L., Wang, J., SAR Image Ship Detection Based on Ant Colony Optimization, International Congress on Image and Signal Processing (CISP 2012), Chongging , China. 2012.

[19] Weigo, X., The Weather Prediction Method Based on Artificial Immune System, International Forum on Information Technology and Applications, Kunming, Chinal , 2010.

[20] Mirjalili, S. and Lewis, A. (2016). The Whale Optimization Algorithm. Advances in Engineering Software, 95, pp.51-67

[21] https://archive.ics.uci.edu/ml/datasets/Parkinson+Dataset+with+replicated+acoustic+features+

[22] Chen L. (2009) Curse of Dimensionality. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA.

[23] Cunningham, P. 2007. Dimension Reduction. Technical Report. UCD-CSI-2007-7. University College Dublin.

[24] Aapo Hyvainen, Erkki OJa, Juha Karhunen. "Independent Component Analysis", (1. ed.). New York: John wiley &sons, 2001.

[25] C. M. Bishop. Pattern Recognition and Machine Learning, Springer, 2006.

[26] J. Shlens, M. View, and I. Introduction, "A Tutorial on Principal Component Analysis," 2014.

[27] K. Kira, L.A. Rendell, A practical approach to feature selection, in: Proceedings of the Ninth International Workshop on Machine Learning, 1992b, pp. 249–256.

[28] Urbanowicz R.J., Meeker M., Cava W., Olson R.S., Moore J.H., Relief-based feature selection: Introduction and review, Journal of Biomedical Informatics, Vol 85, pp.189-203, 2018.

[29] C.-Y. Liou, W.-C. Cheng, J.-W. Liou, and D.-R. Liou, "Autoencoder for words," Neurocomputing, vol. 139, pp. 84–96, 2014.

[30] Yang W., Wang K., Zuo W., Neighborhood Component Feature Selection for High-Dimensional Data, Journal Of Computers, Vol. 7, No. 1, January 2012.

[31] Quoc V Le, Alexandre Karpenko, Jiquan Ngiam, and Andrew Y Ng. Ica with reconstruction cost for efficient overcomplete feature

learning. Advances in Neural Information Processing Systems, 24:1017–1025, 2011.

[32] C. Cortes and V. Vapnik, "Support Vector Networks," Machine Learning, vol. 20, pp. 273-297, 1995.

[33] K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd ed. San Diego, CA: Academic, 1990.

[34] Mao, C., Hu, B., Wang, M. and Moore, P., "Learning from neighborhood for classification with local distribution characteristics", IEEE International Joint Conference on Neural Networks (IJCNN),1-8(2015).

[35] Özkaya, U., & Seyfi, L. "Dimension Forecast in Microstrip Antenna for C/X/Ku Band by Artificial Neural Network", 2nd International Symposium on Innovative Approaches in Scientific Studies, vol. 3, pp. 518-522 (2018).

[36] Özkaya, U., Öztürk, Ş., Seyfi, L., & Akdemir, B. "Non-Magnetic Materials Assignment based on Artificial Neural Network", 1st International Symposium on Innovative Approaches in Scientific Studies, vol. 2, pp. 410-415 (2018)

[37] Ozkaya, U., & Seyfi, L. (2018). A comparative study on parameters of leaf-shaped patch antenna using hybrid artificial intelligence network models. Neural Computing and Applications, 29(8), 35-45.

[38] Ozkaya, U., & Seyfi, L. (2015). Dimension optimization of microstrip patch antenna in X/Ku band via artificial neural network. Procedia-Social and Behavioral Sciences, 195, 2520-2526.