Medical & Biological Engineering & Computing
https://doi.org/10.1007/s11517-020-02260-3

ORIGINAL ARTICLE

Evaluation of train and test performance of machine learning algorithms and Parkinson diagnosis with statistical measurements

Emre Avuçlu[1] · Abdullah Elen[2]

Received: 5 March 2020 / Accepted: 29 August 2020

Abstract
Parkinson's disease is a neurological disorder that causes partial or complete loss of motor reflexes and speech and affects thinking, behavior, and other vital functions affecting the nervous system. Parkinson's disease causes impaired speech and motor abilities (writing, balance, etc.) in about 90% of patients and is often seen in older people. Some signs (deterioration of vocal cords) in medical voice recordings from Parkinson's patients are used to diagnose this disease. The database used in this study contains biomedical speech voice from 31 people of different age and sex related to this disease. The performance comparison of the machine learning algorithms k-Nearest Neighborhood (k-NN), Random Forest, Naive Bayes, and Support Vector Machine classifiers was performed with the used database. Moreover, the best classifier was determined for the diagnosis of Parkinson's disease. Eleven different training and test data (45 × 55, 50 × 50, 55 × 45, 60 × 40, 65 × 35, 70 × 30, 75 × 25, 80 × 20, 85 × 15, 90 × 10, 95 × 5) were processed separately. The data obtained from these training and tests were compared with statistical measurements. The training results of the k-NN classification algorithm were generally 100% successful. The best test result was obtained from Random Forest classifier with 85.81%. All statistical results and measured values are given in detail in the experimental studies section.

Keywords Medical voice recordings · Machine learning · Parkinson's disease · Performance comparison

1 Introduction

Neurological diseases in the world cause more and more human deaths for people. Parkinson's disease is a neurodegenerative disease of the central nervous system that causes loss of motor reflex and speech and affects behavior, mental process, and other vital functions [1]. Parkinson's disease was described and named as shaky paralysis by Doctor James Parkinson in 1817 [2]. It is generally seen in elderly people and causes loss of speech and motor abilities (balance, etc.) in 90% of patients [3]. Parkinson's disease is the second most common neurological health problem following Alzheimer's disease [4]. The incidence and prevalence of the disease vary in different studies. In general, approximately 10 million people in the world complain of this disease [5, 6]. In a recent comprehensive study, the incidence of the disease was reported to be 20/100000 [7]. It is known that there are more than a million Parkinson's patients in North America [8]. In Europe, the prevalence of the disease is 108–257/100,000, whereas the incidence rate is 11–19/100,000 [9]. It has been reported that 20% of patients have not yet been diagnosed [10]. Most studies have shown that the number of male patients is higher than that of female patients [11, 12]. In addition, males have a 4.4% chance of developing the disease and 3.7% of women [13]. There is no known cure for the disease so far [14, 15]. Parkinson's disease is usually diagnosed by invasive methods, although various drug therapies are applied to minimize the difficulties caused by the disease [16]. The association of speech disorders with Parkinson's disease has been proven in several studies [17–19]. In addition, many studies have shown that there is a decrease in voice performance as the disease progresses [20, 21]. Speech samples are ideal for a decision support system where data can be easily collected.

✉ Emre Avuçlu
emreavuclu@aksaray.edu.tr

Abdullah Elen
aelen@karabuk.edu.tr

1 Department of Computer Technology, Aksaray University, Aksaray, Turkey

2 Department of Computer Technology, Karabuk University, Karabuk, Turkey

Published online: 13 September 2020

Table 1 List of measurements applied to acoustic signals recorded from patients

Feature	Descriptions
MDVP: Fo (Hz)	Kay Pentax MDVP average vocal fundamental frequency
MDVP: Fhi (Hz)	Kay Pentax MDVP maximum vocal fundamental frequency
MDVP: Flo (Hz)	Kay Pentax MDVP minimum vocal fundamental frequency
MDVP: Jitter (%)	Kay Pentax MDVP jitter as a percentage
MDVP: Jitter (Abs)	Kay Pentax MDVP absolute jitter in microseconds
MDVP: RAP	Kay Pentax MDVP relative amplitude perturbation
MDVP: PPQ	Kay Pentax MDVP five-point period perturbation quotient
Jitter: DDP	Average absolute difference of differences between cycles, divided by the average period
MDVP: Shimmer	Kay Pentax MDVP local shimmer
MDVP: Shimmer (dB)	Kay Pentax MDVP local shimmer in decibels
Shimmer: APQ3	Three-point amplitude perturbation quotient
Shimmer: APQ5	Five-point amplitude perturbation quotient
MDVP: APQ	Kay Pentax MDVP eleven-point amplitude perturbation quotient
Shimmer: DDA	Average absolute difference between consecutive differences between the amplitudes of consecutive periods
NHR	Noise to harmonics ratio
HNR	Harmonics to noise ratio
RPDE	Recurrence period density entropy
D2	Correlation dimension
DFA	Detrended fluctuation analysis
Spread-1, Spread-2	Non-linear measures of fundamental frequency variation
PPE	Pitch period entropy
Status	Health status of the subject; 1—Parkinson's, 0—healthy

Therefore, speech samples and Parkinson's disease have been the subject of many studies [22–25]. Studies on Parkinson's disease have focused mostly on voice problems [26–29]. The reason for this is that speech samples are preferred in remote diagnosis and monitoring systems and their cost is low [30, 31]. Such systems also help in early diagnosis of the disease [32]. Sound disturbances can be measured simply by acoustic means that detect nonperiodic vibrations in sound [30]. Several recent studies have shown that deterioration of vocal cords resulting from the disease plays an important role in the diagnosis and follow-up of the disease. Little et al. aim to measure severity of Parkinson's disease by measuring the distortion of vocal cords [15]. Tsanas et al. compared the studies that predicted a person's tendency to progress or regress with the results of the Unified Parkinson's Disease Rating Scale, which doctors often use to detect the disease [33]. Revett et al. made some rules using the various frequencies derived from the sound samples of Little et al. [34]. They did studies to monitor Parkinson's disease and collect data [35, 36]. Only 5 to 10% of all Parkinson's patients have a disease onset between 20 and 40 years of age. The incidence of Parkinson's disease is usually the same [37, 38]. Although

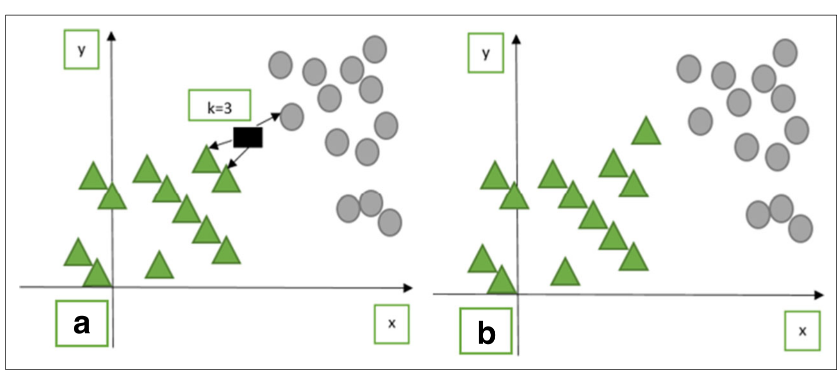


Fig. 1 Classification of a new sample according to its value ($k = 3$)

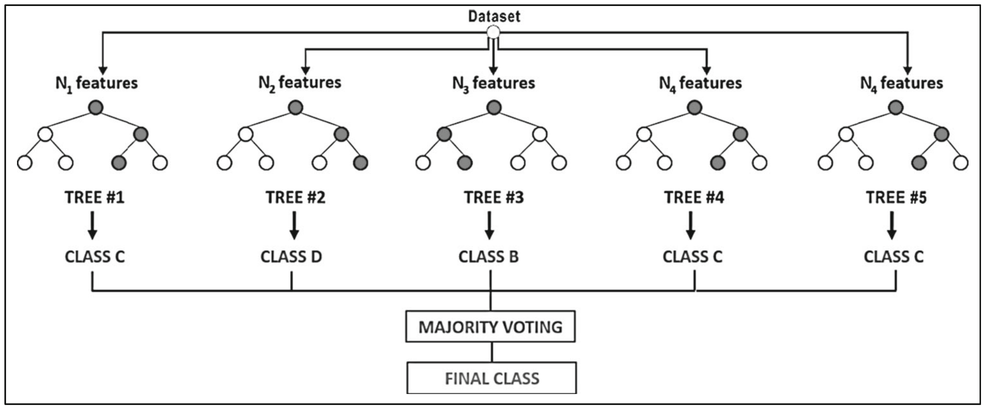


Fig. 2 Random Forest algorithm

studies on the diagnosis of Parkinson's were started in the 2000s, there are very few studies that will determine the patients' interest in Parkinson's disease by using voice and text measurements.

Güriüler has proposed a new hybrid diagnostic system for the diagnosis of Parkinson's disease. The proposed method lies in the proposed approach, which includes a combination of k-averages cluster-based feature weighting method and a complex-value artificial neural network [39]. Hirschauer et al. presented a computer model for the diagnosis of Parkinson's disease using the Enhanced Probabilistic Neural Network. In experimental studies, they compared their proposed method with the commonly used probabilistic neural network, SVM, k-NN algorithm, and DT algorithms [40]. Castro et al. have developed a multi-layer perceptron classifier as an artificial neural network to diagnose Parkinson's disease. Multiple networks were trained to change the number of neurons in the hidden layer from 10 to 6000 in 10 steps [41]. Devarajan and Ravi used the combinatorial Fuzzy K-Nearest Neighbor and case-based reasoning classifier for the diagnosis of Parkinson's disease [42]. Kadam and Jadhav have proposed a feature ensemble learning method based on sparse autoencoders for the diagnosis of Parkinson's disease [43]. A study was conducted on the use of sensitive devices to evaluate the effect of drugs in people with Parkinson's disease [44] They measured physical outcome with gait analysis of Parkinson's patients [45]. In a different study, a study was conducted and demonstrated the differences in the heart rate variability of patients with Parkinson's disease and healthy individuals [46]. Parkinson's disease was diagnosed using electrovestibulography [47]. A machine learning algorithm was applied to diagnose and treat Parkinson's disease in a different study [48]. Machine learning algorithms with biological and medical content were used in different studies [49, 50].

When the studies in the literature are analyzed, it is seen that they are mostly diagnosed with Parkinson's disease. In our study, in addition to these studies, it is to determine how the performance of these algorithms will change according to the input data rate.

This study has 3 main objectives related to machine learning algorithms (k-Nearest Neighborhood (k-NN), Random Forest, Naive Bayes, Support Vector Machine):

1. To determine which of these machine learning algorithms best diagnoses and classifies
2. To determine the rate at which the best training and test result will be obtained by making statistical measurements for Parkinson's disease diagnosis, and thus to contribute to the future studies in the literature
3. To develop a clinical decision support system that will help doctors diagnose Parkinson's disease

The development of a system to assist in the diagnosis of Parkinson's disease will be very beneficial for both the patient and the physician. It is also useful in terms of time to diagnose with such a system.

Table 2 Statistical measurements

Sensitivity or true positive rate	$TPR = \frac{TP}{TP+FN}$	Dice similarity coefficient	$DSC = \frac{2TP}{2TP+FP+FN}$
Specificity or true negative rate	$TNR = \frac{TN}{TN+FP}$	Accuracy	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$
Precision or positive predictive value	$PPV = \frac{TP}{TP+FP}$		
Negative predictive value	$NPV = \frac{TN}{TN+FN}$	F-measurements	$FM = \frac{2}{\frac{1}{TPR}+\frac{1}{PPV}}$
False positive ratio	$FPR = \frac{FP}{TN+FP}$	Matthews Correlation Coefficient	
False negative ratio	$FNR = \frac{FN}{TP+FN}$	$MCC = \frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)}}\ (TP+FN)\ (TN+FP)\ (TN+FN)$	

Table 3 Selected values for training and tests

Machine learning method		
k-NN	Distance method: Euclidean	$k = 5$
Random Forest	Number of trees:100	—
Naive Bayes	Distribution: Gaussian	—
Support Vector Machine	Kernel: Linear, Gaussian	Tolerance = 0.001

2 Materials and methods

In this section, technical specifications of the medical voice recording database used in the study are explained. In addition, general information about machine learning algorithms used in the study was given.

2.1 Dataset and features

This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease. A total of 195 sound samples were collected from each subject with an average of 6 recordings. Little et al. reported that purpose of the dataset was on methods of feature extraction for general voice and speech disorders [15]. Table 1 shows 22 kinds of features and their descriptions for acoustic sound recordings from patients. Sound recordings were made in sound processing cabinet using head-mounted microphone. The audio signals sampled at 16 bit resolution and 44.1 kHz were recorded directly to a computer using Computerized Speech Laboratory (CSL). Amplitude of the audio samples was digitally normalized to overcome discrepancies due to speech pressure. In Table 1, MDVP (Kay Pentax) means multidimensional voice program.

2.2 Machine learning algorithms

Machine Learning makes inferences from existing data using mathematical and statistical methods. These inferences can be defined as the method for estimating unknown values in any subject. The structure of the machine learning algorithms used in this study are explained in the following sections, respectively.

2.2.1 k-NN

In the k-NN algorithm, Euclidean, Manhattan, Minkowski, Chebyshev, Hellinger, and Angular were used as distance measurement methods. Since the formulation and definition of these methods are long, the general formulas and explanations of the k-NN algorithm are given below.

The k-NN algorithm was proposed in 1967 by Cover and Hart [51]. k-NN is one of the most basic pattern recognition and classification methods that classify objects based on the closest educational examples in the attribute space. Accordingly, the new vector is assigned to whichever class is the majority, referring to the classes to which the selected samples belong. There are different methods (Euclid, Manhattan, Minkowski, etc.) for calculating the distance of a new sample from the classified samples. The most common one is the Euclidean distance calculation method.

$$d(i,j) = \sqrt{\sum_{p=1}^{n} \left(X_{ip} - X_{jp}\right)^2} \tag{1}$$

where n represents the dimension; i is a new sample (X_{jp}) to be classified and the nearest k-neighbors $X_{ip}(i = 1, 2, \ldots, k)$. Figure 1 shows the process of classifying a new X_{jp} sample in a space of two dimensions ($n = 2$) according to $k = 3$.

Table 4 Statistical measurement results for 45 × 55

		TPR	SPC	PPV	NPV	FPR	FNR	ACC	MCC	FM
k-NN	Class0	0.67	0.72	0.27	0.93	0.28	0.07	0.71	0.28	0.38
	Class1	0.72	0.67	0.93	0.27	0.33	0.73	0.71	0.28	0.81
Random Forest	Class0	0.5	0.76	0.57	0.71	0.24	0.29	0.66	0.27	0.53
	Class1	0.76	0.5	0.71	0.57	0.5	0.43	0.66	0.27	0.74
Naive Bayes	Class0	0.67	0.79	0.54	0.86	0.21	0.14	0.75	0.43	0.6
	Class1	0.79	0.67	0.86	0.54	0.33	0.46	0.75	0.43	0.82
Support Vector Machine	Class0	1	0.67	0.03	1	0.33	0	0.67	0.13	0.05
	Class1	0.67	1	1	0.03	0	0.97	0.67	0.13	0.8

Table 5 Train and test results for 45×55

	k-NN (%)	Random Forest (%)	Naive Bayes (%)	Support Vector Machine (%)
Train result	100	100	88.24	87.06
Test result	70.91	66.36	75.45	67.27

In the example of Fig. 1, after determining the neighbor distances according to all the entries in the data set of a new sample, it is shown which class to belong to according to the k-neighborhood status (Fig. 1a). As a result, because the number of triangles is higher than $k = 3$, the class of the new sample is also determined as triangle.

2.2.2 Random Forest

This method combines the decisions of many trees, each trained in different sets of education rather than a single decision tree. As a result, the algorithm achieves a high success rate in solving classification problems. Figure 2 shows the general structure of the Random Forest algorithm.

2.2.3 Naive Bayes

Naive Bayes Distribution methods—Gaussian, Exponential, Gumbel, Uniform Continuous, Cauchy, Von Mises, and Log Normal (Dalton)—were used. Since the formulation and definition of these methods are long, the general formulas and explanations of the Naive Bayes algorithm are given below.

The Naive Bayes Classifier is a simple probabilistic classification method based on Bayes' theorem. In Bayes' theorem, where two independent events occur randomly (and) in succession, it is the probability that the second event occurs if one of these two events occur. By means of the change property, the product rule as in Eq. 2 can be written with two different expressions.

$$P(X \cap Y) = P(X|Y)P(Y) = P(Y|X)P(X) \tag{2}$$

Bayes' theorem defines the relationship between a random event occurring in a random process and conditional probabilities for another random event as in Eq. 3.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \tag{3}$$

The probabilities of the dependent states that are likely to occur in any problem are calculated by the Bayes in Eq. 3. In this equation, $P\ (X)$ represents the input probability of the problem, $P\ (X)$ represents the probability of a possible exit status, and $P\ (Y|X)$ represents the probability of a Y output versus input X [52]. In the NB classification technique, it analyzes the relationship between dependent and independent properties to create a conditional probability from each relationship. To classify a new sample, an estimate is made by combining the effects of independent variables on the dependent variable [53].

2.2.4 Support Vector Machine

Support Vector Machine is basically designed to solve duple classification problems. Thus, a hyper-plane is obtained to optimally separate the classes from each other. The classification is generally represented by class labels $\{-1, +1\}$. The data to be classified can be separated linearly ("AND" and "OR" problem) or cannot be separated by a single line (XOR problem). As is known, many classification problems in the real world consist of more than two classes. To solve such problems, there is a need for a multi-class SVM classifier. Multiple classification can be achieved by combining duple classifiers [54].

Table 6 Statistical measurement results for 50×50

		TPR	SPC	PPV	NPV	FPR	FNR	ACC	MCC	FM
k-NN	Class0	0.64	0.69	0.25	0.92	0.31	0.08	0.68	0.24	0.36
	Class1	0.69	0.64	0.92	0.25	0.36	0.75	0.68	0.24	0.79
Random Forest	Class0	0.51	0.74	0.56	0.7	0.26	0.3	0.65	0.25	0.53
	Class1	0.74	0.51	0.7	0.56	0.49	0.44	0.65	0.25	0.72
Naive Bayes	Class0	0.73	0.77	0.53	0.89	0.23	0.11	0.76	0.46	0.61
	Class1	0.77	0.73	0.89	0.53	0.27	0.47	0.76	0.46	0.83
Support Vector Machine	Class0	1	0.66	0.08	1	0.34	0	0.67	0.23	0.15
	Class1	0.66	1	1	0.08	0	0.92	0.67	0.23	0.8

Table 7 Train and test results for 50 × 50

	k-NN (%)	Random Forest (%)	Naive Bayes (%)	Support Vector Machine (%)
Train result	100	100	85.26	90.53
Test result	68	65	76	67

3 Experimental results

In this section, information about the performance of machine learning algorithms is presented using Confusion Matrix (CM). It is a matrix model that provides a holistic approach to the classification performance of an intelligent system algorithm. The CM structure is closely related to the classifier's performance and testing. Deriving a classification performance in CM in this way makes it possible to easily calculate all performance metrics related to experimental work. In CM, each element has a special meaning regarding classification performance. These elements are used as a reference in defining other performance metrics. The CM is structurally expressed as in Eq. 4.

$$CM = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \quad (4)$$

In this study, we applied ten statistical measurements to analyze results in experimental studies. These measurements are shown in Table 2. Where TP, TN, FP, and FN refer to true positive (correctly approved), true negative (correctly rejected), false positive (incorrectly approved), and false negative (incorrectly rejected), respectively.

To obtain results from 195 Parkinson's data in the data set, machine learning algorithms were first determined according to the values shown in Table 3.

After these determinations, numerical values of 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, and 95 were used as training data for each machine learning algorithm, respectively. As the test data, 55, 50, 45, 40, 35, 30, 25, 20, 15, 10, and 5 numerical values were used, respectively. With this data, statistical results of all machine learning algorithms, best and worst classification rates, and other data obtained were recorded in tables. The parameters of the machine learning algorithms (k-NN, RF, NB, SVM) used in experimental studies were adjusted as shown in Table 3. These parameter values were kept constant for each different combination of the dataset (training and test). The parameters in Table 3 were tested for each machine learning algorithm by trial-and-error and determined as the optimum value.

In Table 4, the values obtained from machine learning algorithms are listed separately for training and test results (Class0, healthy; Class1, Parkinson's disease).

The best training classification rate for 45% training and 55% test data was obtained from k-NN and Random Forest in Table 5. The best test result was obtained from Naive Bayes.

Table 8 Statistical measurement results for 55 × 45

		TPR	SPC	PPV	NPV	FPR	FNR	ACC	MCC	FM
k-NN	Class0	1	0.7	0.27	1	0.3	0	0.73	0.44	0.43
	Class1	0.7	1	1	0.27	0	0.73	0.73	0.44	0.83
Random Forest	Class0	0.92	0.73	0.36	0.98	0.27	0.02	0.76	0.47	0.52
	Class1	0.73	0.92	0.98	0.36	0.08	0.64	0.76	0.47	0.84
Naive Bayes	Class0	0.77	0.8	0.61	0.89	0.2	0.11	0.79	0.53	0.68
	Class1	0.8	0.77	0.89	0.61	0.23	0.39	0.79	0.53	0.84
Support Vector Machine	Class0	1	0.7	0.27	1	0.3	0	0.73	0.44	0.43
	Class1	0.7	1	1	0.27	0	0.73	0.73	0.44	0.83

Table 9 Train and test results for 55 × 45

	k-NN (%)	Random Forest (%)	Naive Bayes (%)	Support Vector Machine (%)
Train result	100	100	83.81	91.43
Test result	73.33	75.56	78.89	73.33

Table 10 Statistical measurement results for 60 × 40

		TPR	SPC	PPV	NPV	FPR	FNR	ACC	MCC	FM
k-NN	Class0	1	0.67	0.23	1	0.33	0	0.7	0.39	0.37
	Class1	0.67	1	1	0.23	0	0.77	0.7	0.39	0.8
Random Forest	Class0	0.91	0.7	0.32	0.98	0.3	0.02	0.72	0.43	0.48
	Class1	0.7	0.91	0.98	0.32	0.09	0.68	0.72	0.43	0.81
Naive Bayes	Class0	0.78	0.77	0.58	0.9	0.23	0.1	0.78	0.52	0.67
	Class1	0.77	0.78	0.9	0.58	0.22	0.42	0.78	0.52	0.83
Support Vector Machine	Class0	1	0.67	0.23	1	0.33	0	0.7	0.39	0.37
	Class1	0.67	1	1	0.23	0	0.77	0.7	0.39	0.8

Table 11 Train and test results for 60 × 40

	k-NN (%)	Random Forest (%)	Naive Bayes (%)	Support Vector Machine (%)
Train result	100	99.13	85.22	91.30
Test result	70	72.50	77.50	70

Table 12 Statistical measurement results for 65 × 35

		TPR	SPC	PPV	NPV	FPR	FNR	ACC	MCC	FM
k-NN	Class0	0.86	0.63	0.21	0.98	0.37	0.02	0.66	0.3	0.33
	Class1	0.63	0.86	0.98	0.21	0.14	0.79	0.66	0.3	0.77
Random Forest	Class0	1	0.68	0.34	1	0.32	0	0.73	0.49	0.51
	Class1	0.68	1	1	0.34	0	0.66	0.73	0.49	0.81
Naive Bayes	Class0	0.81	0.76	0.59	0.9	0.24	0.1	0.77	0.53	0.68
	Class1	0.76	0.81	0.9	0.59	0.19	0.41	0.77	0.53	0.82
Support Vector Machine	Class0	1	0.65	0.24	1	0.35	0	0.69	0.4	0.39
	Class1	0.65	1	1	0.24	0	0.76	0.69	0.4	0.79

Table 13 Train and test results for 65 × 35

	k-NN (%)	Random Forest (%)	Naive Bayes (%)	Support Vector Machine (%)
Train result	100	100	84.80	90.40
Test result	65.21	72.86	77.14	68.57

Table 14 Statistical measurement results for 70 × 30

		TPR	SPC	PPV	NPV	FPR	FNR	ACC	MCC	FM
k-NN	Class0	0.71	0.64	0.21	0.94	0.36	0.06	0.65	0.23	0.32
	Class1	0.64	0.71	0.94	0.21	0.29	0.79	0.65	0.23	0.76
Random Forest	Class0	0.83	0.65	0.21	0.97	0.35	0.03	0.67	0.29	0.33
	Class1	0.65	0.83	0.97	0.21	0.17	0.79	0.67	0.29	0.78
Naive Bayes	Class0	0.77	0.7	0.42	0.92	0.3	0.08	0.72	0.4	0.54
	Class1	0.7	0.77	0.92	0.42	0.23	0.58	0.72	0.4	0.8
Support Vector Machine	Class0	1	0.64	0.17	1	0.36	0	0.67	0.33	0.29
	Class1	0.64	1	1	0.17	0	0.83	0.67	0.33	0.78

Table 15 Train and test results for 70 × 30

	k-NN (%)	Random Forest (%)	Naive Bayes (%)	Support Vector Machine (%)
Train result	100	100	87.41	88.89
Test result	65	66.67	71.67	66.67

Table 16 Statistical measurement results for 75 × 25

		TPR	SPC	PPV	NPV	FPR	FNR	ACC	MCC	FM
k-NN	Class0	0.38	0.83	0.45	0.79	0.17	0.21	0.71	0.23	0.42
	Class1	0.83	0.38	0.79	0.45	0.62	0.55	0.71	0.23	0.81
Random Forest	Class0	0.75	0.88	0.55	0.95	0.12	0.05	0.86	0.56	0.63
	Class1	0.88	0.75	0.95	0.55	0.25	0.45	0.86	0.56	0.91
Naive Bayes	Class0	0.5	0.97	0.91	0.74	0.03	0.26	0.78	0.55	0.65
	Class1	0.97	0.5	0.74	0.91	0.5	0.09	0.78	0.55	0.84
Support Vector Machine	Class0	0.62	0.85	0.45	0.92	0.15	0.08	0.82	0.42	0.53
	Class1	0.85	0.62	0.92	0.45	0.38	0.55	0.82	0.42	0.89

Table 17 Train and test results for 75 × 25

	k-NN (%)	Random Forest (%)	Naive Bayes (%)	Support Vector Machine (%)
Train result	100	100	71.23	85.62
Test result	71.43	85.81	77.55	81.63

Table 18 Statistical measurement results for 80 × 20

		TPR	SPC	PPV	NPV	FPR	FNR	ACC	MCC	FM
k-NN	Class0	0.5	0.5	0.05	0.95	0.5	0.05	0.5	0	0.1
	Class1	0.5	0.5	0.95	0.05	0.5	0.95	0.5	0	0.65
Random Forest	Class0	1	0.51	0.05	1	0.49	0	0.53	0.16	0.1
	Class1	0.51	1	1	0.05	0	0.95	0.53	0.16	0.68
Naive Bayes	Class0	1	0.51	0.05	1	0.49	0	0.53	0.16	0.1
	Class1	0.51	1	1	0.05	0	0.95	0.53	0.16	0.68
Support Vector Machine	Class0	1	0.51	0.05	1	0.49	0	0.53	0.16	0.1
	Class1	0.51	1	1	0.05	0	0.95	0.53	0.16	0.68

Table 19 Train and test results for 80 × 20

	k-NN (%)	Random Forest (%)	Naive Bayes (%)	Support Vector Machine (%)
Train result	100	100	78.98	87.90
Test result	50	52.63	57.89	52.63

Table 20 Statistical measurement results for 85 × 15

		TPR	SPC	PPV	NPV	FPR	FNR	ACC	MCC	FM
k-NN	Class0	0	0.48	0	0.93	0.52	0.07	0.47	X	0
	Class1	0.48	0	0.93	0	1	1	0.47	X	0.64
Random Forest	Class0	X	0.5	0	1	0.5	0	0.5	X	X
	Class1	0.5	X	1	0	X	1	0.5	X	0.67
Naive Bayes	Class0	0.54	0.53	0.47	0.6	0.47	0.4	0.53	0.07	0.5
	Class1	0.53	0.54	0.6	0.47	0.46	0.53	0.53	0.07	0.56
Support Vector Machine	Class0	X	0.5	0	1	0.5	0	0.5	X	X
	Class1	0.5	X	1	0	X	1	0.5	X	0.67

Table 21 Train and test results for 85 × 15

	k-NN (%)	Random Forest (%)	Naive Bayes (%)	Support Vector Machine (%)
Train result	100	99.39	77.58	82.42
Test result	46.67	50	53.33	50

Table 22 Statistical measurement results for 90 × 10

		TPR	SPC	PPV	NPV	FPR	FNR	ACC	MCC	FM
k-NN	Class0	X	0.5	0	1	0.5	0	0.5	X	X
	Class1	0.5	X	1	0	X	1	0.5	X	0.67
Random Forest	Class0	X	0.5	0	1	0.5	0	0.5	X	X
	Class1	0.5	X	1	0	X	1	0.5	X	0.67
Naive Bayes	Class0	0.33	0.36	0.3	0.4	0.64	0.6	0.35	X	0.32
	Class1	0.36	0.33	0.4	0.3	0.67	0.7	0.35	X	0.38
Support Vector Machine	Class0	1	0.53	0.1	1	0.47	0	0.55	0.23	0.18
	Class1	0.53	1	1	0.1	0	0.9	0.55	0.23	0.69

Table 23 Train and test results for 90 × 10

	k-NN (%)	Random Forest (%)	Naive Bayes (%)	Support Vector Machine (%)
Train result	100	100	76	81.14
Test result	50	50	30	55

Table 24 Statistical measurement results for 95 × 5

		TPR	SPC	PPV	NPV	FPR	FNR	ACC	MCC	FM
k-NN	Class0	X	0.5	0	1	0.5	0	0.5	X	X
	Class1	0.5	X	1	0	X	1	0.5	X	0.67
Random Forest	Class0	X	0.5	0	1	0.5	0	0.5	X	X
	Class1	0.5	X	1	0	X	1	0.5	X	0.67
Naive Bayes	Class0	0	0	0	0	1	1	0	X	0
	Class1	0	0	0	0	1	1	0	X	0
Support Vector Machine	Class0	1	0.56	0.2	1	0.44	0	0.6	0.33	0.33
	Class1	0.56	1	1	0.2	0	0.8	0.6	0.33	0.71

Table 25 Train and test results for 95 × 5

	k-NN	Random Forest	Naive Bayes	Support Vector Machine
Train result	100	100	75.14	72.97
Test result	50	50	0	60

Table 26 Features of the machine learning algorithms used in experimental studies

Classifiers	Features	
k-NN	Distance method: Euclidean	Number of k = 5
RF	Number of trees = 100	–
NB	Distribution: normal	–
SVM	Kernel: polynomial	–

In Table 6, the values obtained from machine learning algorithms are listed separately for training and test results (Class0, healthy; Class1, Parkinson's disease).

The best training classification rate for 50% training and 50% test data was obtained from k-NN and Random Forest in Table 7. The best test result was obtained from Naive Bayes.

In Table 8, the values obtained from machine learning algorithms are listed separately for training and test results (Class0, healthy; Class1, Parkinson's disease).

Table 27 Classification results of the machine learning algorithms by 10-fold cross-validation

	Accuracy	Sensitivity	Specificity	Precision	Recall	F1measure	MCC
k-NN	8564E-01	7645E−01	7645E−01	8260E−01	7645E−01	7876E−01	5872E−01
NB	7026E−01	7676E−01	7676E−01	6987E−01	7676E−01	6807E−01	4612E−01
RF	8513E−01	8382E−01	8382E−01	7974E−01	8382E−01	8132E−01	6343E−01
SVM	8872E−01	8410E−01	8410E−01	8510E−01	8410E−01	8458E−01	6919E−01

The best training classification rate for 55% training and 45% test data was obtained from k-NN and Random Forest in Table 9. The best test result was obtained from Naive Bayes.

In Table 10, the values obtained from machine learning algorithms are listed separately for training and test results (Class0, healthy; Class1, Parkinson's disease).

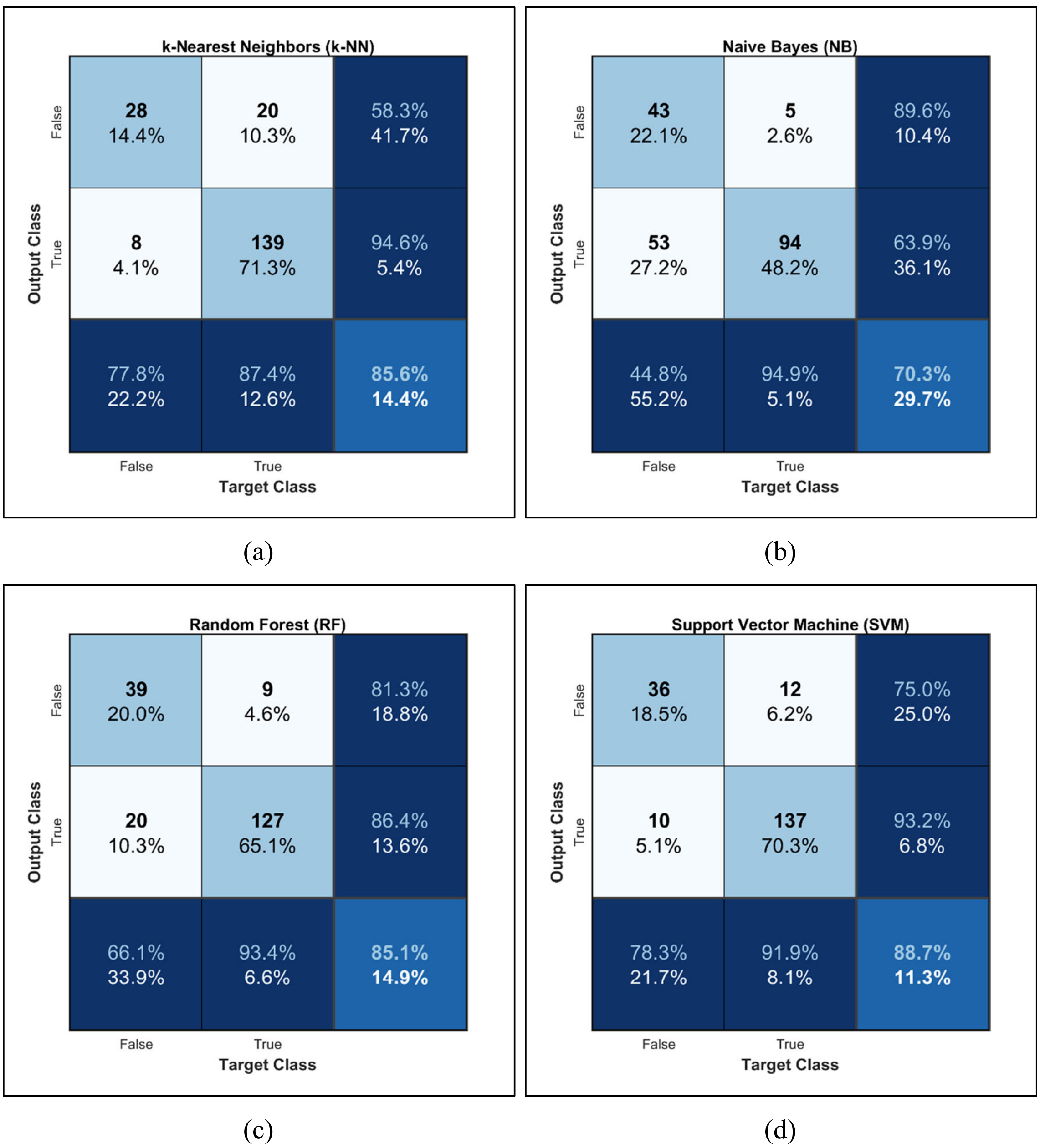


Fig. 3 Confusion matrices of the machine learning algorithms; **a** k-NN, **b** Naive Bayes, **c** Random Forest, **d** SVM

The best training classification rate for 60% training and 40% test data was obtained from k-NN in Table 11. The best test result was obtained from Naive Bayes.

In Table 12, the values obtained from machine learning algorithms are listed separately for training and test results (Class0, healthy; Class1, Parkinson's disease).

The best training classification rate for 65% training and 35% test data was obtained from k-NN and Random Forest in Table 13. The best test result was obtained from Naive Bayes.

In Table 14, the values obtained from machine learning algorithms are listed separately for training and test results (Class0, healthy; Class1, Parkinson's disease).

The best training classification rate for 70% training and 30% test data was obtained from k-NN and Random Forest in Table 15. The best test result was obtained from Naive Bayes.

In Table 16, the values obtained from machine learning algorithms are listed separately for training and test results (Class0, healthy; Class1, Parkinson's disease).

The best training classification rate for 75% training and 25% test data was obtained from k-NN and Random Forest in Table 17. The best test result was obtained from Random Forest.

In Table 18, the values obtained from machine learning algorithms are listed separately for training and test results (Class0, healthy; Class1, Parkinson's disease).

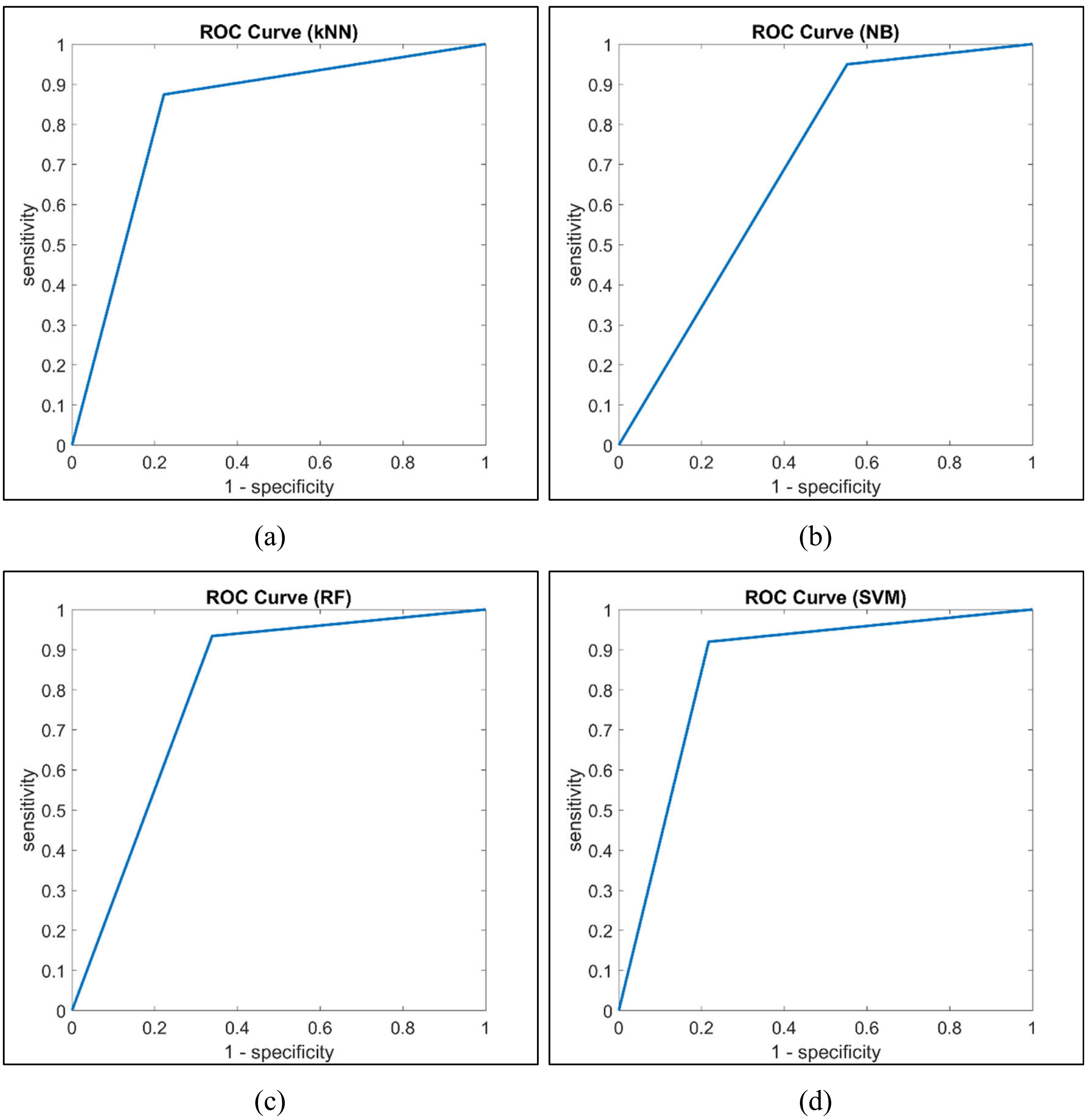


Fig. 4 ROC curves of the machine learning algorithms; **a** k-NN, **b** Naive Bayes, **c** Random Forest, **d** SVM

The best training classification rate for 80% training and 20% test data was obtained from k-NN and Random Forest in Table 19. The best test result was obtained from Naive Bayes.

In Table 20, the values obtained from machine learning algorithms are listed separately for training and test results (Class0, healthy; Class1, Parkinson's disease).

The best training classification rate for 85% training and 15% test data was obtained from k-NN in Table 21. The best test result was obtained from Naive Bayes.

In Table 22, the values obtained from machine learning algorithms are listed separately for training and test results (Class0, healthy; Class1, Parkinson's disease).

The best training classification rate for 90% training and 10% test data was obtained from k-NN and Random Forest in Table 23. The best test result was obtained from k-NN, Random Forest, and Support Vector Machine.

In Table 24, the values obtained from machine learning algorithms are listed separately for training and test results (Class0, healthy; Class1, Parkinson's disease).

The best training classification rate for 95% training and 5% test data was obtained from k-NN and Random Forest in Table 25. The best test result was obtained from the Support Vector Machine.

The meaning of X marks in Tables 20, 22, and 24 occurred due to insufficient test data. Low test data was not a good choice. Naive Bayes generally gave the best results in the test experiments. Naive Bayes' success decreased as the data training data decreased. The training successes of k-NN and Random Forest achieved better results than other algorithms. Overall, the best test result was obtained from the Random Forest machine learning algorithm with 85.81% and 100% training success. Apart from this test data, the Naive Bayes algorithm always gave the best results. It was observed that the algorithms decreased in the test success as the test data decreased.

In this study, 10-fold cross-validation procedures were also performed. Table 26 shows the parameters used for 10-fold cross-validation.

Classification results of machine learning methods with 10-fold cross-validation of the Parkinson data set are shown in Table 27.

When looking at Table 27, the best results in general were obtained from SVM. Figure 3 shows the confusion matrices of the machine learning algorithms.

Figure 4 shows ROC curves of the machine learning algorithms. These ROC curves were obtained from operations with 10-fold cross-validation.

As a result of the tests performed with 10-fold cross-validation, the accuracy value was obtained from the highest 88.72% SVM machine learning algorithm, and the lowest was obtained from 70.26% NB machine learning algorithm.

4 Conclusions

In this study, performance evaluation process was performed by making statistical measurements between 4 different machine learning algorithms. In this study, it was determined from which data rate the best education and test result would be obtained for the diagnosis of Parkinson's disease. The best machine learning algorithm for the diagnosis of Parkinson's disease was determined. A study was carried out using biomedical voice parameters to assist physicians as a decision support system for statistical analysis of these data. As a result, the best classification algorithm was determined by statistical measurements. The best training and test results were obtained from 75% training and 25% test data. In this study, a sample application was developed for which data rate the best training and test results would be obtained for any data set. As a result of this study, in any study to be made in the literature, it was determined which of the best classifiers could be among these algorithms. Moreover, it was determined which of these algorithms should be used for the diagnosis of Parkinson's disease.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

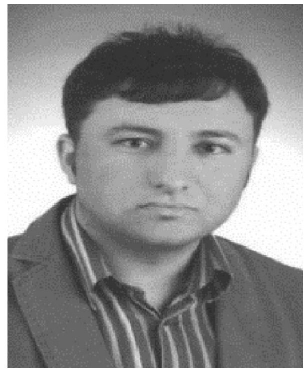
References

1. Jankovic J (2007) Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 79(4):368–376
2. Langston JW (2002) Parkinson's disease: current and future challenges. NeuroToxicology 23(4–5):443–450
3. O'Sullivan, S. B., & Schmitz, T. J. (2007). Physical rehabilitation (5th Edition b., Cilt Parkinson Disease). Philadelphia: F. A. Davis Company.
4. de Rijk MC, Launer LJ, Berger K, Breteler MM, Dartigues JF, Baldereschi M et al (2000) Prevalence of Parkinson's disease in Europe: a collaborative study of population-based cohorts. Neurology 54:21–23
5. Parkinson Derneği. (2011). March 29, 2012 tarihinde Parkinson Nedir? http://www.parkinsondernegi.org/Icerik.aspx?Page=parkinsonnedir&ID=5
6. de Lau LM, Breteler MM (2006) Epidemiology of Parkinson's disease. Lancet Neurol 5(6):525–535

7. Rajput M, Rajput A, Rajput AH (2007) Epidemiology. In: Pahwa R, Lyons (Dü) içinde KE (eds) Handbook of Parkinson's disease (4 b.). Informa Healthcare, USA
8. Lang AE, Lozano AM (1998) Parkinson's disease - first of two parts. N Engl J Med 339:1044–1053
9. von Campenhausen S, Bornschein B, Wick R, Bötzel K, Sampaio C, Poewe W, Oertel W, Siebert U, Berger K, Dodel R (2005) Prevalence and incidence of Parkinson's disease in Europe. Eur Neuropsychopharmacol 15(4):473–490
10. Schrag A, Ben-Schlomo Y, Quinn N (2002) How valid is the clinical diagnosis of Parkinson's disease in the community? J Neurol Neurosurg Pshych 73:529–535
11. Baldereschi M, Di Carlo A, Rocca WA, Vanni P, Maggi S, Perissinotto E et al (2000) Parkinson's disease and parkinsonism in a longitudinal study. Two fold higher Incid Neurol 55:1358–1363
12. Haaxma CA, Bloem BR, Borm GF, Oyen WJ, Leenders KL, Eshuis S et al (2007) Gender differences in Parkinson's disease. J Neurol Neurosurg Psychiatry 78:819–824
13. Elbaz A, Bower JH, Maraganore DM, McDonnell SK, Peterson BJ, Ahlskog JE, Schaid DJ, Rocca WA (2002) Risk tables for parkinsonism and Parkinson's disease. J Clin Epidemiol 55:25–31
14. Singh N, Pillay V, Choonara YE (2007) Advances in the treatment of Parkinson's disease. Prog Neurobiol 81(1):29–44
15. Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO (2009) Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng 56(4):1010–1022
16. National Collaborating Centre for Chronic Conditions (2006) Parkinson's disease. Royal College of Physicians, London
17. Darley FL, Aronson AE, Brown JR (1969) Differential diagnostic patterns of dysarthria. J Speech Hear Res 12:246–269
18. Gamboa J, Jimenez-Jimenez FJ, Nieto A, Montojo J, Orti-Pareja M, Molina JA et al (1997) Acoustic voice analysis in patients with Parkinson's disease treated with dopaminergic drugs. J Voice 11: 314–320
19. Ho A, Bradshaw JL, Iansek R (2008) For better or for worse: the effect of levodopa on speech in Parkinson's disease. Mov Disord 23(4):574–580
20. Harel B, Cannizzaro M, Snyder PJ (2004) Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. Brain Cogn 56:24–29
21. Skodda S, Rinsche H, Schlegel U (2009) Progression of dysprosody in Parkinson's disease over time – a longitudinal study. Mov Disord 24(5):716–722
22. Sakar CO, Kursun O (2010) Telediagnosis of Parkinson's disease using measurements of dysphonia. J Med Syst 34(4):591–599
23. Sapir S, Ramig L, Spielman J, Fox C (2010) Formant centralization ratio (FCR): a proposal for a new acoustic measure of dysarthric speech. J Speech Lang Hear Res 53:114–125
24. Cnockaert L, Schoentgen J, Auzou P, Ozsancak C, Defebve L, Grenez F (2008) Low frequency vocal modulations in vowels produced by Parkinsonian subjects. Speech Comm 50:288–300
25. (2013) Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Informat 17(4):828–834
26. Cunningham L, Mason S, Nugent C, Moore G, Finlay D, Craig D (2011) Home-based monitoring and assessment of Parkinson's disease. IEEE Transac Informat Technol Biomed 15(1):47–53
27. Rigas G, Tzallas A, Tsipouras M, Bougia P, Tripoliti E, Baga DF et al (2012) Assessment of tremor activity in the Parkinson's disease using a set of wearable sensors. IEEE Trans Inf Technol Biomed 16(3):478–487
28. Marino S, Ciurleo R, Lorenzo G, Barresi M, De Salvo S, Giacoppo S et al (2012) Magnetic resonance imaging markers for early diagnosis of Parkinson's disease. Neural Regen Res 7(8):611–619
29. Dastgheib Z, Lithgow B, Moussavi Z (2012) Diagnosis of Parkinson's disease using electrovestibulography. Med Biol Eng Comput 50(3):483–491
30. Little MA, McSharry PE, Roberts SJ, Costello DA, Moroz IM (2007) Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomed Eng Online 6(23):23
31. Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO (2012) Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Trans Biomed Eng 59(5): 1264–1271
32. Tsanas A, Little MA, McSharry PE, Ramige LO (2011) Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. J R Soc Interface 8:842–855
33. Tsanas, A., Little, M.A., Mcsharry, P.E. And Ramig, L.O. (2009). Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. Nature Precedings
34. Revett, K., Gorunescu, F. And Salem, A.B.M. (2009). Feature selection in Parkinson's disease: a rough sets approach. Proceedings of the International Multiconference on Computer Science and Information Technology, Poland, 425–428
35. Tsanas, A., Little, M.A., Mcsharry, P.E. And Ramıg, L.O. (2010). Enhanced classical dysphonia measures and sparse regression for telemonitoring of Parkinson's disease progression. IEEE international conference on acoustics speech and signal, 594–597
36. İsenkul ME (2011) Parkinson Hastalığı'nın Teşhisi İçin Veri Toplama ve Örüntü Tanıma Sistemi, Master Thesis. İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul
37. Fahn S, And Przedborski S (2000) Parkinsonizm. In: Rowland LP (ed) Merritt's textbook of neurology, vol 9789752771819, 10Th edn. Lippincott Williams & Wilkins, Philadelphia, pp 679–693
38. Quinn N, And Critchley P, And Marsden CD (1987) Young onset Parkinson's disease. Mov Disord 2(2):73–91
39. Gürüler H (2016) A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. Neural Comput & Applic 28(7): 1657–1666. https://doi.org/10.1007/s00521-015-2142-2
40. Hirschauer TJ, Adeli H, Buford JA (2015) Computer-aided diagnosis of Parkinson's disease using enhanced probabilistic neural network. J Med Syst 39(11):179. https://doi.org/10.1007/s10916-015-0353-9
41. Devarajan M, Ravi L (2018) Intelligent cyber-physical system for an efficient detection of Parkinson disease using fog computing. Multimed Tools Appl 78:32695–32719. https://doi.org/10.1007/s11042-018-6898-0
42. Devarajan M, Ravi L (2018) Intelligent cyber-physical system for an efficient detection of Parkinson disease using fog computing. Multimed Tools Appl 78:32695–32719. https://doi.org/10.1007/s11042-018-6898-0
43. Kadam VJ, Jadhav SM (2018) Feature ensemble learning based on sparse autoencoders for diagnosis of Parkinson's disease. In: Computing, Communication and Signal Processing, pp 567–581. https://doi.org/10.1007/978-981-13-1513-8_58
44. Pradhan SD, Scherer R, Matsuoka Y, Kelly VE (2011) Use of sensitive devices to assess the effect of medication on attentional demands of precision and power grips in individuals with Parkinson disease. Med Biol Eng Comput 49:1195–1199. https://doi.org/10.1007/s11517-011-0798-z

45. Tan D, Pua Y, Balakrishnan S et al (2019) Automated analysis of gait and modified timed up and go using the Microsoft Kinect in people with Parkinson's disease: associations with physical outcome measures. Med Biol Eng Comput 57:369–377. https://doi.org/10.1007/s11517-018-1868-2
46. Kallio M, Suominen K, Bianchi AM, Mäkikallio T, Haapaniemi T, Astafiev S, Sotaniemi KA, Myllylä VV, Tolonen U (2002) Comparison of heart rate variability analysis methods in patients with Parkinson's disease. Med Biol Eng Comput 40:408–414. https://doi.org/10.1007/BF02345073
47. Dastgheib ZA, Lithgow B, Moussavi Z (2012) Diagnosis of Parkinson's disease using electrovestibulography. Med Biol Eng Comput 50:483–491. https://doi.org/10.1007/s11517-012-0890-z
48. Myszczynska MA, Ojamies PN, Lacoste AMB, Neil D, Saffari A, Mead R, Hautbergue GM, Holbrook JD, Ferraiuolo L (2020) Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. Nat Rev Neurol 16:440–456. https://doi.org/10.1038/s41582-020-0377-8
49. Pio G, Ceci M, Prisciandaro F, Malerba D (2019) Exploiting causality in gene network reconstruction based on graph embedding. Mach Learn 109:1231–1279. https://doi.org/10.1007/s10994-019-05861-8
50. Barracchia EP, Pio G, D'Elia D, Ceci M (2020) Prediction of new associations between ncRNAs and diseases exploiting multi-type hierarchical clustering. BMC Bioinformat 21(1):70. https://doi.org/10.1186/s12859-020-3392-2
51. Cover T, Hart P (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13(1):21–27
52. Orhan U, Adem K (2012) The effects of probability factors in naive Bayes method. Elektrik-Elektronik ve Bilgisayar Mühendisliği Sempozyumu, Bursa, pp 722–724
53. Krishna, P. R. and De, S. K., "Naive-Bayes classification using fuzzy approach", Third International Conference on Intelligent Sensing and Information Processing, Bangalore/India, 61–64 (2005)
54. https://www.saedsayad.com/support_vector_machine.htm, access date: 10 Nov 2018

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Emre Avuçlu was born in Ankara, Turkey, in 1986. He has a master's degree in Information Technology at Selçuk University. He received a PhD degree in Computer Engineering from Selçuk University, Konya, in 2019. He is currently an Assist. Professor in Aksaray University Department of Computer Technology. He is interested in the field of image processing, artificial intelligence, and robotics.

Abdullah Elen was born in Bursa City, Turkey, in 1981. He received an A.Sc. degree in computer hardware from Uludag University, Turkey, in 2003, and a B.Sc. degree in electronics and computer education from Zonguldak Karaelmas (Bulent Ecevit) University, Turkey, in 2008. He received an M.Sc. degree in electronics and computer education from Karabuk University (KBU), Turkey, in 2011, and a PhD degree in computer engineering from KBU in 2018. From February 2009 to December 2018, he was a lecturer in KBU. Meanwhile, he worked at Department of Information Technology of KBU as a full stack developer. He is currently an assistant professor at the department of computer technologies in T.O.B.B. Vocational School of Technical Sciences of KBU. His current research interests include image processing, machine learning, and optimization algorithms.