A Comparison of Classification Methods for Diagnosis of Parkinson's

Abdullah Elen*[1], Emre Avuclu[2]

Abstract: Parkinson's is a neurological health problem and one of the most common diseases affecting more than four million people worldwide. Recent studies have shown that deterioration of vocal cords, especially from Parkinson's, provides important information in the diagnosis and follow-up of the disease. In this study, a database of biomedical voice recordings from 32 people of different ages and genders was used to diagnose Parkinson's disease. With this database, the performance comparison of the machine learning algorithms k-Nearest Neighborhood (k-NN) and Naïve Bayes (NB) classifiers were performed. Seven different distance measurement methods (Chebychev, Correlation, Cosine, Euclidean, Hamming, Mahalanobis, and Spearman) for the k-NN and five different distribution methods (Uniform kernel, Epanechnikov kernel, Gaussian kernel, Triangular kernel and Normal distribution) for the NB classifier were performed in the performance process and separate tests were performed. The data obtained from these tests were compared with statistical measurements. In experimental studies, we used 10-fold cross validation technique for Parkinson dataset. Better results were obtained from k-NN classification algorithm than Naive Bayes classification algorithm. While k-NN mean accuracy score was 82.34%, this ratio was obtained as 74.15% for NB. Mahalanobis distance measurement method was found to give better results.

Keywords: Parkinson's disease, Machine learning, k-NN, Naive Bayes, Classification, Performance comparison.

1. Introduction

Parkinson's disease is a disorder of the central nervous system, which usually causes damage to the patient's motor skills, speech, and other body functions. It was first described by British Doctor James Parkinson in 1817 and was described as shaking or shaky paralysis [1]. The disease occurs in one out of every 100 people aged 65 and over. Parkinson's is an insidious and progressive disease [2]. The most important symptom of the disease, whose incidence increases with age, is impairments in voice and motor functions. There are three main symptoms of Parkinson's disease, tremor, rigidity, and bradykinesia. In addition to these basic symptoms, it affects the person with gait disturbance, dysphagia, signs of autonomic dysfunction, burning in the eyes, visual disturbances, pain and sensory complaints and depression [3-7]. The results of the studies were obtained by using speech disorder, low voice and dull speech which are the most important symptoms of the disease. Among these studies, Little et al. aim to measure the severity of the disease by measuring the dysphonia associated with Parkinson's disease [8]. They used 31 subjects, 23 of whom had Parkinson's disease. By applying speech tests to these people through the microphone, the vibration in the voice and tone determined which one could be used in this disease. Using these features, they aimed to determine the stage of the person's disease and achieved a successful outcome of 94%. Tsanas et al. conducted studies aimed at predicting the rate of change of a person's disease progression or regression [9]. Revett et al. set out rules from the sound samples of Little et al [10]. In another study, Tsanas et al.

converted the audio signals logarithmically [11]. They matched the results of the patient's home management with the United Parkinson's disease rating scale. 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 [12, 13]. The incidence of this disease is higher in males and females by 4/3 [2]. The main symptoms of the disease are slowness of movement, tremor, stiffness of muscles, posture, and balance disorders. For clinical diagnosis, it is sufficient to add one of the other three main symptoms to slow motion [14-16]. In the literature of different algorithms, there are different studies on performance measurement [17-19]. Although studies on the diagnosis of Parkinson's were started in the 2000s, there are very few studies that will detect using voice and text measurements.

In this study, performance evaluation process was performed between k-NN and NB classifier. In the experiments, Parkinson's dataset was used for training classifiers and test processes. The methods of each classifier are evaluated separately, and the results are recorded in tables and graphs. The aim of this study is to find out which of the k-NN and NB machine learning algorithms will give better results. Moreover, the learning methods used in these algorithms (for k-NN: Chebychev, Correlation, Cosine, Euclidean, Hamming, Mahalanobis, Spearman distance, and for Naive Bayes: Uniform kernel, Epanechnikov kernel, Gaussian kernel, Triangular kernel, and Normal distribution) to determine which ones will make better classification process as a result of statistical measurements.

Technical details about biomedical voice recordings and machine learning methods used in this study are given in the second section. Experimental results obtained in the section 3. In the last section, this paper is summarized.

[1] *Software Eng., Bandirma Onyedi Eylul University, Bandirma, Balıkesir - 10200, Turkey,*
  ORCID ID: 0000-0003-1644-0476

[2] *Computer Tech., Aksaray University, Aksaray - 68000, Turkey*
  ORCID ID: 0000-0002-1622-9059

* *Corresponding Author Email: aelen@bandirma.edu.tr*

2. Materials and Methods

This section describes the features of the dataset used to diagnose Parkinson's disease and general information about machine learning algorithms.

2.1. Dataset

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 person 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 [8]. The database was obtained from UCI with reference to the work of Little et al. Fig. 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 Figure 1, MDVP (Kay Pentax) means multi-dimensional voice program.

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.

**Fig. 1.** List of measurements applied to acoustic signals recorded from patients.

2.2. Machine Learning Methods

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 study structure of the two different machine learning algorithms used in this study is explained in the following titles, respectively.

The k-NN algorithm was proposed in 1967 by Cover and Hart [20]. 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.

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

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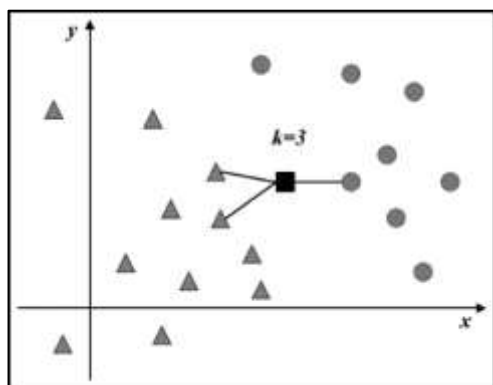
(a)
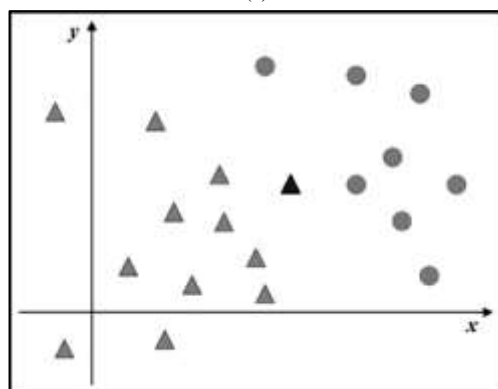
(b)

Fig. 2. Classification of a new sample according to its value ($k = 3$)

In the example of Fig. 2; 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. 2a). As a result, because the number of triangles is higher than $k = 3$, the class of the new sample is also determined as triangle (Fig. 2b).

The Naïve 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.

$$P(X \cap Y) = P(X|Y)P(Y)$$
$$P(X \cap Y) = P(Y|X)P(X)$$
(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.

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

The probabilities of the dependent states that are likely to occur in any problem are calculated by the Bayes Eq. 3 given above. In this equation, $P(X)$ represents the input probability of the problem, $P(Y)$ represents the probability of a possible exit status, and $P(Y|X)$ represents the probability of a Y output versus input X [21]. 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 [22].

The configurations of the machine learning algorithms used in experimental studies are shown in Table 1. The values of these features were determined as constant during the study process.

Table 1. Features of the ML methods.

Methods	Features
k-NN	Euclidean distance, **number of k**:5
Naive Bayes	**Distribution**: Gaussian

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. A CM is structurally expressed as in Eq. 4.

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

In this study, we applied five 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.

Table 2. Statistical measurements.

Accuracy	$ACC = \dfrac{TP+TN}{TP+TN+FP+FN}$
Precision	$PRC = \dfrac{TP}{TP+FP}$
Recall	$Recall = \dfrac{TP}{TP+FN}$
F1-score	$FM = \dfrac{2}{\dfrac{1}{TPR}+\dfrac{1}{PPV}}$
Matthews Correlation Coefficient	
$MCC = \dfrac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$	

Accuracy stands for the degree of proximity between the measurement result and the actual value. Precision is a measure of the reproducibility of the system. That is the closeness of the results obtained in the same way in an analysis. Recall is the ratio of the number of correctly classified positive samples to the total number of positive samples. That is, it shows how well the positive situations are predicted. F1-score is the harmonic mean of precision and recall. The precision and recall criteria alone are not enough to make a meaningful comparison. Evaluating both criteria

together gives more accurate results. Therefore, F1-score is defined. MCC is used as a measure of quality of the binary classification. It is essentially a correlation coefficient between observed and predicted binary classifications; Returns a value between -1 and +1. The +1 coefficient represents an excellent prediction, 0 is no better than a random estimate, and -1 indicates a mismatch between prediction and observation.

Table 3 shows the classification results obtained by using the 10-fold cross-validation technique of the Parkinson dataset with the k-NN algorithm. When the Parkinson dataset is classified using the 10-fold cross validation technique with the k-NN algorithm, the highest accuracy rate was obtained with the Mahalanobis distance metric. The lowest classification success belongs to the Spearman distance. Table 4 shows the classification results obtained by using the 10-fold cross-validation technique of the Parkinson dataset with the NB method. When the Parkinson dataset is classified into different distribution attributes in NB algorithm, it is seen that

Uniform, Epanechnikov, Gaussian and Triangular distributions used as kernel smoothing density estimation are more successful. When the average scores of both (NB and k-NN) classifiers are compared, it is seen that the k-NN algorithm has 82.34% accuracy and the NB algorithm has 74.15% accuracy. However, despite this, both algorithms have an MCC ratio of 49%.
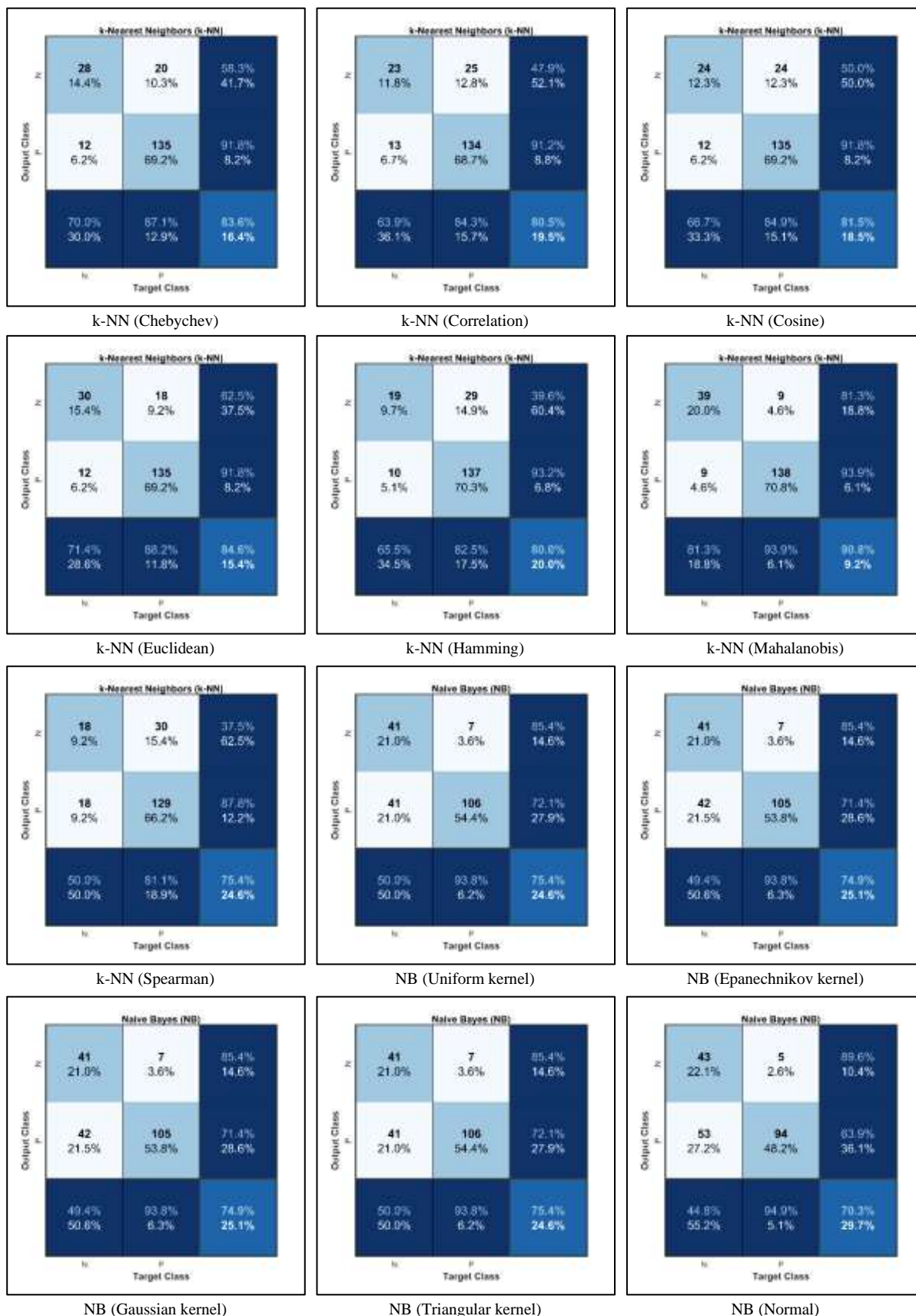
The classification results of the k-NN and NB algorithms with confusion matrix are shown in Fig. 3. Among the k-NN distance methods, Mahalanobis gave the best results. Triangular kernel and Uniform kernel gave the best result among NB distributions. ROC curves of the k-NN and NB classifiers are shown in Fig. 4. Looking at this table, the closest drawing to 1 was realized in k-NN because Mahalanobis distance gave the most accurate results compared to others. Likewise, in NB, Triangular kernel and Uniform kernel distributions are closest to 1.

Table 3. Classification results of the kNN by 10-fold cross validation of the Parkinson dataset.

Distances	Accuracy	Precision	Recall	F1-score	MCC
Chebychev	83.59%	78.55%	75.09%	76.52%	53.52%
Correlation	80.51%	74.08%	69.54%	71.17%	43.38%
Cosine	81.54%	75.79%	70.92%	72.69%	46.45%
Euclidean	84.62%	79.83%	77.17%	78.33%	56.94%
Hamming	80.00%	74.02%	66.39%	68.45%	39.69%
Mahalanobis	90.77%	87.56%	87.56%	87.56%	75.13%
Spearman	75.38%	65.57%	62.63%	63.59%	28.04%
Mean scores	*82.34%*	*76.49%*	*72.76%*	*74.04%*	*49.02%*

Table 4. Classification results of the NB by 10-fold cross validation of the Parkinson dataset.

Distributions	Kernel	Accuracy	Precision	Recall	F1-score	MCC
Box (Uniform)	Yes	75.38%	71.90%	78.76%	72.31%	50.20%
Epanechnikov	Yes	74.87%	71.57%	78.42%	71.84%	49.53%
Gaussian	Yes	74.87%	71.57%	78.42%	71.84%	49.53%
Triangular	Yes	75.38%	71.90%	78.76%	72.31%	50.20%
Normal	No	70.26%	69.87%	76.76%	68.07%	46.12%
Mean scores	*—*	*74.15%*	*71.36%*	*78.23%*	*71.27%*	*49.11%*

k-NN (Chebychev)

k-NN (Correlation)

k-NN (Cosine)

k-NN (Euclidean)

k-NN (Hamming)

k-NN (Mahalanobis)

k-NN (Spearman)

NB (Uniform kernel)

NB (Epanechnikov kernel)

NB (Gaussian kernel)

NB (Triangular kernel)

NB (Normal)

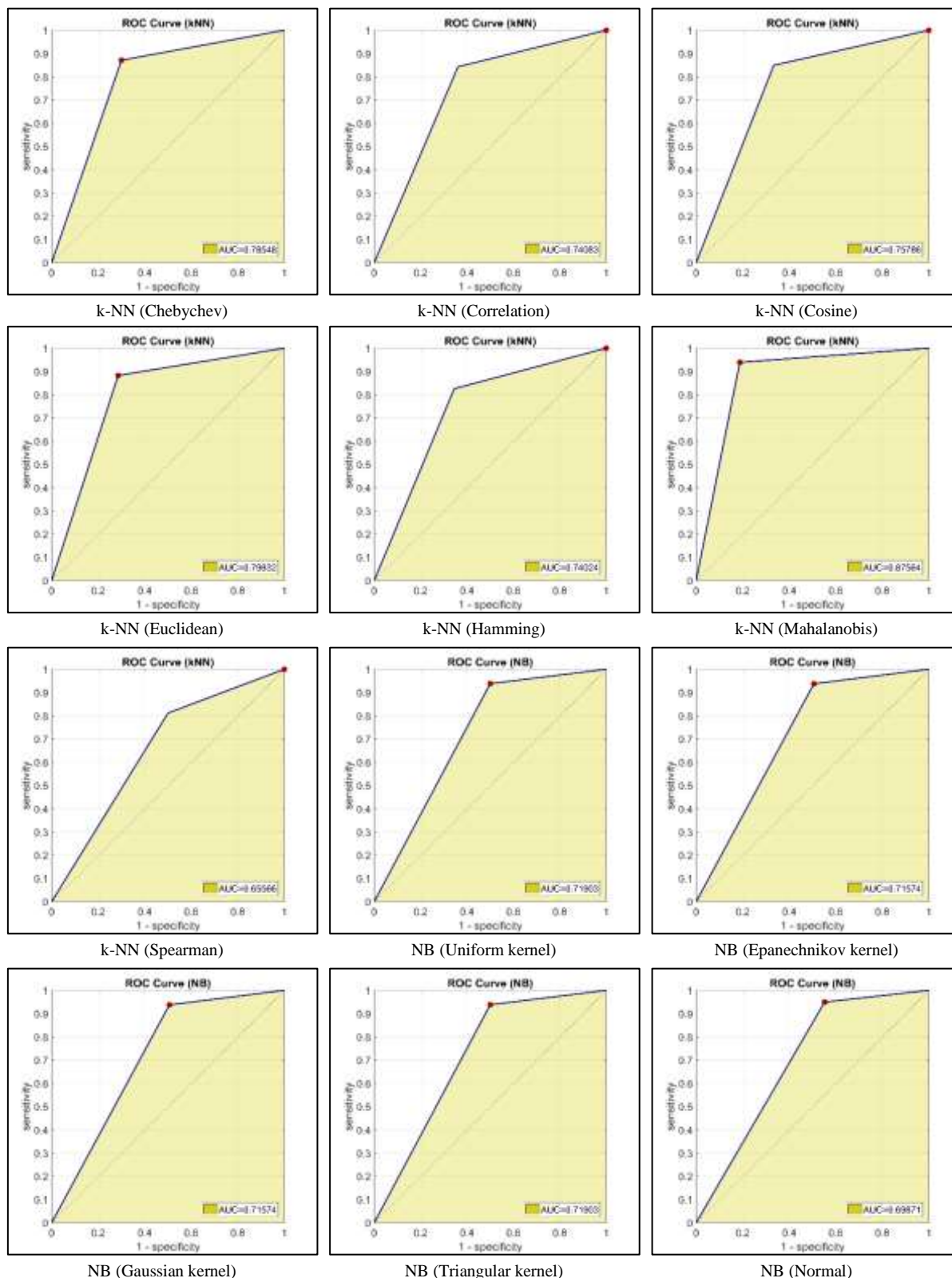**Fig. 3.** Confusion matrix of the k-NN and NB classifiers.

Fig. 4. ROC curves of the k-NN and NB classifiers.

4. Conclusions

There is no standard treatment for Parkinson's disease currently. Symptoms can be controlled by medicine and surgery. The ability to control this disease depends largely on early and accurate diagnosis of the disease. Over the years, several studies focusing on biological, chemical, and genetic areas have been published that

involve the fields of computer sciences. Machine Learning techniques have played a key role in development of speech-based computer-aided diagnostic techniques for Parkinson's disease. In this study, a study that can be used as a decision support system that performs statistical analysis of these data by using biomedical sound parameters of Parkinson's patients was performed. Diagnosis of Parkinson's disease was performed using machine

learning algorithms. In addition, performance evaluation of the methods used in the k-NN and Naive Bayes algorithms was performed. The best score in the classifiers was obtained from the Mahalanobis method for k-NN. AUC value 0.87564 was obtained. In Naive Bayes algorithm, the best result was obtained from Triangular kernel and Uniform kernel. AUC value 0.71574 was obtained. In general, the k-NN classification algorithm had better results than Naive Bayes classification algorithm. As a result of this study, it was determined which of these two algorithms to choose. Moreover, which methods of these algorithms gave better results was determined by statistical values.

References

[1] Adams, R.D., And Victor, M., And Ropper, AH., 1997, Principles of Neurology 6th ed., NY: McGraw-Hill, New York, 1067-1078.

[2] Waters, CH., Çev: Büyükkal B., 2000, Parkinson Hastalığının Tanı ve Tedavisi , Turgut Yayıncılık ve Tic. A.ş., İstanbul.

[3] Apaydın, H. (2001). Parkinson hastalığının klinik özellikleri. Parkinson Hastalığı ve Hareket Bozuklukları Dergisi, 4(2), 114-24.

[4] Hoehn, M.M., Yahr, M.D. (1967). Parkinsonism: onset, progression and mortality. Neurology, 17, 427-42.

[5] Tugwell, C. (2008). Parkinson's Disease on Focus. Chicago: Pharmaceutical Press, 7-8.

[6] Aydınlı, I. (2005). Ağrının Patofizyolojisi. Türk Fiz. Tıp Rehab. Derg., 51 (özel ek B), 8-13.

[7] Ford, B., Louis, E.D., Greene, P. and Fahn, S. (1996). Oral and genital pain syndromes in Parkinson's disease. Movement Disorder, 11(4), 421-6.

[8] Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., Ramig, L.O. (2009). Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. IEEE Transactions on Biomedical Engineering, 56(4), 1015-1022.

[9] 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.

[10] 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.

[11] Tsanas, A., Little, M.A., Mcsharry, P.E. And Ramig, 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.

[12] Fahn, S., and Przedborski, S., 2000, Parkinsonizm. In: Rowland Lp Ed. Merritt's Textbook of Neurology. 10th Ed., Lippincott Williams & Wilkins, Philadelphia, 9789752771819, pp. 679-693.

[13] Quinn, N., And Critchley, P., And Marsden, CD., 1987, Young onset Parkinson's Disease. Mov Disord., 2(2), pp.73-91.

[14] Apaydın, H., Özekmekci, S. (2008). Parkinson Hastalığı: Hasta ve Yakınları için El Kitabı. İstanbul: Parkinson Hastalığı Derneği, 6-15.

[15] Elen, A., & Avuçlu, E. (2019). Diagnosis of Parkinson's Disease Using K-NN Classifier with Various Distance Measurements, 1st International Science and Innovation Congress, pp. 2-8, Pamukkale, Denizli/TR.

[16] Avuçlu, E., & Elen, A. (2020). Evaluation of train and test performance of machine learning algorithms and Parkinson diagnosis with statistical measurements. Med Biol Eng Comput 58, 2775–2788. doi: 10.1007/s11517-020-02260-3

[17] Zhou, J., Li, C., Arslan, C.A. et al. Performance evaluation of hybrid FFA-ANFIS and GA-ANFIS models to predict particle size distribution of a muck-pile after blasting. Engineering with Computers (2019) doi: 10.1007/s00366-019-00822-0

[18] Kishor, N., Singh, S.P., Raghuvanshi, A.S. et al. Engineering with Computers (2007) 23: 71. doi: 10.1007/s00366-006-0024-z

[19] Salim Lahmiri, Amir Shmuel, Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer's disease, Biomedical Signal Processing and Control, Volume 52, 2019, Pages 414-419, ISSN 1746-8094, doi: 10.1016/j.bspc.2018.08.009.

[20] Cover, T., and Hart, P., "Nearest Neighbor Pattern Classification", IEEE Transactions on Information Theory, 13(1): 21-27 (1967).

[21] Orhan, U. and Adem, K., "The Effects of Probability Factors in Naive Bayes Method", Elektrik-Elektronik ve Bilgisayar Mühendisliği Sempozyumu, Bursa, 722-724 (2012).

[22] 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).