

**Research Article****A heuristic approach with artificial neural network for Parkinson's disease****Tuba Parlar** ^a ^aMustafa Kemal Üniversitesi, Bilgisayar Teknolojileri Bölümü, Antakya, Hatay, Turkey

ARTICLE INFO

Article history:

Received 30 September 2020

Accepted 23 December 2020

*Keywords:*Artificial Neural Network,
Feature Selection,
Machine Learning,
Parkinson,
Wolf Search Algorithm

ABSTRACT

Parkinson's disease is a common neurological disorder. Its symptoms are more commonly in the form of motor malfunctioning. As the disease progresses, non-motor symptoms are also observed. In previous studies, feature selection methods have been used and shown significant benefits in the diagnosis of the Parkinson's Disease in patients. Feature selection methods aim to improve the classification performance by eliminating non-valuable or less-valuable features. In this study, we aim to analyze, for diagnosing Parkinson's disease, the voice recordings of the patients with applying a recent bio-inspired optimization technique namely the Wolf Search Algorithm (WSA). WSA is a bio-inspired heuristic optimization algorithm which has been inspired by the natural behavior of wolves in daily life. We also use an artificial neural network model with feature selection methods, for the purpose of classification of the Parkinson's Disease in patients. We investigate the classification performances of the combinations of WSA-based feature selection method with well-known feature selection methods namely Information Gain and ReliefF feature selection methods. Experimental results show that ReliefF feature selection method outperform than the other feature selection method combinations for the diagnosis of the Parkinson's Disease in patients.

This is an open access article under the CC BY-SA 4.0 license.
(<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Introduction

Parkinson's disease is caused by the breakdown of cells that produce the substance called dopamine in the brain that allows brain cells to communicate with each other. Cells that produce dopamine in the brain are responsible for the control, alignment and fluency of movements.

Parkinson's disease commonly manifests itself with slowness in movements, tremor while resting, psychiatric disorders. The disease affects a population of more than 10 million people worldwide. Incidence of Parkinson's disease has been observed to increase with age. It is estimated that four percent of people with PD are diagnosed before age 50 [1].

Besides the movement related issues PD patients experience, they may have many different complaints. These complaints include fatigue, decreased cognitive functions, changes in speech, depression, anxiety, behavioural disorders, visual impairments, weight loss, sleep abnormalities and pain [2].

Early treatment slows the progression of the disease and

improves the patient's quality of life. Many different treatment methods can be applied by the specialist physician according to the stages of the disease. Firstly, it is aimed that the patient can continue his life without addiction. The decrease in dopaminergic nerve signals that develops with the loss of dopamine-producing cells is ensured to be balanced with the drug.

There are many studies based on voice recordings for the diagnosis of Parkinson's disease. These studies apply various machine learning methods such as Decision Tree, Naive Bayes, Support Vector Machines, Random Forest and Neural Network [3], [4].

Many researchers have been studied machine learning methods in different domains and approaches since they can be automatically trained and improved with the training datasets. Sakar et al. [5] collect the Parkinson's Disease Classification dataset and examine some experiments on it. They perform naïve bayes, k nearest neighbour, random forest, multilayer perception, and support vector machines classifiers using different feature subsets in the dataset. They obtain the best result using

minimum redundancy maximum relevance (mRMR) feature selection method with support vector machine (SVM) classifier with an accuracy of 86%.

Solana-Lavalle et al. [6] examine wrappers feature selection method to choose the most important features. Then they classify the features using four machine learning classifiers namely k nearest neighbour, multi-layer perceptron, support vector machine and random forest. Their study shows that an accuracy of 94.7% by using support vector machine classifier.

Almeida et al. [7] analyse eighteen feature extraction methods using four machine learning classifiers on Parkinson's disease dataset obtained from sustained phonation and speech tasks. They present the task of phonation is more efficient than speech tasks in the detection of Parkinson's disease.

Shahbakhhi et al. [4] propose an approach to diagnose Parkinson's disease based on speech analysis. They use genetic algorithm to select informative features. Then they classify the features using support vector machine algorithm. Their method achieves with accuracy 94.5% for 4 optimized features, 93.66% for 7 optimized features 94.22% for 9 optimized features.

Peker et al. [8] analyse biomedical sound measurements obtained from continuous phonation samples to diagnose PD. They apply minimum redundancy maximum relevance (mRMR) feature selection method to identify informative features. They classify these features using artificial neural network (ANN) and complex-valued artificial neural network (CVANN). They obtain 95% using ANN and 99% using CVANN in F score.

In this study, we aim to analyze the best combination of feature selection and classification methods for diagnosing Parkinson's disease. For this purpose, we investigate different combinations of two filters feature selection method with a bio-inspired search algorithm using machine learning classifiers.

The next section introduces research methodology that is used for the study. Section III presents the experimental results. Finally, Section IV concludes the paper.

2. Methodology Used

2.1. Wolf Search Algorithm

The wolf search algorithm (WSA) is an algorithm system that is inspired by the hierarchy and hunting mechanisms of wolves in nature. The algorithm was developed based on the hunting behaviour of wolf packs. Unlike other bio-inspired algorithms like particle swarm optimization [9], it is a method in which multiple wolves without a single leader go to find the best solution in multiple directions. The task of a group of wolves is to scout for prey. Each wolf in the WSA moves independently of each other to achieve a better position according to their own characteristics. Unlike the ant colony optimization algorithm [10] that uses pheromones

to communicate with each other, WSA does not engage in such communication. This results in a shorter searching run time for the WSA. While hunting, wolves try to hide themselves when they approach their prey. This feature encourages hunters in WSA to constantly change their position in an effort to find and move to better positions with less vulnerabilities. During the search, wolves simultaneously look for prey and pay attention to threats. Each wolf in the pack picks its own position, constantly moving to better positions while tracking for potential threats. With the excellent sense of smell that wolves possess they often find their prey by tracking their smell. Similar to this tracking feature, in the WSA each wolf has a visual range that creates a sensing radius. The wolves in WSA are limited by this visual range with regards to their search for food, their awareness of peers in the pack in the search for a better position, and awareness of enemies that may be close by (to jump from the visual range). When the wolves sense a prey is close, they act fast, quietly and carefully, in an effort to hide their presence from the prey.

The pseudo code of the WSA is presented in Figure 1 and summarized basic rules of the algorithm as follows:

1. Each wolf in the WSA has set a fixed visual range with the radius (r). In the hyper plane dominated by multiple features, the distance is estimated by the Minkowski distance. Each wolf can only perceive wolves in his visual range. The distance taken by the wolf at a time is usually smaller than the wolf's visual range.
2. The result of the objective function shows wolf's current location. Wolf constantly works in an effort to move to better positions. However, the wolf prefers to move to the best position that includes a companion. The wolf chooses the best position within the given options in the case where there is more than one better position taken control by a companion. If there are no better options than the wolf's current location, the wolf continues to change position randomly.
3. If the wolf detects an enemy, it will try to avoid it by escaping to a random location that is beyond the enemy's visual range.

```

Algorithm Wolf Search
1. initialize the population of wolves  $x_i$  ( $i=1,2,\dots,W$ )
2. Define and initialize parameters: ( $r$ : radius of the visual range,  $s$ : step size by which a wolf moves at a time,  $\alpha$ : velocity factor of wolf,  $p_a$ = a user defined threshold [0..1])
3. while  $t <$  generations && stopping criteria do
4.   for  $i= 1$  to  $W$  do
5.     initiative_preay()
6.     generate_location()
7.     if  $(\text{dist}(x_i, x_j) < r \ \&\& \ x_j$ 
8.        $x_i$  moves towards  $x_j$ 
9.     else if
10.       $x_i =$  initiative_preay()
11.    end if
12.    generate_location()
13.    if  $\text{rand}() > p_a$ 
14.       $x_i = x_i + \text{rand}() + v$ 
15.    end if
16.  end for
17. end while

```

Figure 1. Pseudo code of Wolf Search Algorithm.

2.2. Information Gain Feature Selection Method

Information gain feature selection method is an entropy-based feature selection method and it is calculated by the feature's contribution to overall entropy. The expected information is calculated to classify a tuple in D as follows [11]:

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

where m is the number of classes and p_i is the probability that a tuple in D . If we have to partition the samples in D on some feature A having v distinct values, $\{a_1, a_2, \dots, a_v\}$, D will split into v subsets $\{D_1, D_2, \dots, D_v\}$.

The expected information required to classify a tuple from D based on the subset on A can be calculated as follows:

$$Info_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \quad (2)$$

where the term $\frac{|D_j|}{|D|}$ is the weight of j th subset.

The information gain is defined as the difference between the original information based on the subset of classes and the obtained information after partitioning on A in the following:

$$Gain(A) = Info(D) - Info_A(D) \quad (3)$$

The information gain method ranks features according to their highest information gain score. The top ranked features are selected to reduce the feature size for better classification results.

2.3. ReliefF Feature Selection Method

ReliefF algorithm is proposed by Kononenko [12][13] by extending Relief algorithm for multi class problems. As shown in Figure 2, the algorithm by random selection picks an instance R_i then it searches for k nearest neighbours from the same class (H_j). The algorithm also searches for k nearest neighbours from each of class (nearest misses $M_j(C)$). The algorithm updates the vector $W[A]$ of estimations of the qualities of attributes depending on R_i , H_j , and $M_j(C)$. The whole process is repeated m times.

The main purpose of the ReliefF algorithm is to separate each pair of classes regardless of which two classes are closest to each other.

Algorithm ReliefF

Input: for each training instance, a vector of feature values and the class value

```

1. initialize vector  $W$ 
2. for  $i=1$  to  $m$  do
3.   randomly select a target instance  $R_i$ 
4.   find  $k$  nearest hits  $H_j$ 
5.   for each class  $C \neq class(R_i)$  do
6.     from class  $C$  find  $k$  nearest misses  $M_j(C)$ 
7.   end for
8. end for
9. for  $A=1$  to  $a$  do
10.   $W[A] = W[A] - \sum_{j=1}^k \text{diff} \frac{A_{R_i} H_j}{m.k} +$ 
       $\sum_{C \neq class(R_i)} \left[ \frac{P(C)}{1-P(class(R_i))} \sum_{j=1}^k \text{diff} (A, R_i, M_j(C)) / (m.k) \right]$ 
11. end for
12. return  $W$ 

```

Figure 2. Pseudo code of ReliefF feature selection algorithm.

2.4. Machine Learning Algorithms for Classification

In this study, we apply four machine learning classifiers, namely logistic regression, support vector machines, random forest, and artificial neural networks, and the system is evaluated using five-fold cross validation. We compare the classification results of using different feature selection methods.

Logistic regression is one of the most popular algorithms in statistics and belongs to the exponential classifiers' family. It extracts some set of features multiplying by a weight from the input dataset. If feature f_1 and feature f_2 are correlated, regression assigns half the weight to w_1 and half to w_2 . Logistic regression classifier estimates $p(c/x)$ by extracting features from the datasets as:

$$p(c|x) = \frac{\exp(\sum_{i=1}^N w_i f_i)}{\sum_c \exp(\sum_{i=1}^N w_i f_i)} \quad (4)$$

where N is the number of features, and features are a property of the observation x and output class c [14].

The main goal of the support vector machines algorithm is to find the maximum-margin hyperplane linear model that separates two classes. The maximum-margin hyperplane identifies the largest separation between two classes. As shown in Figure 3, open and filled circles represent different classes. Samples that are closest to the maximum-margin hyperplane are called support vectors. There is always at least one support vector for each class [15].

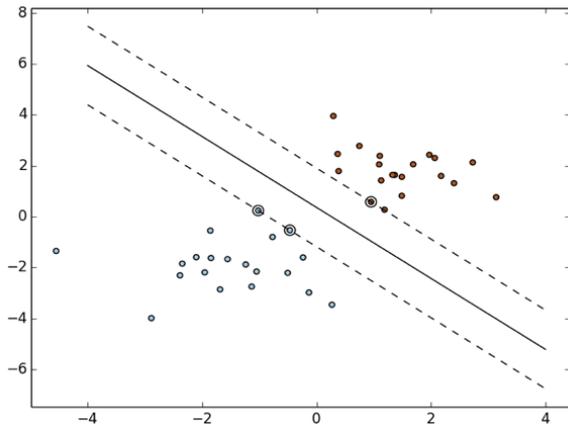


Figure 3. Maximum-margin hyperplane and margins for the SVM

Random forest is an ensemble algorithm that includes same or different kind of algorithms for classifying objects. Random forest classifier randomly selects subsets from the training set to create decision trees and classifies them based on the predictions from each of the decision trees to make an average prediction. The low correlation is important to produce ensemble predictions because de-correlated models protect from individual errors [14].

In artificial neural networks contains multiple layers of neurons. A neuron feeds all the neurons in the next layer. Sigmoid function is used to activate each neuron. This neural network is trained with the back-propagation algorithm based on recursively using gradient descent to adjust the weights. Tuning of the parameters is performed backwards from the output layer to the input layer. An artificial neural network consists of an input layer, one or more hidden layers and an output layer. Figure 4 shows an example of an artificial neural network. The only disadvantage of this method is that it is 20-700 times slower than the other methods.

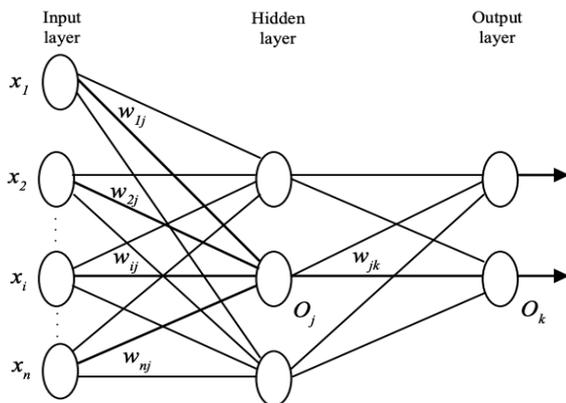


Figure 4. An artificial neural network

3. Dataset Used and Experimental Results

3.1. Datasets

We use a publicly available dataset for the study from the University of California at Irvine Machine Learning Repository by Sakar et al.[5]. Parkinson's Disease

Classification dataset is collected from the voice records of 252 patients 188 of which are Parkinson's disease at the Department of Neurology in Cerrahpaşa Faculty of Medicine, Istanbul University. All the features in the dataset are excluded from the speculum of different voice signals. In Table I, feature groups are summarized with their descriptions [6].

Table 1. Feature set of the Parkinson's disease dataset

Features	Description
Baseline	Jitter variants, shimmer variants, mean, median, standard deviation, minimum, maximum, harmonicity parameters, pitch period entropy, detrended fluctuation analysis, recurrence period density entropy.
Time frequency	Bandwidth, formant frequencies, intensity parameters.
MFCCs	Mel frequency cepstral coefficients.
Wavelet transform	Wavelet transform features related with F0.
Vocal fold	Vocal fold excitation ratio, glottal to noise excitation ratio, glottal to noise decomposition, empirical mode decomposition, glottis quotient.

3.2. Performance Evaluation

The performance of the classification experiments was evaluated with the F-score. The F-score is an evaluation criterion that takes into account both the precision and recall of the classifier [11]. Precision (P) is an evaluation criterion about how sensitive the classifier is in the classifications made for a class. Recall (R) is the criterion that evaluates the proper classification of samples. The F -score is the harmonic mean of precision and recall values as follows:

$$F - score = 2 * \frac{P * R}{P + R} \tag{5}$$

3.3. Experimental Results and Discussions

In this study, we investigate which feature selection method or combination produces the most accurate classification result. Firstly, we use some common feature selection methods namely information gain and reliefF methods. Secondly, we investigate using a heuristic method to select the most important features with these commonly used feature selection methods. For this purpose, we use Wolf Search Algorithm to select the most important features. Finally, we apply information gain and reliefF feature selection methods on these features.

Furthermore, we apply logistic regression (LR), support vector machines (SVM), random forest (RF) and artificial neural network (ANN) to classify the PD dataset. We use the classifiers from the Weka data mining tool version 3.8.1. We evaluate the classification performance by using the micro-average of F-score with five-fold cross validations.

As you can observe from Tables 2-5, ANN has two wildcards (a) and (a, t). In Weka, multilayer perception function uses some wildcard values as following:

$$a = (\# \text{ of attributes} + \# \text{ of classes})/2$$

$$t = \# \text{ of attributes} + \# \text{ of classes} \quad (6)$$

ANN(a) means we have (a) units in the hidden layer of the neural network. ANN(a, t) means we have (a) units in the first hidden layer and (t) units in the second hidden layer of the neural network. We apply ANN classifier to the feature sizes less than 300 because of computation times needed.

ReliefF algorithm performs feature selection by the contribution of k nearest neighbours. In this paper, we set $k=10$ as suggested in [13]. We also use default parameters in weka for WSA as suggested in [16].

For each feature selection method, we tried five feature sizes at 100, 200, 300, 400, and 500. As observed in Tables 2-5, reliefF feature selection method is the best performer with ANN classifier. ANN method achieved 0.925 F-score with 100 feature size. Although WSA is a robust method [16] it does not do well among the other methods. The disadvantage of ANN method is this method needs 20-700 times more computation times than the other methods. As you can see in Tables 2 and 3 we do not apply ANN classifier to the feature sizes more than 300 because of computation times needed.

Table 2. Information gain feature selection method

Feature size	LR	SVM	RF	ANN(a)	ANN(a,t)
100	0.831	0.852	0.861	0.865	0.849
200	0.821	0.864	0.876	0.863	0.883
300	0.795	0.866	0.875	-	-
400	0.786	0.855	0.866	-	-
500	0.786	0.851	0.874	-	-
All features	0.785	0.850	0.847	-	-

Table 3. ReliefF feature selection method

Feature size	LR	SVM	RF	ANN(a)	ANN(a,t)
100	0.846	0.861	0.878	0.925	0.898
200	0.829	0.867	0.883	0.920	0.903
300	0.823	0.861	0.879	-	-
400	0.807	0.856	0.871	-	-
500	0.786	0.851	0.874	-	-
All features	0.785	0.850	0.847	-	-

Table 4. Wolf Search and ReliefF feature selection methods

Feature size	LR	SVM	RF	ANN(a)	ANN(a,t)
100	0.833	0.858	0.876	0.901	0.874
200	0.816	0.865	0.880	0.897	0.884
All features (301)	0.819	0.864	0.865	0.818	0.873

Table 5. Wolf Search and information gain feature selection methods

Feature size	LR	SVM	RF	ANN(a)	ANN(a,t)
100	0.824	0.849	0.875	0.867	0.860
200	0.796	0.867	0.870	0.823	0.892
All features (301)	0.819	0.864	0.865	0.818	0.873

4. Conclusions

Parkinson's disease is a common neurological disorder that affects the patients in many different complaints such as changes in speech. Many researchers study applying machine learning methods to diagnose PD based on the voice recordings. We examined the impacts of different filters feature selection methods with a bio-inspired optimization method namely Wolf Search Algorithm (WSA). WSA is a heuristic algorithm which has been inspired by the behaviours of wolf packs.

For feature selection, we obtained the most accurate performance classification results with reliefF method among all the other methods. When running the four classifiers, ANN classifier presented the most accurate classification results of all classifiers for the PD dataset. ANN classifier achieved 0.925 F-score with 100 feature size. As can be observed in Tables 4 and 5, WSA selects 301 features and works well compared with information gain method with the same feature sizes. Although WSA is a robust method [16] it does not do well among the other methods. ReliefF algorithm also works well when it performs with WSA.

Our experimental results show that ReliefF method with ANN achieved the best performance with all other feature selection methods. The only disadvantage of ANN method is this method needs more computation times than the other methods.

Author's Note

Abstract version of this paper was presented at 9th International Conference on Advanced Technologies (ICAT'20), 10-12 August 2020, Istanbul, Turkey with the title of "A Heuristic Approach with Artificial Neural Network For Parkinson's Disease".

References

- [1] "Statistics | Parkinson's Foundation". Available at: <https://www.parkinson.org/Understanding-Parkinsons/Statistics>. [Erişim: Tem-2020].
- [2] C. Blauwendraat, M. A. Nalls, ve A. B. Singleton, "The genetic architecture of Parkinson's disease", *The Lancet Neurology*, c. 19, sayı 2. Lancet Publishing Group, ss. 170–178, 01-Şub-2020.
- [3] M. Belić, V. Bobić, M. Badža, N. Šolaja, M. Đurić-Jovičić, ve V. S. Kostić, "Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease—A review",

- Clinical Neurology and Neurosurgery*, c. 184. Elsevier B.V., s. 105442, 01-Eyl-2019.
- [4] M. Shahbakhi, D. T. Far, ve E. Tahami, “Speech Analysis for Diagnosis of Parkinson’s Disease Using Genetic Algorithm and Support Vector Machine”, *J. Biomed. Sci. Eng.*, c. 07, sayı 04, ss. 147–156, 2014.
- [5] C. O. Sakar *vd.*, “A comparative analysis of speech signal processing algorithms for Parkinson’s disease classification and the use of the tunable Q-factor wavelet transform”, *Appl. Soft Comput. J.*, c. 74, ss. 255–263, Oca. 2019.
- [6] G. Solana-Lavalle, J. C. Galán-Hernández, ve R. Rosas-Romero, “Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features”, *Biocybern. Biomed. Eng.*, c. 40, sayı 1, ss. 505–516, Oca. 2020.
- [7] J. S. Almeida *vd.*, “Detecting Parkinson’s disease with sustained phonation and speech signals using machine learning techniques”, *Pattern Recognit. Lett.*, c. 125, ss. 55–62, Tem. 2019.
- [8] M. Peker, B. Şen, ve D. Delen, “Computer-aided diagnosis of Parkinson’s disease using complex-valued neural networks and mRMR feature selection algorithm”, *J. Healthc. Eng.*, c. 6, sayı 3, ss. 281–302, Ağu. 2015.
- [9] S. Kiranyaz, “Particle swarm optimization”, *Adapt. Learn. Optim.*, c. 15, ss. 45–82, 2014.
- [10] M. Dorigo ve C. Blum, “Ant colony optimization theory: A survey”, *Theor. Comput. Sci.*, c. 344, ss. 243–278, 2005.
- [11] J. Han, M. Kamber, ve J. Pei, “Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)”, 2011.
- [12] M. Robnik-Šikonja ve I. Kononenko, “Theoretical and Empirical Analysis of ReliefF and RReliefF”, *Mach. Learn.*, c. 53, sayı 1/2, ss. 23–69, 2003.
- [13] I. Kononenko, “Estimating Attributes: Analysis and Extensions of RELIEF”. in Proceedings of the Seventh European Conference on Machine Learning. 1994, ss. 171–182, Springer-Verlag.
- [14] T. Hastie, R. Tibshirani, ve J. Friedman, “Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction”.
- [15] I. Witten, E. Frank, M. Hall, C.J. Pall, “Data Mining: Practical Machine Learning Tools and Techniques”. Available at [Erişim: Aralık-2020]: <https://www.cs.waikato.ac.nz/ml/weka/book.htm>.
- [16] S. Fong, R. P. Biuk-Aghai, R. C. Millham: Swarm Search Methods in Weka for Data Mining. In: 10th International Conference on Machine Learning and Computing (ICMLC 2018), Macau, China, 2018.