Diagnosis of Parkinson's Disease with Acoustic Sounds by Rule Based Model

Kılıçarslan Yıldırım[1], Muhammed Kürşad Uçar[1(✉)], Ferda Bozkurt[2], and Mehmet Recep Bozkurt[1]

[1] Faculty of Engineering, Electrical-Electronics Engineering, Sakarya University, Sakarya, Turkey
{mucar,mbozkurt}@sakarya.edu.tr

[2] Vocational School of Sakarya, Computer Programming, Sakarya University of Applied Sciences, Sakarya, Turkey
fbozkurt@subu.edu.tr

Abstract. Parkinson's disease causes disruption in many vital functions such as speech, walking, sleeping, and movement, which are the basic functions of a human being. Early diagnosis is very important for the treatment of this disease. In order to diagnose Parkinson's disease, doctors need brain tomography, and some biochemical and physical tests. In addition, the majority of those suffering from this disease are over 60 years of age, make it difficult to carry out the tests necessary for the diagnosis of the disease. This difficult process of diagnosing Parkinson's disease triggers new researches. In our study, rule-based diagnosis of parkinson's disease with the help of acoustic sounds was aimed. For this purpose, 188 (107 Male-81 Female) individuals with Parkinson's disease and 64 healthy (23 Male-41 Female) individuals were asked to say the letter 'a' three times and their measurements were made and recorded. In this study, the data set of recorded 756 measurements was used. Baseline, Time, Vocal, MFCC and Wavelet that are extracted from the voice recording was used. The data set was balanced in terms of the "Patient/Healthy" feature. Then, with the help of Eta correlation coefficient based feature selection algorithm (E-Score), the best 20% feature was selected for each property group. For the machine learning step, the data were divided into two groups as 75% training, 25% test group with the help of systematic sampling method. The accuracy of model performance was evaluated with Sensivity, Specifitiy, F-Measurement, AUC and Kapa values. As a result of the study, it was found that the disease could be detected accurately with an accuracy rate of 84.66% and a sensitivity rate of 0.96. High success rates indicate that patients can be diagnosed with Parkinson's disease with the help of their voice recordings.

Keywords: Parkinson's disease · Acoustic sounds · Systematic feature selection · E-score · Decision tree

J. Hemanth et al. (Eds.): ICAIAME 2020, LNDECT 76, pp. 59–75, 2021.
https://doi.org/10.1007/978-3-030-79357-9_7

1 Introduction

Parkinson's disease is the most common neurodegenerative disorder of the central nervous system after Alzheimer's disease that causes the partial or complete loss of brain-induced motor reflex, speech, behavior, mental process, and other vital functions [1]. This disease was first described by Doctor James Parkinson in 1817 as shaky paralysis and named after him [2]. Today, Parkinson's disease has become a universal neurological disease affecting more than 10 million people worldwide [3]. The incidence of this disease is 1 in 1000, while this rate increases to 1% in the age range of 60–85, 5% in the age range of 85 and above [4]. Since the course of the disease is mild in the first years, patients may not experience significant motor disorders. However, as it is a progressive disease, it leads to important health problems in later times. Although there is no definitive treatment for Parkinson's disease, medication is often used to reduce symptoms that affect patients' daily lives [5].

Early diagnosis of the disease plays an important role in increasing the effectiveness of the treatment. However, there is no known diagnostic method for Parkinson's disease. For the diagnosis of the disease, physicians require some physical tests to assess the functional competence of the legs and arms, muscle status, free walking and balance. Physicians may also request other diagnostic procedures, such as blood tests and neuroimaging techniques to diagnose [6]. Unfortunately, in the early stages of the disease as they may show similar symptoms to other neurological disorders such as Multi system atrophy (MSA) and Progressive supranuclear palsy (PSP) the symptoms are misleading and may cause misdiagnosis. [7]. In addition, the age of the patients to be diagnosed is usually 60 years and older, making these procedures even more difficult. All these facts have led to the need for an easier and more reliable method of diagnosing the disease [8–10].

One of the most common symptoms in the early stages of Parkinson's disease is vocal (speech and voice) problems [11–14]. Recently, speeches (voices) of individuals have been recorded and studies have been carried out on the diagnosis of Parkinson's disease [2,8–12,14–22]. Generally, two different data sets were used in these studies. The first of these is 195 sound measurements taken from 23 individuals with Parkinson's disease and 8 healthy individuals and the second data set was a public data set consisting of multiple speech records from 20 patients with Parkinson's disease and 20 healthy individuals [2,14–20]. Other new studies have also attempted to diagnose Parkinson's disease using small data sets [9–12,21,22]. Although the data sets in these studies have uneven distribution in terms of the number of people with Parkinson's disease and healthy individuals, the researches were carried out without any upper or lower sampling [9–12,14,22].

Numerous scientific studies have been conducted in recent years in order to make a reliable and rapid diagnosis of Parkinson's disease. The most important objective of these studies is to eliminate physical difficulties for patients, reduce the workload of clinical staff and provide support for physicians to diagnose [9–12,14,22].

In 2018–2019, many new machine-based systems for the diagnosis of Parkinson's disease were developed [10,12,22,23]. The most recent approach among these is the model developed based on deep learning and predicting the severity of Parkinson's disease [22]. Developed Tensorflow-based system works with an accuracy rate of 62–81% using open UCI Machine Learning Repository data [22]. In a study conducted by Sadek et al. In 2019 based on artificial neural networks, they identified Parkinson's disease with a success rate of 93% [12]. 195 sound measurements from 31 individuals were used in the study. B. Karan et al. Established a model with support vector machines and random forest algorithm with 150 sound measurements taken from 45 people. 100% accuracy rate has been obtained from the model established with these small data sets [10]. However, when the training and test data are tested together, the machine moves towards memorization and makes biased decisions. When the data set is divided into training/test sets, the accuracy of the proposed model decreases effectively [10,12,14,16,19,20]. High accuracy rates from such small data sets cannot be obtained from large data sets. In order to obtain more appropriate results, multiple data groups should be used in the larger data set and balancing operation must be performed in the data set [24–26].

Features extracted to use in machine learning; vocal basic frequency, the amount of variability in frequency, the amount of variability in amplitude, the ratio of components between noise and sound tone, non-linear dynamic values and non-linear basic frequency values include similar features [14,16,17]. These features help us achieve very high accuracy rates (98–99%) in distinguishing between patients and healthy individuals in the diagnosis of Parkinson's disease [12,14,16,19,20].

In this study, we aimed to diagnose Parkinson's disease by designing rule based models with the help of acoustic sounds. For this purpose, 756 measurements obtained from 252 individuals and data set containing 752 features were used. This data set is more extensive than the data sets used in other studies in the literature. Therefore, this data set will contribute to the creation of a more successful model.

The organization of our article is as follows. In Sect. 2, the data used in the study, data preprocessing operations, feature selection algorithm, decision trees (DT) and performance evaluation criteria were explained. The results obtained in Sect. 3 and the interpretation of results obtained in Sect. 4 were given. In the Sect. 5, future studies are mentioned in order to develop this study.

2 Materials and Methods

The study was carried out as shown in the flow diagram in Fig. 1. Accordingly, the data set was first balanced by systematic sampling method. Then, the data sets created were grouped according to feature extraction methods. Then, 20% of the properties of the data sets were selected with the help of Eta correlation coefficient based feature selection algorithm (E-Score). A rule-based model was created with 75% of these data sets and the system we designed with 25% of the data sets was tested. Finally, the performance of the system was evaluated.

2.1 Data Set

The data set used in this study was obtained from Machine Learning Repository (UCI) of Istanbul University Cerrahpaşa Faculty of Medicine Department of Neurology. The data set consisted of 756 measurements of 188 patients (107 males and 81 females) and 64 healthy individuals (23 males and 41 females). They were asked to say the letter 'a' three times and their measurements were made and recorded [27]. One for each measurement being a label, a total of 753 properties have been created. The label group consists of 1 (patient) and 0 (healthy). This data set is shown in Table 1. ID number in Table 1; is the number to specify the persons to be measured given in sequence starting from zero. Gender is shown as 1 (Male) and 0 (Female). 3 records were taken from each individual.

Table 1. Data set

			Number of Data Attributes		
NM	**ID**	**Gender**	1 2 . . .	752	**Label**
1	0	1	. . . . .	.	1
2	0	1	. . . . .	.	1
3	0	1	. . . . .	.	1
.	.	.	. . . . .	.	.
.	.	.	. . . . .	.	.
.	.	.	. . . . .	.	.
756	251	0	. . . . .	.	0

NM Number of measurements, ID ID Number, Gender Female(0)-Male(1), Label Patient(1)-Healthy(0)

2.2 Data Preprocessing

In the literature, there are some steps created by Han and Kamber (2006) to prepare the data set for analysis [28]. Data preprocessing steps performed in this study are described below.

Dividing a Data Set into Feature Groups. The 752 properties in the raw dataset consist of some property groups. These include Baseline features, Intensity Parameters, Formant Frequencies, Bandwidth Parameters, Vocal Fold, MFCC, and Wavelet Features. Intensity Parameters, Formant Frequencies and Bandwidth Parameters are combined as Time frequency features because they are close properties. Thus, 5 basic property groups to be processed and a total of 6 data groups consisting of all properties were created and these data groups are shown in Table 2.

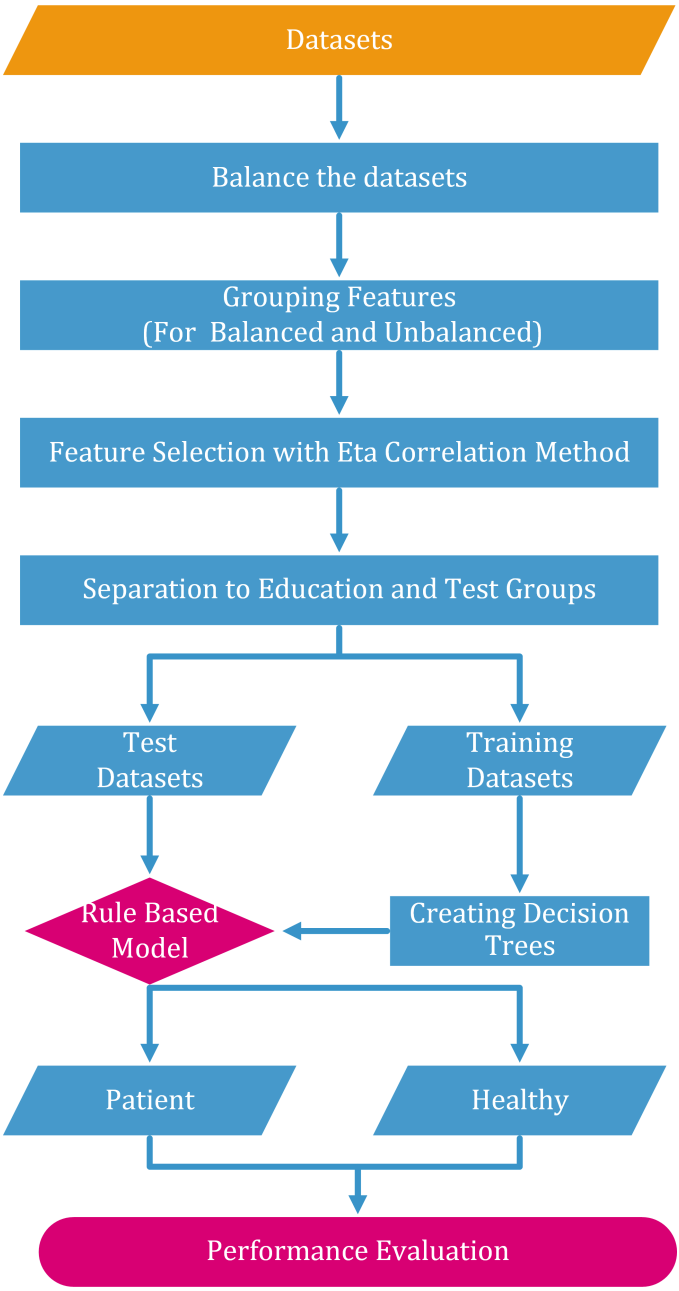

Fig. 1. Flow diagram

Table 2. Distribution of features

Data property groups	Number of features
Baseline	21
MFCC	84
Time	11
Vokal	22
Wavelet	614
All	752

Balancing the Data Set. If the number of label class values in the data set is not equal, the data set is unstable. If the data set used in the studies is unbalanced, the accuracy values may cause misleading decision making in performance evaluation [24]. To prevent this adverse situation, the data sets were balanced using systematic sampling method [25]. In the data set of this study, 756 measurements had 192 healthy (0) labels and 564 patient (1) labels. When we offset this unbalanced data set, we obtained 192 healthy and 192 patients. This balanced data set is shown in Table 3.

Table 3. Balanced data

	Raw data set	Balanced data set
Patient	564	192
Healthy	192	192
Total	756	384

2.3 Feature Selection Algorithm

Feature selection algorithms are of great importance in the field of machine learning. Significantly reducing very large data is the main function of the feature selection algorithm. As data increases, more advanced and better performance selection algorithms will be needed. A good feature selection algorithm will further improve the performance and speed of the designed system [29].

Eta Correlation Coefficient Based Feature Selection Algorithm. Correlation coefficients are the criteria that give information about the strength and direction of the relationship between the variables. The correlation coefficient formula to be used varies according to the type of variables compared. In the field of machine learning, the overall data is continuous numerical data. In this study, correlation coefficient was calculated between qualitative and continuous numerical variables [30]. In this study, Eta correlation coefficient (r_{pb}) which is

suitable for these variables was calculated as shown in Eq. 1. Using these calculated values, properties were selected from the data set.

$$r_{pb} = \frac{\bar{Y}_1 - \bar{Y}_0}{s_y}\sqrt{p_0 p_1} \tag{1}$$

$\bar{Y}_0$ and $\bar{Y}_1$ is the average of data labeled 0 and 1, respectively. s_y is the standard deviation of data in both classes and is calculated from the expression in Eq. 2.

$$s_y = \sqrt{\frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{n}} \tag{2}$$

p_0 ve p_1 is calculated by the expression in Eq. 3, wherein N, N_0 and N_1 represent the total number of labels, the number of elements labeled 0, and the number of elements labeled 1, respectively.

$$p_0 = \frac{N_0}{N}, p_1 = \frac{N_1}{N} \tag{3}$$

The number of properties selected by performing the above steps is shown in Table 4. Separate models were created for the whole data set shown in Table 4 and the best results were selected by calculating the performance values of the system for each data set.

Table 4. Number of selected properties by the Eta correlation coefficient

	Balanced		Unbalanced	
	All	Eta	All	Eta
Baseline	21	4	21	4
MFCC	84	17	84	17
Time	11	2	11	2
Vocal	22	4	22	4
Wavelet	614	123	614	123
Total	752	150	752	150

2.4 Classification

In our study, the classification process was performed with decision trees. Without classification, the data sets were divided into two groups, 75% training and 25% testing. The training and test datasets were shown in Table 5. This separation was carried out by systematic sampling method. Rule-based models were established for each training data set. The performance of the system was evaluated by running the designed models with test data sets.

Table 5. Training (75%) and test (25%) data set distribution

Label	Balanced				Unbalanced			
	All features		FSE		All features		FSE	
	Training	Test	Training	Test	Training	Test	Training	Test
Patient	144	48	144	48	423	141	423	141
Healthy	144	48	144	48	144	48	144	48

FSE Features Selected with Eta

Decision Tree. The basic structure of the decision tree algorithm includes roots, branches, nodes and leaves, and the basic structure of decision trees is shown in Fig. 2. Each attribute is associated with a node when creating the tree structure. There are branches between the root and nodes. Each node switches to the other node through branches. The decision in the tree is made according to the final leaf [31]. The basic logic in forming a decision tree structure can be summarized as; asking the related questions in each node reached and reaching the final leaf in the shortest way and time according to the answers given. Thus, models are created according to the answers obtained from the questions. The performance of this trained tree structure is calculated with test data and the model is used if it produces appropriate results.

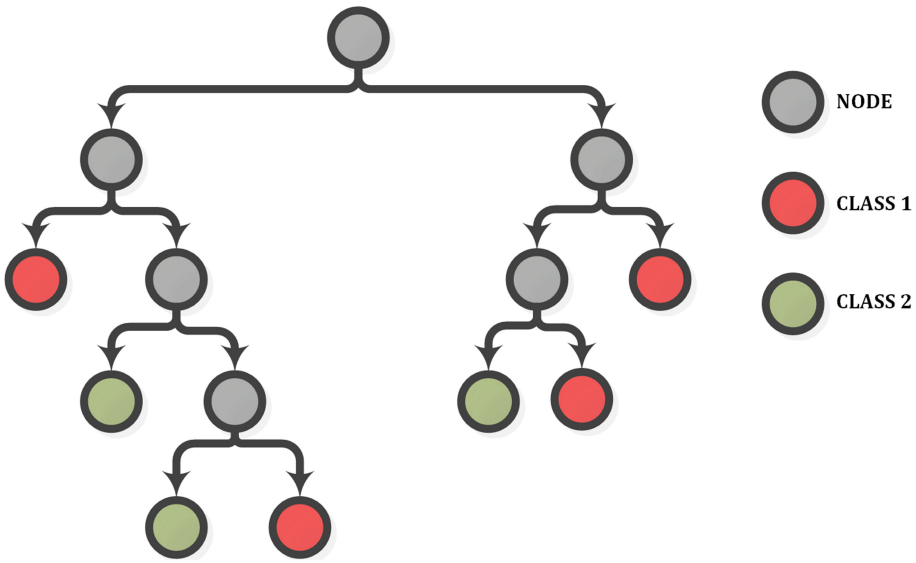

Fig. 2. Decision tree

2.5 Performance Evaluation Criteria

Different performance evaluation criteria were used to test the accuracy of the proposed systems. These are accuracy rates, sensitivity, specificity, F-Measurement, Area Under an ROC (AUC) and kappa coefficient value. These performance criteria are described in detail in the subheadings [32].

Accuracy. The accuracy is calculated as in Eq. 4. The expressions TP, TN, FP and FN in Eq. 4 are True Positives (TP), True Negatives (TN), False Positives (FP) and False Negative (FN) respectively. The accuracy rate of a system is intended to be 100. An accuracy of 100 means that the designed model answers all the questions correctly [32].

$$\text{Accuracy Rates} = \frac{TP + TN}{TP + TN + FN + FP} * 100 \tag{4}$$

Sensitivity. Sensitivity indicates the ability of the test to distinguish real patients within patients and is calculated as in Eq. 5. The sensitivity ranges from 0 to 1. A diagnostic test is required to have a sensitivity value of 1. A sensitivity value of 1 indicates that the test can correctly diagnose all patients [32].

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5}$$

Specificity. Specificity is the ability of the system to find what is actually healthy among the people who are actually healthy and is calculated as in Eq. 6. Specificity ranges from 0 to 1. Used in cases where the disease needs to be confirmed. A specificity value of 1 indicates that the test is able to correctly identify all healthy people [32].

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{6}$$

F-Measure. F-Measure is used to determine the effectiveness of the model being created. The value obtained is the weighted average of sensitivity and specificity values. The F-measure is calculated as in Eq. 7. It gets a value ranging from 0 to 1. 1 indicates that the created model is excellent, and 0 indicates that it is very poor [32].

$$F = 2 * \frac{\text{specificity} * \text{sensitivity}}{\text{specificity} + \text{sensitivity}} \tag{7}$$

AUC. The AUC value is used to evaluate the performance of diagnostic tests used to diagnose a disease. When analyzing with the ROC curve, the curves of the different tests are drawn on top of each other and then the comparison is made. In each ROC curve given in the results, the ideal ROC curve is also given. In this way, it can be seen how close the designed system is to the ideal. The fact that the ROC curve is close to the left or top axis is an indication that it is better for diagnosing the location [32].

Kappa. The Kappa coefficient is a coefficient which provides information about reliability by correcting the "chance matches" that occur entirely depending on chance. Different limit values have been defined in the literature regarding the degree of agreement of the Kappa coefficient [32].

3 Results

In this study, a rule-based model was developed by using acoustic sounds for the diagnosis of Parkinson's disease. First, the data were balanced and then divided into groups according to feature types. E-Score based feature was selected for each feature group and decision tree based diagnostic system was created.

Table 6 shows the change in performance according to the distribution of the test data set in the decision tree process. The Average column represents the average of the accuracy rates for each row. The mean highest accuracy was obtained when the test data set was 25%. Therefore, in this study, the data set was divided into two groups as Training (75%) and Test (25%). The results obtained were obtained at this rate unless otherwise stated.

As shown in Table 7, the best test accuracy rate of unbalanced data was 81.48% from all properties. Sensitivity value is also 0.86. Although these values are good, the Specificity value is 0.67. Sensitivity values are much better, although the Specificity values are low, as shown in Table 7. The reason for this is that the number of healthy data in the unbalanced data set is less than the number of patient data. Therefore, this model does not work stable. As can be seen in Table 9, the accuracy ratio, specificity and sensitivity values obtained with the model generated with all unbalanced selected property data are 84.66%, 0.70 and 0.90 respectively. As it can be understood from here, the performance of the model designed by selecting feature with Eta correlation method is higher.

As shown in Table 8, the best test accuracy ratio of the balanced data was obtained from the Wavelet property with a rate of 75%. Specificity and sensitivity values of this property are 0.77 and 0.73, respectively. Although the accuracy rate (75%) obtained with the model created with balanced data is lower than the accuracy rate (81.48%) with the model created with unbalanced data, it is more stable. The reason for this is that we perform the balancing process and equalize the number of patients and healthy data. As we can see in Table 8, the performance evaluation criteria we obtained when we perform the balancing process are closer to each other. Therefore, this model works more stable. As

Table 6. Change in test ratio versus overall performance change

Test %	U	US	B	BS	Average
	AR	AR	AR	AR	
50.00	68.41	67.96	64.69	62.26	65.83
45.00	69.72	66.90	63.67	64.63	66.23
40.00	67.36	67.80	62.31	61.76	64.81
35.00	68.25	67.62	63.98	62.13	65.50
30.00	68.48	67.39	67.18	64.71	66.94
25.00	70.78	70.94	67.09	65.51	68.58
20.00	63.90	61.63	63.89	58.84	62.07
15.00	68.04	70.12	68.77	70.05	69.24
10.00	62.66	59.49	69.46	70.91	65.63

BS Balanced Selection, US Unbalanced Selection,
U Unbalanced, B Balanced, AR Accuracy Rate

Table 7. Classification results of unbalanced data

		AR	Specificity	Sensitivity	F-measure	AUC	Kappa
Baseline	Training	82.89	0.53	0.93	0.68	0.73	0.51
	Test	79.37	0.54	0.87	0.67	0.71	0.43
MFCC	Training	83.60	0.64	0.90	0.75	0.77	0.56
	Test	80.42	0.61	0.87	0.72	0.74	0.47
Time	Training	80.78	0.50	0.91	0.65	0.71	0.45
	Test	78.84	0.48	0.89	0.62	0.68	0.39
Vocal	Training	82.72	0.58	0.91	0.71	0.75	0.52
	Test	72.49	0.41	0.83	0.55	0.62	0.24
Wavelet	Training	94.89	0.92	0.96	0.94	0.94	0.87
	Test	75.13	0.61	0.80	0.69	0.70	0.38
AF	Training	97.88	0.97	0.98	0.97	0.97	0.94
	Test	81.48	0.67	0.86	0.76	0.77	0.52

AF All Features, AR Accuracy Rate

shown in the Table 10, the highest accuracy rate in balanced selected properties was calculated as 81.25% in the model designed with the entire property data set.

Test performance is improved in models designed by applying feature selection process with Eta correlation method to all properties of both balanced and unbalanced data sets. While the test accuracy ratio of the model designed with "all balanced property data" shown in Table 8 is 66.67%, when we apply the feature selection process, this ratio increased to 81.25% as shown in Table 10. Similarly, while the test accuracy ratio of the model we designed with all

Table 8. Classification results of balanced data

		AR	Specificity	Sensitivity	F-measure	AUC	Kappa
Baseline	Training	77.08	0.78	0.76	0.77	0.77	0.54
	Test	70.83	0.73	0.69	0.71	0.71	0.42
MFCC	Training	73.96	0.68	0.80	0.73	0.74	0.48
	Test	66.67	0.63	0.71	0.66	0.67	0.33
Time	Training	83.68	0.87	0.81	0.84	0.84	0.67
	Test	71.88	0.83	0.60	0.70	0.72	0.44
Vocal	Training	70.49	0.79	0.62	0.69	0.70	0.41
	Test	69.79	0.79	0.60	0.69	0.70	0.40
Wavelet	Training	90.97	0.90	0.92	0.91	0.91	0.82
	Test	75.00	0.77	0.73	0.75	0.75	0.50
AF	Training	73.96	0.68	0.80	0.73	0.74	0.48
	Test	66.67	0.63	0.71	0.66	0.67	0.33

AF All Features, AR Accuracy Rate

Table 9. Classification results of unbalanced selected properties

		AR	Specificity	Sensitivity	F-measure	AUC	Kappa
Baseline	Training	78.13	0.36	0.93	0.52	0.64	0.34
	Test	79.37	0.39	0.92	0.55	0.66	0.36
MFCC	Training	85.19	0.58	0.95	0.72	0.76	0.57
	Test	80.95	0.50	0.91	0.65	0.70	0.44
Time	Training	79.54	0.32	0.96	0.47	0.64	0.34
	Test	79.37	0.28	0.96	0.44	0.62	0.30
Vocal	Training	77.78	0.36	0.92	0.51	0.64	0.33
	Test	71.96	0.28	0.86	0.43	0.57	0.16
Wavelet	Training	93.30	0.84	0.97	0.90	0.90	0.82
	Test	84.13	0.74	0.87	0.80	0.81	0.59
AF	Training	95.06	0.86	0.98	0.92	0.92	0.87
	Test	84.66	0.70	0.90	0.78	0.80	0.59

AF All Features, AR Accuracy Rate

unbalanced property data in Table 7 was 81.48%, this ratio increased to 84.66% when we applied feature selection process as shown in Table 9.

When we apply feature selection process to datasets with low number of Time, Baseline Vocal and MFCC features, the sensitivity value decreases while the specificity value of the system increases. In other words, while the patient detection ability of the system increases, its ability to determine healthy individuals decreases. This may be due to the fact that the E-Score features in the unbalanced data set have chosen more features that represent patient individuals.

Table 10. Classification results of balanced selected properties

		AR	Specificity	Sensitivity	F-measure	AUC	Kappa
Baseline	Training	79.17	0.82	0.76	0.79	0.79	0.58
	Test	71.88	0.73	0.71	0.72	0.72	0.44
MFCC	Training	79.51	0.88	0.72	0.79	0.80	0.59
	Test	64.58	0.81	0.48	0.60	0.65	0.29
Time	Training	68.75	0.64	0.74	0.68	0.69	0.38
	Test	62.50	0.63	0.63	0.63	0.63	0.25
Vocal	Training	65.97	0.56	0.76	0.64	0.66	0.32
	Test	67.71	0.60	0.75	0.67	0.68	0.35
Wavelet	Training	95.49	0.93	0.98	0.95	0.95	0.91
	Test	64.58	0.73	0.56	0.64	0.65	0.29
AF	Training	90.97	0.94	0.88	0.91	0.91	0.82
	Test	81.25	0.88	0.75	0.81	0.81	0.63

AF All Features, AR Accuracy Rate

Table 11. Best results

			AR	Specificity	Sensitivity	F-Ölçümü	AUC	Kappa
US	W	Eğitim	93.30	0.84	0.97	0.90	0.90	0.82
		Test	84.13	0.74	0.87	0.80	0.81	0.59
US	AF	Eğitim	95.06	0.86	0.98	0.92	0.92	0.87
		Test	84.66	0.70	0.90	0.78	0.80	0.59
BS	AF	Eğitim	90.97	0.94	0.88	0.91	0.91	0.82
		Test	81.25	0.88	0.75	0.81	0.81	0.63
U	AF	Eğitim	97.88	0.97	0.98	0.97	0.97	0.94
		Test	81.48	0.67	0.86	0.76	0.77	0.52
U	MFCC	Eğitim	83.60	0.64	0.90	0.75	0.77	0.56
		Test	80.42	0.61	0.87	0.72	0.74	0.47
B	W	Eğitim	90.97	0.90	0.92	0.91	0.91	0.82
		Test	75.00	0.77	0.73	0.75	0.75	0.50

BS Balanced Selection, US Unbalanced Selection, U Unbalanced, B Balanced, AF All Features, W Wavelet, AR Accuracy Rate

Table 11 shows the performance values of the six best performing models. The best model is a model designed with the unbalanced selection of Wavelet data set. The decision tree of this model is shown in Fig. 3 and reached the result in three steps. In addition, as shown in Table 11, all three of the top three best models are feature-selective models. It was concluded that the models designed with Eta correlation method gave better results. In addition, it was concluded that the models designed with All Feature Data Set and Wavelet Feature Data

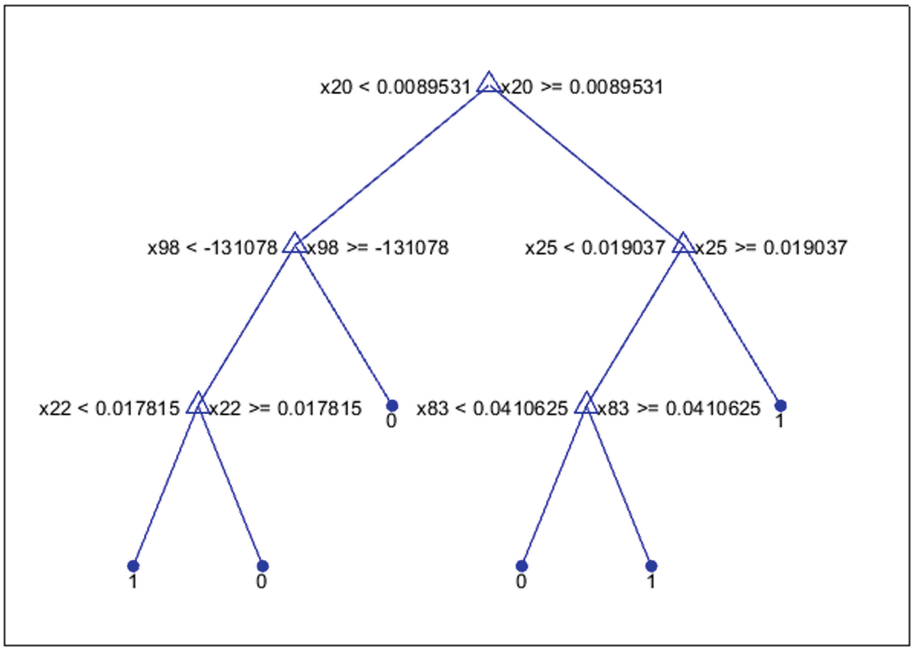

Fig. 3. Decision tree for best classification

Set from the data feature groups gave better results than the models designed with other data feature sets.

4 Discussion and Conclusion

There are many studies with high accuracy rate in the literature for the diagnosis of Parkinson's disease [2,9–12,14–22]. However, smaller data sets and less data characteristics were used in these studies. Therefore, they will not be able to obtain the high accuracy they obtain in large data sets. The data set and features we use are larger. For this reason, even though the accuracy rate of our designed system is lower than these studies, the system we designed is more stable. So the system we design gives more reliable results.

Many models designed in the literature have used unstable data sets [8–12,14,18–22]. In models created with unstable data, the system yields results close to the data in excess of quantity [25,26,28]. In this study, we used both balanced and unbalanced data sets. In addition, 20% of the characteristics of these datasets were selected by Eta correlation method. Thus, more specific data characteristics were used. In this study, we created models with data sets whose features were not selected. Thus, it was possible to compare the models created with Eta correlation method and normal models.

The results given in some studies in the literature are the average of training and test performances [16,33]. Therefore, the results of these studies are controversial. In this study, we performed the performance evaluation based on the test results. We also obtained the arithmetic average of all performance evaluation criteria while performing the performance evaluation. For this purpose, since the performance values of sensitivity, specificity, F-measurement, AUC and Kappa values are calculated over 1, we multiplied them by 100 and evaluated the performance on a scale of 100. Thus, the same performance evaluation size (100) as the Accuracy Ratio was obtained. After that, we obtained the arithmatic average of all performance evaluation criteria. The performance values obtained by these processes are more stable and more reliable than the other studies.

As a result, the model designed with this study requires only the voice recordings of the person to be diagnosed. In this way, the diagnostic process will be faster and at a lower cost. In addition, the workload of doctors will be reduced and patients will be provided with an easier diagnosis process.

5 Future Work

Our work in this area is ongoing and this work can be improved in many ways. Advances in signal processing applications occur with the development of signal processing techniques. The innovations for these steps can be listed as follows.

- The number of features can be increased with different processes.
- Different feature selection algorithms can be used to reveal useful features.
- Different machine learning algorithms can be used to increase the accuracy of decision making.
- The balancing process can be performed during and after the feature selection stage so that the location of the balancing process can be determined.
- Data groups for the diagnosis of gender-based Parkinson's disease can be classified as gender-based.

References

1. Kurt, I., Ulukaya, S., Erdem, O.: Musical feature based classification of Parkinson's disease using dysphonic speech. In: 2018 41st International Conference on Telecommunications and Signal Processing, TSP 2018. Institute of Electrical and Electronics Engineers Inc., August 2018
2. Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., Ramig, L.O.: Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans. Biomed. Eng. **56**(4), 1015–1022 (2009)
3. Tysnes, O.B., Storstein, A.: Epidemiology of Parkinson's disease, August 2017
4. Wood-Kaczmar, A., Gandhi, S., Wood, N.W.: Understanding the molecular causes of Parkinson's disease, November 2006
5. Jankovic, J.: Parkinson's disease: clinical features and diagnosis
6. Poewe, W., Scherfler, C.: Role of dopamine transporter imaging in investigation of parkinsonian syndromes in routine clinical practice. Mov. Disorders **18**(SUPPL. 7) (2003)

7. Perju-Dumbrava, L.D., et al.: Dopamine transporter imaging in autopsy-confirmed Parkinson's disease and multiple system atrophy. Mov. Disorders Offic. J. Mov. Disorder Soc. **27**(1), 65–71 (2012)
8. Okan Sakar, C., et al.: 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. **74**, 255–263 (2019)
9. Mathur, R., Pathak, V., Bandil, D.: Parkinson disease prediction using machine learning algorithm. In: Rathore, V.S., Worring, M., Mishra, D.K., Joshi, A., Maheshwari, S. (eds.) Emerging Trends in Expert Applications and Security. AISC, vol. 841, pp. 357–363. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-2285-3_42
10. Karan, B., Sahu, S.S., Mahto, K.: Parkinson disease prediction using intrinsic mode function based features from speech signal. Biocybern. Biomed. Eng. (2019)
11. Tang, J., et al.: Artificial neural network-based prediction of outcome in Parkinson's disease patients using DaTscan SPECT imaging features. Mol. Imaging Biol. **21**, 1165–1173 (2019)
12. Sadek, R.M., et al.: Parkinson's Disease Prediction Using Artificial Neural Network (2019)
13. Torun, S.: Parkinsonlularda Konuşma Fonksiyonunun Subjektif ve Objektif (Elektrolaringografik) Yöntemlerle incelenmesi (1991)
14. Gürüler, H.: A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. Neural Comput. Appl. **28**(7), 1657–1666 (2017)
15. Eskidere, Ö.: A comparison of feature selection methods for diagnosis of Parkinson's disease from vocal measurements. Technical report
16. Okan Sakar, C., Kursun, O.: Telediagnosis of Parkinson's disease using measurements of dysphonia. J. Med. Syst. **34**(4), 591–599 (2010)
17. Tsanas, A., Little, M.A., McSharry, P.E., Ramig, L.O.: 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**(59), 842–855 (2011)
18. Peker, M.: A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM. J. Med. Syst. **40**(5), 1–16 (2016). https://doi.org/10.1007/s10916-016-0477-6
19. Sakar, B.E., Serbes, G., Okan Sakar, C.: Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease. PloS one **12**(8), e0182428 (2017)
20. Peker, M., Şen, B., Delen, D.: Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm. J. Healthcare Eng. **6**(3), 281–302 (2015)
21. Pak, K., et al.: Prediction of future weight change with dopamine transporter in patients with Parkinson's disease. J. Neural Trans. **126**(6), 723–729 (2019)
22. Grover, S., Bhartia, S., Akshama, A.Y., Seeja, K.R.: Predicting severity of Parkinson's disease using deep learning. Procedia Comput. Sci. **132**, 1788–1794 (2018)
23. Wroge, T.J., Özkanca, Y., Demiroglu, C., Si, D., Atkins, D.C., Ghomi, R.H.: Parkinson's disease diagnosis using machine learning and voice. In: 2018 IEEE Signal Processing in Medicine and Biology Symposium, SPMB 2018 - Proceedings. Institute of Electrical and Electronics Engineers Inc., January 2019
24. Özen, Z., Kartal, E.: Dengesiz Veri Setlerinde Sınıflandırma (2017)
25. Alpar, R.: Reha Alpar. Spor Sağlık Ve Eğitim Bilimlerinden Örneklerle UYGULAMALI İSTATİSTİK VE GEÇERLİK GÜVENİRLİK. Detay Yayincilik, 2018 edition (2018)

26. García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J.M., Herrera, F.: Big data preprocessing: methods and prospects. Big Data Anal. **1**(1) (2016)
27. UCI. UCI Makine Öğrenimi Havuzu: Parkinson Hastalığı Sınıflandırma Veri Seti (2018)
28. Kartal, E.: Sınıflandırmaya Dayalı Makine Öğrenmesi Teknikleri ve Kardiyolojik Risk Değerlendirmesine İlişkin bir Uygulama. Ph.D. thesis (2015)
29. Kürşat Uçar, M.: Makine Öğrenimi için Eta Korelasyon Katsayısı Tabanlı Özellik Seçimi Algoritması: E-Skor Özellik Seçimi Algoritması - Semantik Bilgin (2019)
30. Alpar, R.: Applied statistic and validation-reliability (2010)
31. (John Ross) Quinlan, J.R., Ross, J.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Francisco (1993)
32. Kürşat Uçar, M.: OBSTRÜKTİF UYKU APNE TEŞHİSİ İÇİN MAKİNE ÖGRENMESİ TABANLI YENİ BİR YÖNTEM GELİŞTİRİLMESİ. Ph.D. thesis (2017)
33. Betul Erdogdu Sakar, M., et al.: Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J. Biomed. Health Inf. **17**(4), 828–834 (2013)