



You have downloaded a document from  
**RE-BUŚ**  
repository of the University of Silesia in Katowice

**Title:** Diagnosing Parkinson's disease by means of ensemble classification of patients' voice samples

**Author:** Krzysztof Wróbel

**Citation style:** Wróbel Krzysztof. (2021). Diagnosing Parkinson's disease by means of ensemble classification of patients' voice samples. "Procedia Computer Science" (Vol. 192 (2021), s. 3905-3914), DOI:10.1016/j.procs.2021.09.165



Uznanie autorstwa - Użycie niekomercyjne - Bez utworów zależnych Polska - Licencja ta zezwala na rozpowszechnianie, przedstawianie i wykonywanie utworu jedynie w celach niekomercyjnych oraz pod warunkiem zachowania go w oryginalnej postaci (nie tworzenia utworów zależnych).



UNIwersYTET ŚLĄSKI  
W KATOWICACH



Biblioteka  
Uniwersytetu Śląskiego



Ministerstwo Nauki  
i Szkolnictwa Wyższego



25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

# Diagnosing Parkinson's disease by means of ensemble classification of patients' voice samples

Krzysztof Wrobel\*

*University of Silesia, Faculty of Science and Technology, Institute of Computer Science, Bedzinska 39, 41-200 Sosnowiec*

---

## Abstract

This paper proposes a method for diagnosing Parkinson's disease using ensemble classification of patient voice samples. Conducted research concerned testing the parameters of the ensemble of classifiers, in terms of types and numbers of classifiers included in it. More than a dozen popular classifiers were considered in the study. Additionally, for each of the tested classifiers, a set of features of voice samples were selected, for which a given classifier showed the highest classification efficiency. The Sequential Backward Selection (SBS), which belongs to the wrapper methods, was used for feature selection. The ensemble of classifiers was then tested in two cases, with all features considered and including only those indicated by the SBS method. The obtained results were compared with each other.

All experiments were performed on a publicly available database containing voice samples of Parkinson's patients and healthy patients. This database can be found in the UCI archives.

© 2021 The Authors. Published by Elsevier B.V.

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

Peer-review under responsibility of the scientific committee of KES International.

*Keywords:* Parkinson's disease; Medical diagnosis; Ensemble classification; Feature selection

---

## 1. Introduction

Parkinson's disease (PD) is a disease of the central nervous system. It develops when there is a decrease in brain's nerve cells that produce dopamine. Dopamine is a chemical messenger that transmits impulses between nerve cells to control body movements and balance. Thus, PD is a neurological condition whose main symptoms are tremors that occur, slowed movements, and difficulty in movement [1, 2].

---

\* Corresponding author. Tel.: +48 32 3689766.

*E-mail address:* [krzysztof.wrobel@us.edu.pl](mailto:krzysztof.wrobel@us.edu.pl)

The risk of developing PD increases with age, and the first symptoms of the disease can occur at almost any age [3]. It is therefore important to detect Parkinson's disease as early as possible to initiate immediate therapy to stop or slow the disease process.

Parkinson's disease also affects other areas of activity, such as speaking. Common problems include a weak, monotone, or nasal voice, slow speech, difficulty in starting speech, rhythm or accent disorders, and stuttering [4, 5]. It is therefore possible to diagnose Parkinson's disease using human voice analysis [6]. The advantage of this method is that it is non-invasive and non-contact. Voice samples can be pre-recorded and properly prepared, then analysed at any time, without the use of specialized, and often expensive, medical equipment. Recording of samples does not have to take place in specialized hospital departments but can be done at home, and the samples can be sent for testing, e.g., via e-mail [7]. The registration and diagnosis process can also be done online via the Internet [8]. From the recorded sample, appropriate voice characteristics can be extracted and subjected to a classification process. As a result, it is possible to determine whether the analyzed sample comes from a sick or healthy person.

With the emergence of the COVID-19 virus pandemic, non-contact voice recording has become particularly important, and is an additional advantage of the method. The lack of contact with the recording device means that the likelihood of a sick person transmitting the virus to such a device is reduced, and in the case of healthy people, infection with the virus is limited. This prevents the spread of the pandemic.

The above-presented advantages were motivation for developing a method that would effectively diagnose Parkinson's disease using patients' voice samples. The proposed method uses an ensemble of classifiers [9, 10, 11] and a feature reduction method [12, 13, 14] to minimize the number of features. In the following section, it is shown that a properly configured ensemble of classifiers along with a reduced set of voice sample features enables the correct diagnosis of Parkinson's disease.

This paper is organized as follows: Section 2 presents the theoretical background related to the classification of patient voice samples, using both a single classifier and an ensemble of classifiers. Section 3 focuses on the feature selection method, Sequential Backward Selection. For each classifier, the voice features for which it shows the highest classification performance are indicated using this method. Section 4 outlines the tests performed, and the analysis of the results obtained. And Section 5 sums up the conclusions and provides a summary of the work and further possible directions for development.

## 2. Theoretical background

Let a set of  $m$  patients  $P = \{P_1, P_2, \dots, P_m\}$  be given. For each patient  $P_i$ , where  $i = 1, \dots, m$ , a set of  $n$  samples of his voice  $P_i = \{S_1, S_2, \dots, S_n\}$  is given. Each sample  $S_j$ , where  $j = 1, \dots, n$ , is described by a set of  $r$  features  $S_j = \{f_1, f_2, \dots, f_r\}$ . Based on the analysis of the set of voice samples of patient  $P_i$  and the features of these samples, the task is to diagnose whether patient  $P_i$  is healthy or has Parkinson's disease.

Let us introduce the class  $C = \{0, 1\}$  to which patient  $P_i$  belongs. Class 0 indicates that the patient is healthy, while class 1 indicates Parkinson's disease. Additionally, let  $c_{ij} \in C$  denote the class of a single sample  $j$  of patient  $i$ . Let us denote the decision that the sample  $S_{ij}$  belongs to class  $c_{ij} \in C$  as:

$$d_{ij} = \Psi(S_{ij}), \quad (1)$$

where  $\Psi$  denotes the classifier model, and  $S_{ij}$ , the  $j$ -th sample, from the  $i$ -th patient.

$N$  samples of patient  $P_i$  are analyzed, and each sample is classified separately. The final decision  $d_i^{final} \in C$  whether patient  $P_i$  has Parkinson's disease or is healthy is made based on majority voting:

$$d_i^{final} = \begin{cases} 0, & \text{if } d_{ij}^0 \geq d_{ij}^1 \\ 1, & \text{if } d_{ij}^1 > d_{ij}^0 \end{cases}, \quad (2)$$

where:

$d_{ij}^1 = \sum_{j=0}^n d_{ij}$  - is the decision that the sample is from a patient with Parkinson's disease,

$d_{ij}^0 = n - d_{ij}^1$  - is the decision that the sample is from a healthy patient.

2.1. A single classifier approach

In this approach, the decision of whether a patient is healthy or has Parkinson’s disease is made using a single classifier. First, a single classifier  $\Psi$  is trained on a set of voice samples present in the database but from patients other than the diagnosed patient  $P_i$ . Then, the classifier is tested on a set of  $n$  voice samples  $S$  of patient  $P_i$ . Each sample  $S_j$ , where  $j = 1, \dots, n$ , is classified separately. And finally, a decision is made as to whether the diagnosed patient is healthy or has Parkinson’s disease. This is determined by using a majority voting system. The flow chart of the single classifier approach is shown in Figure 1.

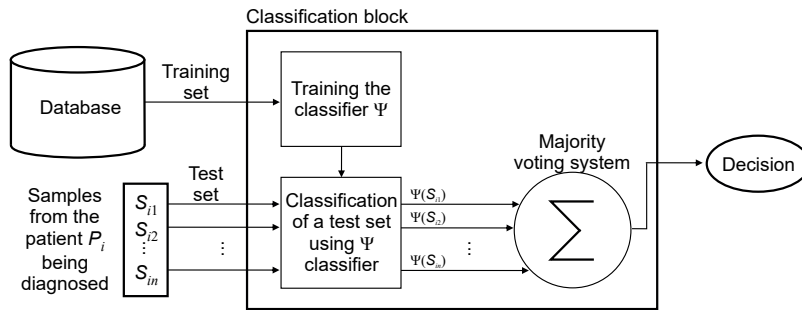


Fig. 1. The flow chart of decision determination in a single classifier approach.

The presented approach can be extended to the feature selection method. For this part of the test, the popular Sequential Backward Selection (SBS) method was used, which is described in detail in Section 3. The feature selection was performed on both the voice samples from the training set and the samples from the test set. This process leaves only selected features of voice samples in both the training and test sets for the further classification process.

2.2. The proposed ensemble approach

In the proposed approach, the ensemble of classifier is built based on different types of  $\Psi$  classifiers. Each classifier  $\Psi$  was trained by use of the same learning set. Such a situation is shown in Figure 2, where  $k$  classification blocks are given. In each of these  $k$  blocks, there was a classifier  $\Psi_i$ , where  $i = 1, \dots, k$ . Furthermore,  $\Psi_1 \neq \Psi_2 \neq \dots \neq \Psi_k$ , where  $\Psi_1, \Psi_2, \dots, \Psi_k$  are the classifiers located in blocks 1, 2, ...,  $k$  respectively.

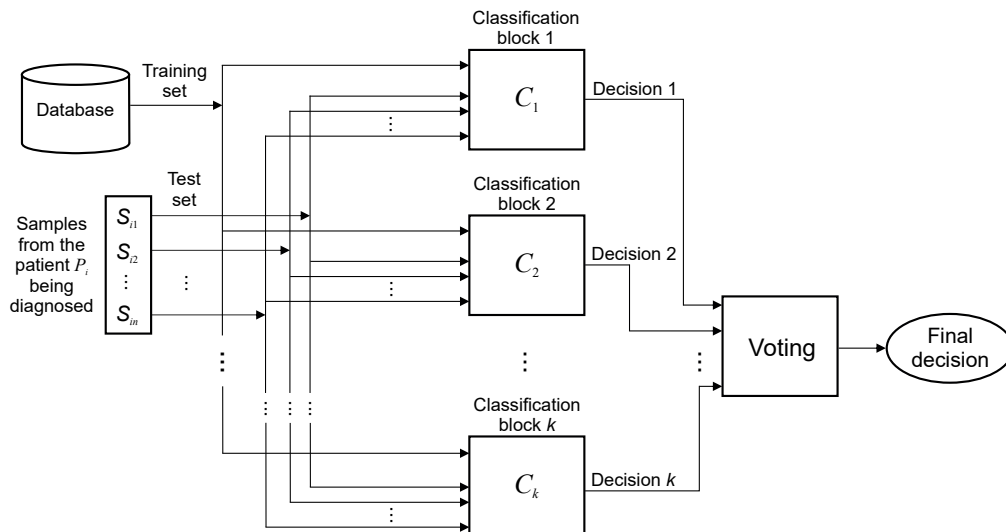


Fig. 2. The flow chart of decision determination in the proposed ensemble approach.

Each of the classification blocks generates a  $d^C$  decision. So, an ensemble of  $k$  classification blocks will generate  $d^{C_1}, \dots, d^{C_k}$  decisions. Having these data, the final decision  $D$  is determined from the following formula:

$$D = \begin{cases} 0, & \text{if } \sum_{i=1}^k d^{C_i} \leq t \\ 1, & \text{if } \sum_{i=1}^k d^{C_i} > t \end{cases}, \tag{3}$$

where:

$d^{C_i}$  - is the decision generated by the  $i$ -th classification block,

$t = [k/2, \dots, k - 1]$  - is a parameter that determines how many classification blocks determine whether a patient has Parkinson's disease or is healthy.

The value of parameter  $t$  was examined in the Experiments and results section.

### 3. Feature selection

The Sequential Backward Selection (SBS) algorithm was used for feature selection. Its choice was dictated by its simple implementation and speed. However, it should be noted that this algorithm searches only a subset of combinations of all features, smaller than all possibilities, which does not ensure the selection of optimal features [13, 15]. this method proved to be sufficient.

The SBS algorithm belongs to the group of wrapper methods. It means that to find the set of optimal features it uses a specific machine learning algorithm [14]. It belongs to the family of greedy feature space search algorithms and is used to reduce the initial  $d$ -dimensional feature space  $S$  to its  $r$ -dimensional subspace  $S_r$ , where  $r < d$ . The input parameter of the method is a given number of features contained in the new subspace.

The working of the SBS algorithm is explained using an example. Subspace  $S_r$  is initially set as the full feature space  $S$ . Then, in each iteration step, a feature is eliminated, whose removal results in the highest classification accuracy. Completion of the algorithm occurs when a set number of output features is reached. An example of an SBS algorithm, for five input features and two output features, is presented in Figure 3.

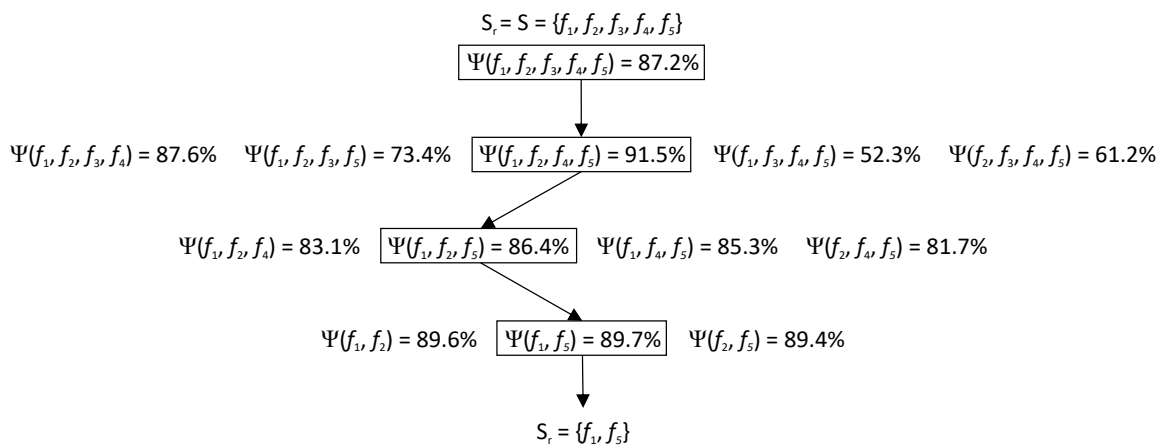


Fig. 3. The principle of operation of the SBS algorithm.

This algorithm was used for all the classifiers tested and an optimal set of features was selected for each classifier. The input parameter, denoting the number of features, was given a value of 1. This meant that the algorithm had to analyze a "path" starting from a set containing all the features to a set containing only one feature. This gave a set of results from which the best result was chosen and the features for which it was achieved. In Figure 3, the best result was achieved for 4 features  $\Psi(f_1, f_2, f_4, f_5) = 91.5\%$ .

## 4. Experiments and results

### 4.1. Database

The experiments used a patient database from the UCI archives, available at <https://archive.ics.uci.edu/ml/index.php> [16]. The tested database contains voice samples from 32 people - 24 people were Parkinson's patients, 8 people were healthy. For 29 of the individuals, 6 voice samples were recorded, while for 3 individuals, 7 voice samples were recorded. A total of 195 feature vectors are included in the database out of which 147 vectors are labelled as feature vectors from diseased persons, while 48 are vectors from healthy persons. Each voice sample is described by the features in Table 1.

Table 1. Analyzed features of voice samples.

No.	Feature	Description
1.	MDVP:Fo(Hz)	Average vocal fundamental frequency.
2.	MDVP:Fhi(Hz)	Maximum vocal fundamental frequency.
3.	MDVP:Flo(Hz)	Minimum vocal fundamental frequency.
4.	MDVP:Jitter(%),	Several measures of variation in fundamental frequency.
5.	MDVP:Jitter(Abs),	
6.	MDVP:RAP,	
7.	MDVP:PPQ,	
8.	Jitter:DDP	
9.	MDVP:Shimmer,	Several measures of variation in amplitude.
10.	MDVP:Shimmer(dB),	
11.	Shimmer:APQ3,	
12.	Shimmer:APQ5,	
13.	MDVP:APQ,	
14.	Shimmer:DDA	
15.	NHR,	Two measures of ratio of noise to tonal components in the voice.
16.	HNR	
17.	RPDE,	Two nonlinear dynamical complexity measures.
18.	D2	
19.	DFA	Signal fractal scaling exponent.
20.	spread1,	Three nonlinear measures of fundamental frequency variation.
21.	spread2,	
22.	PPE	

### 4.2. Tested classifiers and their parameters

The experiments were performed with software written in Python (version 3.8) using the Scikit-learn library (version 0.24.1). This library contains implementations of popular and frequently used classifiers [17]. For some of the classifiers, the classification accuracy was enhanced by the method of pre-scaling the data. The Standard Scaler class method of the Scikit-learn library was used for scaling data. This method is described in detail in [17].

The classification performance was tested using the leave-one-out method, i.e., records belonging to one patient were removed from the database and treated as test data. The records remaining in the database were treated as training data. The procedure was repeated for each patient in the database.

Tested classifiers, their parameters, and whether the tested data were scaled or not are shown in Table 2.

### 4.3. Experiment 1

In the first conducted experiment, the classification accuracy was determined for single classifiers without feature reduction and with feature reduction using the SBS algorithm. Table 3 shows the obtained results.

Table 2. Tested classifiers and their parameters.

Classifier	Input parameters	Data scaling
Perceptron	max_iter=1000	Yes
Logistic Regression	solver='liblinear'	Yes
Multi-layer Perceptron	max_iter=1000	Yes
SVM (linear kernel)	kernel='linear'	Yes
SVM (non-linear kernel)	kernel='rbf', gamma='auto'	Yes
SVM (polynomial kernel)	kernel='poly', gamma='auto', degree=1	Yes
Decision Tree	criterion='entropy', max_depth=5	No
k-NN	n_neighbors=10, metric='euclidean', weights='distance'	Yes
Naive Bayes	default	No
Stochastic Gradient	max_iter=1000	No
Gaussian Process	default	Yes
Random Forest	default	Yes
Ada Boost	default	No
Gradient Boosting	default	No
Extra Trees	default	No
Bagging	DecisionTreeClassifier	No
Linear Discriminant	default	No
Quadratic Discriminant	default	No

Table 3. Classification accuracy for single classifier without and with feature reduction.

Classifier	Without feature reduction [%]	With feature reduction [%]
Perceptron	75.0	87.5
Logistic Regression	81.2	84.4
Multi-layer Perceptron	90.6	90.6
SVM (linear kernel)	81.2	87.5
SVM (non-linear kernel)	84.4	87.5
SVM (polynomial kernel)	84.4	87.5
Decision Tree	84.4	90.6
k-NN	81.2	90.6
Naive Bayes	68.8	90.6
Stochastic Gradient	78.1	90.6
Gaussian Process	84.4	87.5
Random Forest	81.2	87.5
Ada Boost	81.2	87.5
Gradient Boosting	84.4	87.5
Extra Trees	84.4	84.4
Bagging	84.4	87.5
Linear Discriminant	81.2	90.6
Quadratic Discriminant	81.2	87.5

Analyzing Table 3 it can be observed, that practically, for each of the tested classifiers, higher accuracy was obtained by a reduced set of features. The exceptions are the Multi-layer Perceptron and Extra Trees classifier, where the same classification accuracy was obtained for the whole and reduced set of features. The best accuracy in classification of all features was achieved using a Multi-layer Perceptron classifier and it was 90.6%. Whereas in the case of feature reduction the best result of 90.6% was achieved for Multi-layer Perceptron, Decision tree, k-NN, Naive Bayes, Stochastic Gradient Descent and Linear Discriminant classifiers.

4.4. Experiment 2

Based on single classifiers, ensembles of classifiers were constructed in the second experiment conducted. The ensembles contained respectively: 3, 5, and 7 different classifiers.

The number of classifier combinations in an ensemble can be calculated from the following formula:

$$l_k = \binom{n}{k} \tag{4}$$

where:

$n$  - the pool of all classifiers. In the presented work,  $n = 18$  classifiers were tested (see Table 2),

$k$  - number of classifiers in the ensemble. Respectively,  $k = 3, 5, 7$ .

Similarly, to Experiment 1, tests were performed for a set of all features (without reduction) and a reduced set of features.

By substituting the values of  $n = 18$  and  $k = 3, 5, 7$ , into equation (4), we obtain the number of all tested combinations:  $l_3 = 816, l_5 = 8568, l_7 = 31824$ . Due to such a large number of combinations, the study shows the probability of obtaining a certain classification accuracy for a given configuration and different parameters  $t$ , without specifying the classifiers with which a given result was obtained.

The values of the parameter  $t$  were dependent on the number of classifiers in the ensemble. Thus, for  $k = 3, t = 1, 2, k = 5, t = 2, \dots, 4$ , and  $k = 7, t = 3, \dots, 6$ .

Figure 4 shows the probability of obtaining an accuracy of classification for 3 – Fig. 4(a), 5 – Fig. 4(b), and 7 – Fig. 4(c) classifiers in the ensemble, without feature reduction. Also, the influence of the parameter  $t$  (see equation 3) on the obtained results was investigated.

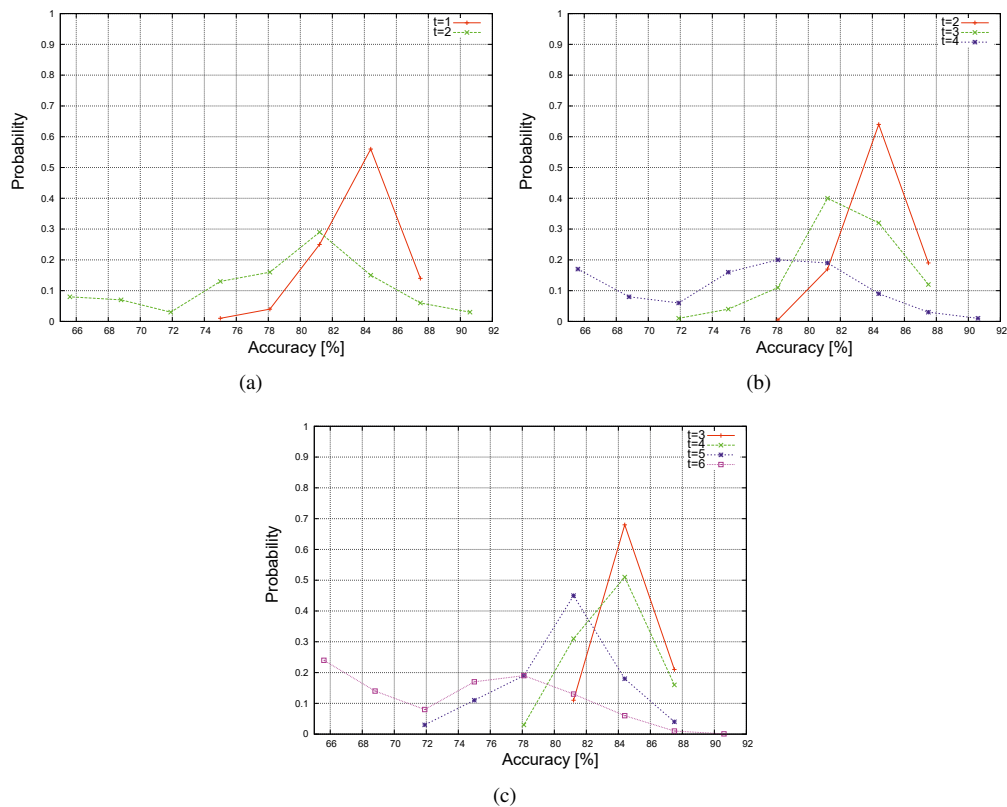


Fig. 4. Probability of obtaining a given classification accuracy and the influence of the parameter  $t$  on the obtained results for an ensemble of (a) 3, (b) 5, (c) 7 classifiers. Results for the case without feature reduction.



Analyzing the presented charts, it can be observed that the highest classification accuracy was achieved for the maximum value of the parameter  $t$ . It was 90.6% accuracy for all tested ensembles. For each ensemble, the most probable classification accuracy was 84.4% for a minimum value of the parameter  $t$ . Thus, it can be concluded that building an ensemble of classifiers without feature reduction does not significantly improve the classification accuracy compared to using single classifiers with and without feature reduction. The highest classification accuracy of 90.6% was also obtained for these cases.

Figure 5 shows the probability of obtaining the specified classification accuracy for 3 – Fig. 5(a), 5 – Fig. 5(b), and 7 – Fig. 5(c) classifiers in the ensemble with feature reduction. As in the previous case, the influence of the parameter  $t$  on the obtained results was investigated.

By analyzing the charts shown in Figure 5, it can be observed that the classification accuracy increases with increasing the parameter  $t$ . For an ensemble of 3 classifiers, the maximum accuracy was 96.9%, while for ensembles of 5 and 7 classifiers, the highest accuracy was 100%. For each ensemble, the most probable classification accuracy was 87.5% for a minimum value of the parameter  $t$ .

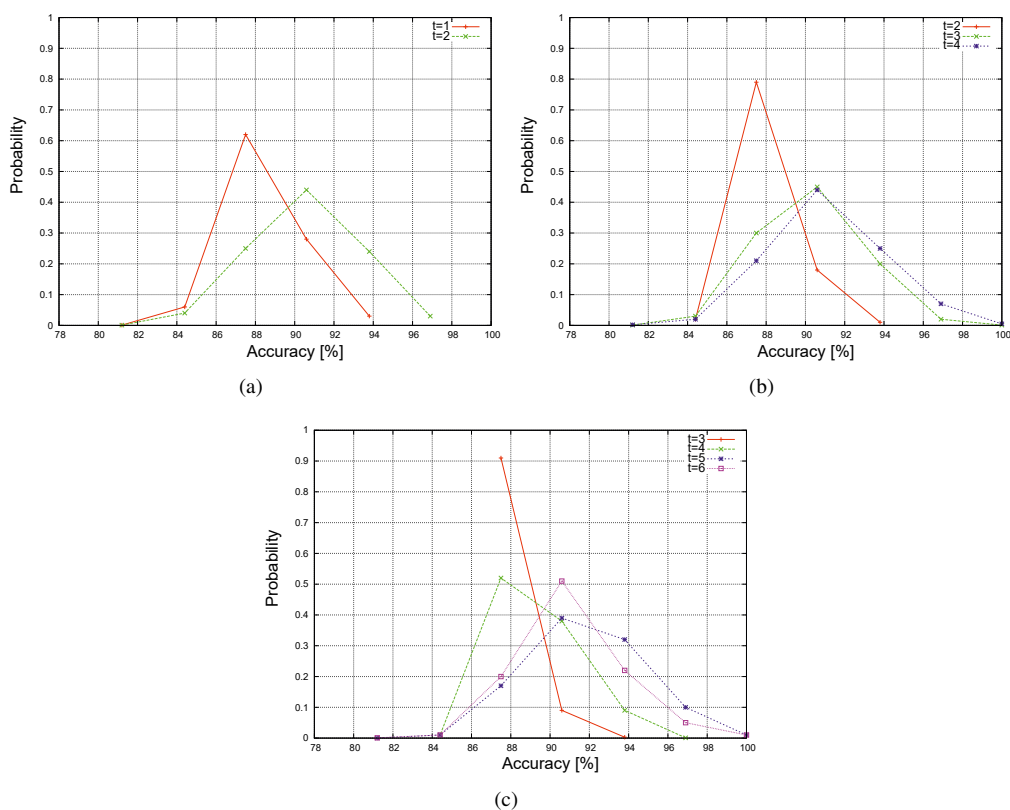


Fig. 5. Probability of obtaining a given classification accuracy and the influence of the parameter  $t$  on the obtained results for an ensemble of (a) 3, (b) 5, (c) 7 classifiers. Results for the case with feature reduction.

This result is higher than for the case in which the feature set was not reduced. The conclusion is that building an ensemble of classifiers with feature reduction improves the classification effectiveness over using an ensemble of classifiers without feature reduction and single classifiers (without and with feature reduction). One hundred percent classification accuracy was obtained when the classifiers were properly configured, and the parameter  $t$  was carefully chosen.

Table 4 presents a summary of the study performed in this experiment. The best classification performance obtained for each configuration is highlighted.

Table 4. Summary of results performed for ensembles of classifiers.

Number of classifiers in an ensemble	The value of the parameter t	Without feature reduction		With feature reduction	
		Most probable result [%]	The best result obtained [%]	Most probable result [%]	The best result obtained [%]
3	1	84.4	87.5	87.5	93.8
	2	81.2	90.6	90.6	96.9
5	2	84.4	87.5	87.5	93.8
	3	81.2	87.5	90.6	100.0
	4	78.1	90.6	90.6	100.0
7	3	84.4	87.5	87.5	93.8
	4	84.4	87.5	87.5	96.9
	5	81.2	87.5	90.6	100.0
	6	65.6	90.6	90.6	100.0

## 5. Conclusions

During the COVID-19 virus pandemic, non-contact recording of voice samples and their use to diagnose Parkinson's patients is an important advantage. This paper proposes a method for diagnosing PD using ensemble classification of patient voice samples.

The conducted research has shown that ensemble classification allows increasing the effectiveness of classification in comparison to classification using a single classifier, but on the condition that features reduction has been performed. Considering an ensemble of 5 or 7 classifiers and an appropriate choice of t parameter, it is possible to achieve classification effectiveness equal even to 100%. However, the presented research does not indicate which classifiers allow to obtain such a high result. The automatic selection of classifiers, using which the maximum accuracy can be achieved, will be the subject of further research. Further work will also include testing other feature selection/reduction methods. The research will also be conducted on other databases.

## References

- [1] Golbe, L.I., Mark, M.H., Sage, J.I. (2010) "Parkinson's disease handbook.", The American Parkinson Disease Association, Inc.
- [2] Parkinson, J. (1817) "An essay on the shaking palsy.", London.
- [3] Grosset, D., Fernandez, H., Grosset, K., Okun, M. (2009) "Parkinson's Disease Clinician's Desk Reference.", CRC Press.
- [4] Rouzbahani, H.K., Daliri, M.R. (2011) "Diagnosis of Parkinson's Disease in Human Using Voice Signals." *Basic and Clinical Neuroscience* **2** (3): 12-20.
- [5] Tsanas, A., Little, M.A., McSharry, P.E., Ramig, L.O. (2011) "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** (59): 842-55.
- [6] Froelich, W., Wrobel, K., Porwik, P. (2015) "Diagnosis of Parkinson's disease using speech samples and threshold-based classification." *Journal of Medical Imaging and Health Informatics* **5** (6): 1358-1363.
- [7] Sakar, B.E., Kursun, O. (2014) "Telemonitoring of changes of unified Parkinson's disease rating scale using severity of voice symptoms." *In Proc. 2nd International Conference on E-Health and TeleMedicine*, Istanbul, 114-119.
- [8] Tsanas, A., Little, M.A., McSharry, P.E., Ramig, L.O. (2010) "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests." *IEEE Transactions on Biomedical Engineering* **57** (4): 884-93.
- [9] Opitz, D., Maclin, R. (1999) "Popular ensemble methods: An empirical study." *Journal of Artificial Intelligence Research* **11**: 169-198.
- [10] Polikar, R. (2006) "Ensemble based systems in decision making." *IEEE Circuits and Systems Magazine* **6** (3): 21-45.
- [11] Porwik, P., Doroz, R., Wrobel, K. (2019) "An ensemble learning approach to lip-based biometric verification, with a dynamic selection of classifiers." *Expert Systems with Applications* **115**: 673-683.
- [12] Porwik, P., Doroz, R. (2014) "Self-adaptive biometric classifier working on the reduced dataset." *Hybrid Artificial Intelligence Systems (HAIS), Book Series: Lecture Notes in Computer Science* **8480**: 377-388.

- [13] Wrobel, K. (2019) "Diagnosing Parkinson's Disease with the Use of a Reduced Set of Patients' Voice Features Samples." *Computer Information Systems and Industrial Management (CISIM)*. *Book Series: Lecture Notes in Computer Science* **11703**: 84-95.
- [14] Pudil, P., Novovicova, J., Kittler, J. (1994) "Floating search methods in feature selection." *Pattern recognition letters* **15** (11): 1119-1125.
- [15] Palys, M., Doroz, R., Porwik, P. (2013) "On-line signature recognition based on an analysis of dynamic feature." *IEEE International Conference on Biometrics and Kansei Engineering (ICBAKE)*: 103-107.
- [16] Little, M.A., McSharry, P.E., Roberts, S.J., Costello, D.A., Moroz, I.M. (2007) "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection." *Biomedical Engineering Online* **6** (1): 23.
- [17] Pedregosa, F., et al. (2011) "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* **12**: 2825-2830.