



A new approach: information gain algorithm-based k-nearest neighbors hybrid diagnostic system for Parkinson's disease

Cüneyt Yücelbaş¹

Received: 18 February 2021 / Accepted: 9 April 2021 / Published online: 14 April 2021
© Australasian College of Physical Scientists and Engineers in Medicine 2021

Abstract

Parkinson's disease (PD) is a slow and insidiously progressive neurological brain disorder. The development of expert systems capable of automatically and highly accurately diagnosing early stages of PD based on speech signals would provide an important contribution to the health sector. For this purpose, the Information Gain Algorithm-based K-Nearest Neighbors (IGKNN) model was developed. This approach was applied to the feature data sets formed using the Tunable Q-factor Wavelet Transform (TQWT) method. First, 12 sub-feature data sets forming the TQWT feature group were analyzed separately after which the one with the best performance was selected, and the IGKNN model was applied to this sub-feature data set. Finally, it was observed that the performance results provided with the IGKNN system for this sub-feature data set were better than those for the complete set of data. According to the results, values of receiver operating characteristic and precision-recall curves exceeded 0.95, and a classification accuracy of almost 98% was obtained with the 22 features selected from this subgroup. In addition, the kappa coefficient was 0.933 and showed a perfect agreement between actual and predicted values. The performance of the IGKNN system was also compared with results from other studies in the literature in which the same data were used, and the approach proposed in this study far outperformed any approaches reported in the literature. Also, as in this IGKNN approach, an expert system that can diagnose PD and achieve maximum performance with fewer features from the audio signals has not been previously encountered.

Keywords Parkinson's disease · Information gain approach · KNN · Artificial intelligence systems · Speech signals

Introduction

Dopamine deficiency is the most significance reason for the occurrence of Parkinson's disease (PD) [1]. Dopamine, a chemical that transmits information between brain regions controlling the body movements, is produced by a subset of cells found in a specific part of the human brain [1]. In short, dopamine enables people to perform their movements fluently and harmoniously [1]. In humans, the cells that produce this chemical start to decrease in the later years. When this loss is between 60 and 80%, dopamine cannot be produced in sufficient quantity and motor disorders, one of the symptoms of PD, occurs [1]. Symptoms of the disease are more prominent in people between 40 and 70 years of age and mostly occur in the 60s [1]. The incidence of this

disease is higher in males than in females, and it is accepted that one in every 100 male individuals over the age of 65 in the community is suffering from PD [1]. The first symptom indicating a possible diagnosis of Parkinson's disease is slow movements and, in addition, the presence of tremors during periods of rest [1]. In this disease, symptoms, such as slow movements, mask-like expressions, cramped handwriting, tremors, muscle contraction, postural, gait, speech, and smelling disorders, kyphosis, the feeling of discomfort, restless leg syndrome, and forgetfulness are observed [2]. In a previous study, it was emphasized that early diagnosis of PD is possible with a simple blood test in which the risk of presenting with this disease can be revealed before the associated symptoms occur, and the necessary medical measures can be taken accordingly [3].

In the field of engineering, voice or walking recordings are mainly used for automatic detection of PD [4]. A speech disorder is an early distinctive symptom of this disease and develops in the majority of people with PD (about 90%) [5, 6]. For this reason, sound-related features have been

✉ Cüneyt Yücelbaş
cuneytyucelbas@hakkari.edu.tr; cuneytyucelbas@gmail.com

¹ Electrical-Electronics Engineering Department, Hakkari University, 30000 Hakkari, Turkey

widely used in systems for automatic recognition of PD. The purpose of these studies was to automatically differentiate patients from healthy individuals by using the relevant audio features [7]. For instance, Sakar and Kursun designed a tele-diagnosis system for automatic recognition of PD [7]. During the testing of this system, various attributes were extracted from the records of patients and healthy people and then were evaluated with the Support Vector Machine (SVM) method [7]. Considering the characteristics of the data set on which they were working, the authors attempted to obtain the maximum classification accuracy with the least set of features [7]. In another example, the effectiveness of vocal characteristics in PD diagnosis was analyzed using machine learning techniques [8]. As a result of these analyses, the highest classification accuracy was obtained with an SVM of 96.4% [8]. After collecting various audio recordings of people with and without PD and extracting the necessary features, Sakar et al. gave these records to several classifiers and analyzed the results [9]. Braga et al. conducted research on various available data sets in which audio signals were processed for the automatic detection of PD [10]. In another research, an auto-diagnostic system using the fuzzy k-nearest neighbors (FKNN) classification method for PD diagnosis was presented [11]. In a study by Parisi et al., a new hybrid artificial intelligence system was presented for early diagnosis of Parkinson [12]. In a similar study, SVM based on bacterial foraging optimization (BFO-SVM), a new hybrid diagnostic method, was developed for PD recognition [13]. In addition, the Random Forest-BFO-SVM structure in which the feature selection stage was performed was applied to the data and a result of 97.42% was obtained [13]. Another researcher used four different classification systems: Neural Networks (NN), DMneural, Regression, and Decision Tree for effective detection of PD. The best result (92.2%) was obtained with the NN [14]. Lahmiri et al. tested various classification systems, such as linear discriminant analysis (LDA), k nearest-neighbors (KNN), naive Bayes (NB), regression trees (RT), radial basis function neural networks (RBFNN), SVM, and Mahalanobis distance and found that the best result was obtained with SVM with a rate of 92% in the automatic detection of PD [15]. In another study conducted with respect to classification systems, parallel feed-forward neural network architecture was presented for the same purpose [16]. As a result, it was emphasized that a 9-parallel neural network system works better by 8.4% compared to a single network structure [16]. Eskidere et al. tested the methods of SVM, Least Square SVM, (LS-SVM), Multilayer Perceptron NN (MLPNN), and General Regression NN (GRNN) on an available data set for the purpose of Parkinson's follow-up and concluded that LS-SVM gave the best result [17]. In another study, Benba et al. first applied the technique of mel frequency cepstral coefficients (MFCCs) to multi-type audio recordings

taken from healthy and PD subjects [18]. They then gave the resulting data to the SVM classifier and evaluated the results in which they emphasized that the record of /u/ letter contains more discriminatory analysis than other types of audio signals [18]. A cloud-based framework was presented and achieved a classification performance of 96.6% by the authors in reference [19]. In another study conducted in this area, an FKNN system based on a particle swarm optimization (PSO), named PSO-FKNN, was used to automatically diagnose PD, and an average of 97.47% accuracy was obtained [20]. When studies in this area are examined, data registration, processing, feature extraction, selection, and classification processes were carried out in almost all of them. In studies conducted in this field of engineering, the focus has been on automatic recognition of PD with high classification performance.

In this study, the combined Information Gain Algorithm-based K-Nearest Neighbors (IGKNN) approach was proposed for operating with high accuracy and automatically diagnosing PD from the audio signals of the individual. For the presented system, the attributes extracted from the previously recorded audio recordings from 252 people [21] were used as a data set. These data was taken from the University of California Irvine (UCI) Machine Learning Repository. The selected data were separated as a training-test by virtue of the stratified cross-validation (CV) method. The KNN classifier, which exhibits high performance against noisy data such as audio signals, was used in the automatic PD diagnostic system. The performance results obtained from the selected algorithm were evaluated with many statistical criteria. This study aimed at investigating the effect of the Information Gain approach, which has not previously been used in PD diagnoses in the literature. Also, as in this IGKNN approach, an expert system that can diagnose PD and achieve maximum performance with fewer features from the audio signals has not previously been encountered. Moreover, the stratified CV method, which was used as a data segmentation method, has also been viewed as an innovation for PD studies. Considering the low number of subjects used in the studies so far, another purpose of this study was to depict all of the details of this success rate obtained from these 252 subjects.

Methods

Speech data set

Speech disorders, when used in the diagnosis of PD, can be seen as a symptom that can be understood by an expert or even by the surroundings in nearly 90% of the patients. Because the brain's signals controlling the speech and the muscles providing the speech are affected by this

disease, the voice of PD patients is generally softer and monotonous. For this reason, the symptoms that must be quickly noticed by families are changes in speech. When the muscles of the face are stiff or take longer to move, people have difficulty in speaking, and words can be slurred or mumbled [22]. For this reason, it is possible to determine the difference between a PD patient and a healthy person in an early stage using a specialist system although the changes in the audio signals are among the secondary symptoms for the medical diagnosis of PD. For this purpose, the attributes resulting from the use of dissimilar methods based on the speech recordings of 252 subjects (188 PD patients with 107 males and 81 females and 64 healthy subjects with 23 males and 41 females) in the Department of Neurology at Cerrahpaşa Faculty of Medicine, Istanbul University were taken from UCI [21]. The age range of the subjects ranged from 33 to 87. Accompanied by a specialist, the subjects were asked to say the letter /a/ three times, and the necessary data record was provided. The frequency of the microphone used during recording was fixed at 44.1 kHz. As stated in Ref. [21]; after providing information about the data collection process, signed informed consent was taken from all individual participants in accordance with the approval of Clinical Research Ethics Committee, Bahçeşehir University, İstanbul, Turkey. More detailed information was provided by the creators of the data in [21].

Certain characteristics received from the audio signals of patients and healthy people in the diagnosis of PD facilitate the separation of these classes. For instance, the sample curves of standard deviation characteristics extracted from the 36 sub-bands obtained after applying the Tunable Q-factor Wavelet Transform (TQWT) method to the signals are shown in Fig. 1. In the figure, these features are distinctly separated from each other in specific sub-bands.

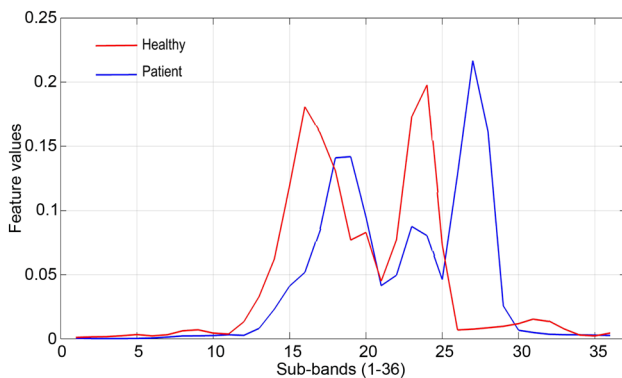


Fig. 1 Sample curves of speech signal attributes in healthy and PD subjects

Analysis of speech signals

Acoustic measurements of sounds obtained from PD patients for analyzing speech signals can be easily obtained without disturbing the patient under the supervision of a specialist physician [22]. Considering the studies in this area, many methods that have been suggested for acoustic sound measurements are outstanding. The most commonly used measurement were jitter, shimmer, and basic frequency irregularities that occur in patient syllables [23]. In addition to these measurements, the harmonic noise ratio parameter, which can reveal hoarseness occurring over time, was also presented as an effective tool [24]. Cepstral peak importance measurements [25], linear prediction modeling [26, 27], auditory modeling [28], and Mel frequency cepstral coefficients [29] could be also preferred in acoustic sounds measurements in PD patients.

The presence of noise in signal processing applications is one of the main factors affecting the system result. The noise in these signals can occur for many reasons, such as the environment, data transmission, and the subject's own body actions. In speech signals, which have an important role in PD detection, irregular vibrations and breathiness often produce noise. It is very important to determine the noise source correctly in the first stage. Noises whose source cannot be detected directly will negatively affect the system performance. Most of the noises have more irregular, random, and high frequency contents compared to the basic signal frequency. However, classical signal processing techniques may be insufficient for detecting and eliminating the above-mentioned noises. Instead, using of the adaptive, adjustable advanced signal processing techniques such as TQWT [30] can achieve a high level of noise cancellation. The performance of the system proposed in this study may be adversely affected by the noise in the data presented as input as done in any algorithm. However, thanks to the TQWT method used in [21], from which the data used in the study were obtained, the rate of exposure of the proposed system was minimized. When Ref. [21] is examined, it could be seen that the feature group obtained as a result of TQWT had more successful classification results by algorithm than the feature groups obtained by other methods.

Attributes of the used data

In this study, the TQWT feature group previously obtained from the speech signals of 252 subjects by the authors as described in reference [21] was used. This feature group consists of sub-bands obtain from the TQWT process. Detailed information on this feature group was given in [21]. The TQWT method is a new discrete wavelet transform form consisting of three basic parameters: Q (Q-factor), j (the number of levels), and r (redundancy). Band-pass filters with

different Q -factors can be generated, and the low and high frequency values of the signal are also separated using this method. The Q from which the method derives its name is derived by dividing the bandwidth of the center frequency by the band-pass filter. This factor can be adjusted according to the oscillation of the signal being processed, thus creating a non-linear separation. When the frequency distribution of the signal is examined, the frequency spectrum of the sudden changing finite signals for the low Q widens. The frequency spectra of the oscillating signals are more localized for a high Q . In short, the Q refers to the oscillation of the signals being processed. j is described as number of levels that will have $j+1$ sub-bands after obtaining high-pass filter and last low-pass filter outputs. The last parameter, r , determines the frequency of the band-pass filters, and as a result of this parameter, TQWT starts to resemble a continuous wavelet transform [30]. Decomposition stages for

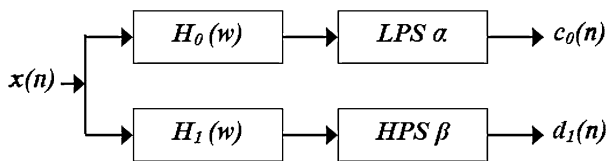


Fig. 2 Analysis steps for single level TQWT [21, 30]

a single level TQWT are given as an example in Fig. 2. In this figure, $x(n)$, $H_0(w)$, $H_1(w)$, LPS , HPS , α , β , $c_0(n)$, and $d_1(n)$ represent input signal, frequency responses of low-pass filter, frequency responses of high-pass filter, low-pass scaling, high-pass scaling, low-pass scaling parameter, high-pass scaling parameter, low-pass sub-band signal, and high-pass sub-band signal, respectively.

In Fig. 3, the decomposition of the speech signals from PD and healthy subjects into sub-bands using TQWT is given. Samples of the signal in this figure were taken from Ref. [31]. Details about this dataset can be found in reference [31].

The performance of the TQWT algorithm is directly dependent on the Q , r , and j parameters. As stated in Ref. [21] from which the dataset used in this study was taken, a large number of trial and error experiments have been conducted to achieve high accuracy rates. In these experiments performed to determine the optimum values, the r parameter was chosen as 3, 4, and 5, respectively. The Q parameter was analyzed for values between 1 and 10. Finally, in order to determine the most appropriate number of levels, the j parameter was tested between 5 and 50 for different Q values. As a result of all these long-term processes, these parameters were determined as $Q=2$, $r=4$, and $j=35$ for the best system performance according to Ref. [21].

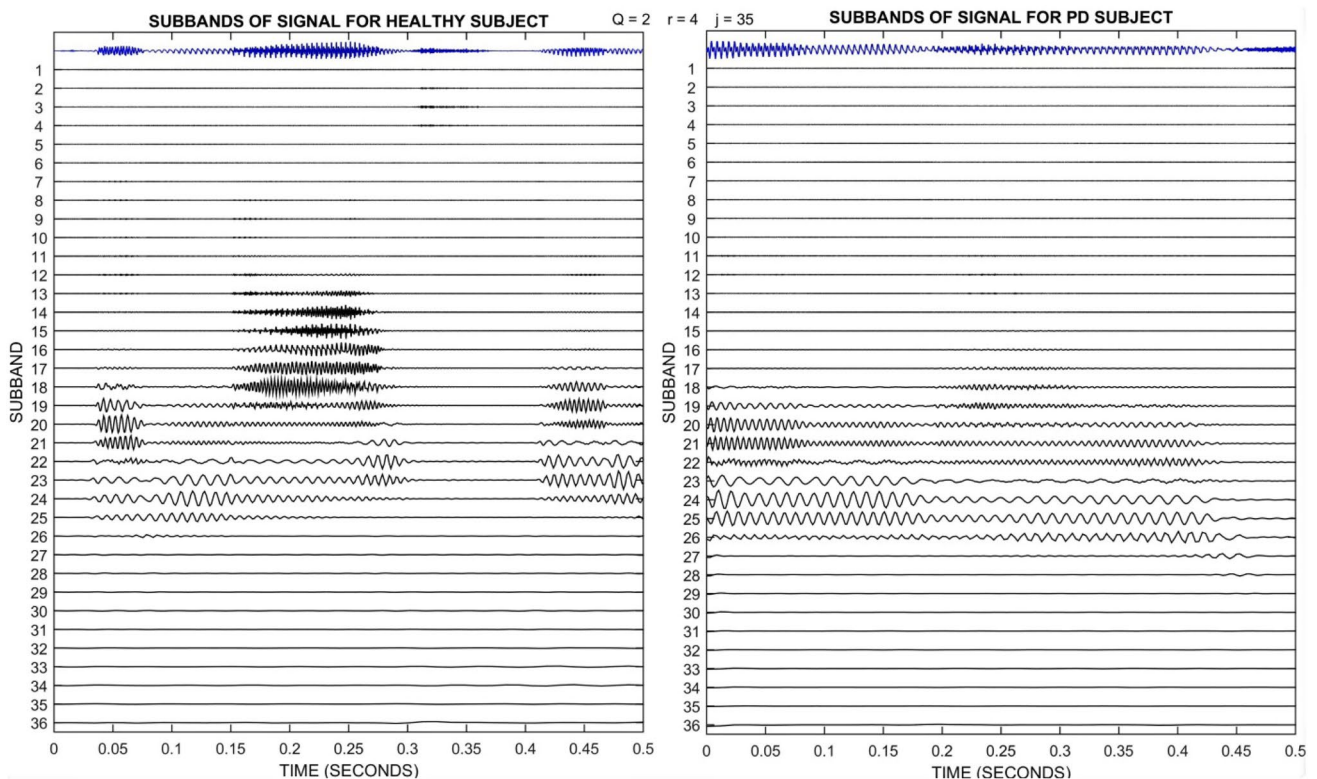


Fig. 3 Decomposition of the sample speech signals into 36 sub-bands using TQWT method

Power spectrum density/power spectrum (PSD) is an analytical method that shows the power of any signal or time series, such as sound, as a distribution on a frequency axis. Thanks to the analysis, information about the noise components in any signal or at which frequencies the signal is effective can be accessed. In this way, preliminary information about the signals can be obtained before the classification of patient and healthy data. The PSD method will make a positive contribution to the classification performance as a result of the highly distinctive features determined by the method if the analyses are sufficiently sensitive. In Fig. 4, the results of the 4th level TQWT application on the sample speech signals in Ref. [31] and the power spectra of the fifth sub-band with the highest energy are given in order to obtain the PSD outputs.

When the sub-bands, their energy ratios, and power spectra obtained from the sample signals [31] in Fig. 4 are examined, it can be seen that patients and healthy subjects can be separated from each other. The sound signals of the healthy subject have a wider range of power values, while those from the patient occur over a more limited range. In addition, this separation will be understood more clearly when the most basic statistical calculations, such as the average, standard deviation, and maximum value of these power spectra are obtained.

TQWT attributes used for this study contain 12×36 parameters. In other words, a total of 432 parameters were obtained by extracting 12 attributes (energy, Shannon entropy, Log Energy entropy, mean Teager–Kaiser energy operator (TKEO), TKEO standard deviation (std), median, mean, std, minimum (min), maximum (max), skewness and kurtosis values) from 36 sub-bands reached as a result of applying the TQWT [21].

Entropy is a measure of the complexity of the data being studied. This criterion cannot be negative [32]. Besides that Shannon entropy (E) is defined by Formula 1 [32]:

$$E = - \sum_{i=1}^N P_i \log_2 P_i \tag{1}$$

In this formula, P_i symbolizes the possibility of the i . data type in the whole data set to be present in all of the data [32].

The Log-Energy entropy (H) attribute formulation is stated below [33]:

$$H(x) = - \sum_{i=1}^{N-1} (\log_2 (P_i(x)))^2 \tag{2}$$

TKEO is a method used to monitor energy in audio signals [34–38]. Formula 3 shows a discrete TKEO formulation:

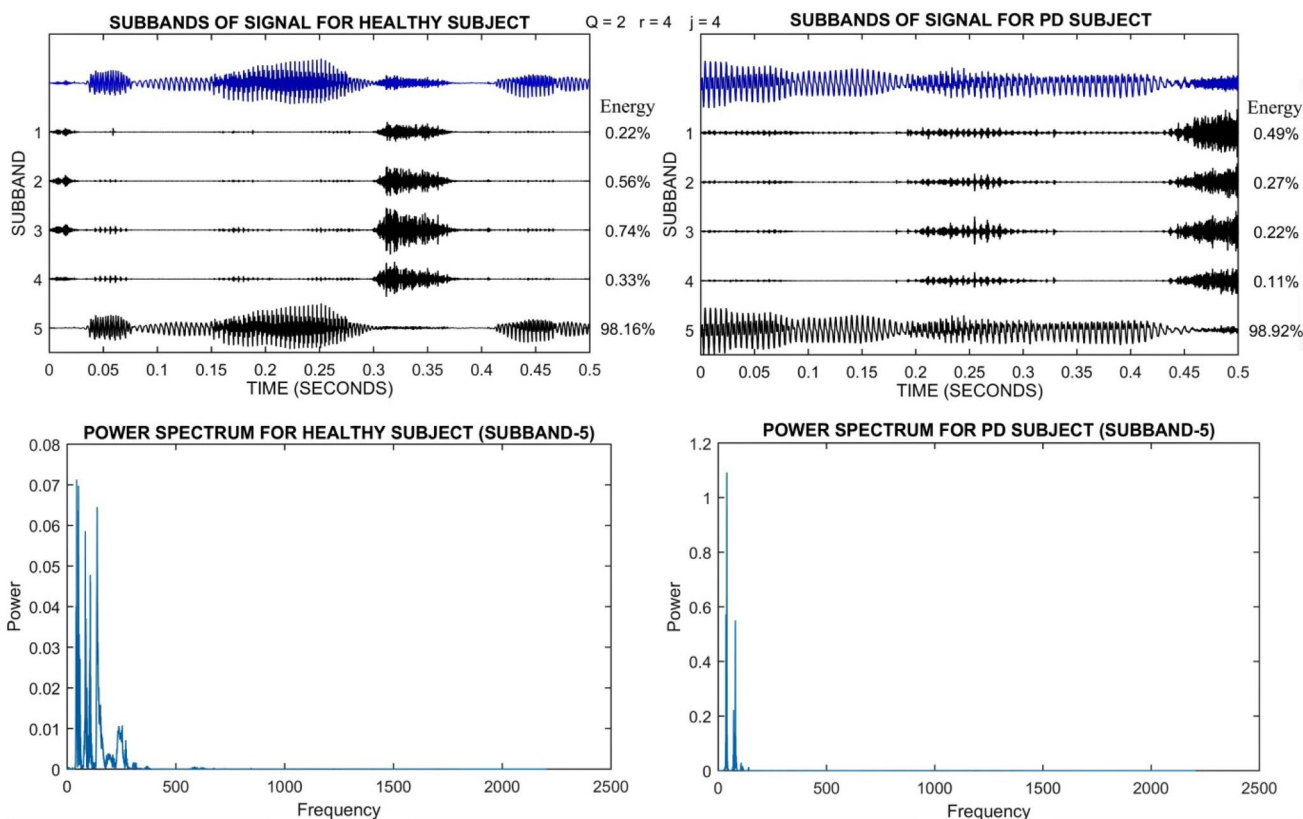


Fig. 4 Power spectrum graphs and 4th level TQWT decomposition of the sample speech signals

$$\psi[x(n)] = x^2(n) - x(n+1)x(n+1) \quad (3)$$

In this formula, data are defined as x and n is the number of samples.

Information gain algorithm-based KNN hybrid model (IGKNN)

Automatic analysis systems using artificial intelligence algorithms can easily diagnose diseases with high accuracy, similar to diagnosis made by a physician. These systems analyze the data entered by a selected classifier or a clustering algorithm. They also statistically demonstrate the accuracy of the result of the evaluation system. In this study, a new combined IGKNN approach was proposed for the features analysis process. The information gain (IG) algorithm and KNN classifier, which exhibits high performance against noisy data such as audio signals, were chosen for this model. Figure 5 shows the pseudo-code diagram of the IGKNN hybrid feature analysis system.

The basis of the IGKNN method is the assignment of the feature subset in which the lowest error will be obtained for the used classifier. Thus, the IG method was selected for this purpose. This method is often used in data mining and artificial intelligence topics. IG can be expressed as the opposite of the Entropy concept. This criterion, which takes a value between “0” and “1”, shows how much value can be gained as a result of classification according to the given feature. The fact that the calculation is close to “1” is proof that the related feature plays an active role in the parsing of classes [39, 40]. In order to calculate the IG criterion, entropy for

each class label must be calculated. Entropy, a measure of uncertainty in the system, was calculated using formula 4:

$$(T) = - \sum_{i=1}^n P_i \log_2(P_i) \quad (4)$$

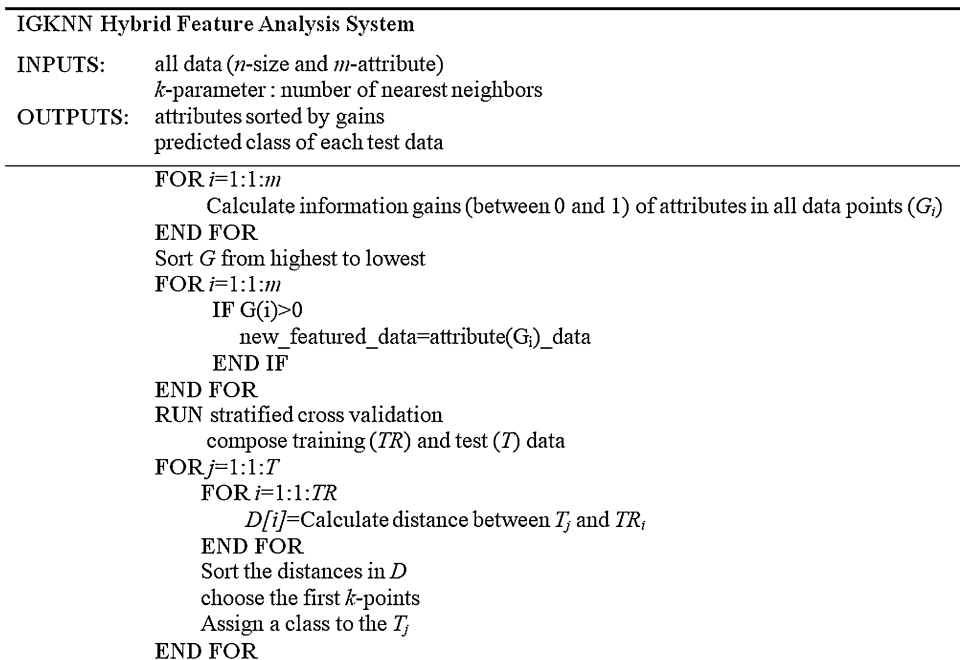
P_i shows the probability that each class tag is contained in a data set with n class tags. Also, formula 5 was used to find IG ranging from “0” to “1”.

$$(x, T) = (T) - \sum_{i=1}^n \frac{|T_i|}{|T|} H(T_i) \quad (5)$$

In addition, T and x are the data set and class type to be calculated, respectively [39, 40]. Besides the feature elimination algorithm, KNN algorithm [41] was selected as the classifier for the classification process of the IGKNN hybrid system. In the KNN algorithm, Euclidean, Manhattan, and Minkowski functions have been tried in distance calculations. As a result of these trials, the best performance results were obtained with the Euclidean function. In addition, the k -parameter, which is the algorithm input, was tried from “1” to “20”, and the best result was achieved with a value of “1”. The classes of healthy and patient were labeled as 0 and 1, respectively. Each data was divided tenfold by the Stratified cross validation (CV) method. In this CV method, each fold has an approximately equal percent sample for each class.

Big-O notation is used for the computational complexity calculation of the proposed combined IGKNN approach. This notation is often used in computer science to refer to the worst case scenario of an algorithm [42]. The computational

Fig. 5 The pseudo code diagram of IGKNN approach



complexity calculation of the KNN algorithm used as a classifier is the same as the analysis system proposed in this study. The computational complexity of KNN is proportional to the number and dimensions of training data. Accordingly, assuming p is the number of sizes and m is the number of dimensions in training data, the computational complexity value of the proposed method is obtained as $O(pm)$. A low value means that the algorithms respond quickly and take up less memory space. Since the KNN classification algorithm operates on the basis of “instance-based” [42], it needs more memory. This situation negatively affects the computational complexity value of the algorithm or the system and causes it to increase. If data with hundreds/thousands of dimensions are presented to the system in operations such as text classification, the computational complexity will increase. Consequently, poor classification performance will likely be achieved. Methods, such as feature selection/elimination and dimension reduction, are used to solve this problem, which is generally caused by the size of the data. In the proposed system in this study, IG was used as a feature analysis algorithm. In this way, the analysis of each feature of the samples was made directly. As a result, the computational complexity ($O(pm)$) of the proposed method can be reduced by changing the value of m .

Statistical evaluation processes

In expert systems designed with artificial intelligence algorithms, it is necessary to form a confusion matrix to reveal the number of true and false labeling obtained by automatic detection. Various statistical criteria may be calculated using the number of labels specified in this matrix. In this study, statistically valid results were obtained using several criteria: True Positive rate (TP rate), also named as sensitivity or recall; False Positive rate (FP rate) F measure (F) Matthews Correlation Coefficient (MCC) classification accuracy rate (ACC); Precision (Prec) [43–46] and Cohen’s Kappa Coefficient (Kappa) [47]. Besides these statistical criteria, the values of receiver operating characteristic and precision-recall curves (ROC and PRC, respectively) [48] were computed.

Results

For this study, the attributes obtained from the audio signals of 252 subjects (188 PD patients and 64 healthy people) [21] were under physician supervision, the /a/ letter was repeated three times by these people and the data recording process was obtained. As a result, the total number of recordings was 756 (252×3).

In the first phase of the study, the analysis of the sub-feature sets of the TQWT feature group was started. This feature group consisted of 36 sub-bands, and 12 features were extracted from each sub-band, so 12 sub-feature groups were formed [21]. The tenfold stratified CV method was applied to create training-test data in all analysis processes after this stage. Initially, all of the 12 sub-feature sets were presented to KNN and IGKNN in order to compare the system performances. The results obtained for this comparison are given in Table 1.

When Table 1 is examined, the success rate for KNN was obtained as 90.74% using all 432 sub-bands in total, while it was 94.97% for IGKNN using only 108 sub-bands. These results alone prove the superiority of the proposed combined IGKNN system. In addition, when the proposed system was used, a higher performance was obtained with less sub-band value by performing a feature analysis. This situation contributes to the reduction of the computational complexity of the system. In the next step, which of the 12 sub-feature sets was more effective was investigated. Thus, Table 2 shows the classification results for each of the 12 sub-feature sets of TQWT with the KNN algorithm. As shown in Table 2, the best statistical result among all groups was obtained with Log Energy entropy (LEE) sub-feature group. The ACC rate of this sub-feature group was calculated as 95.76% at maximum. Also, the values of ROC and PRC reached almost the 0.95 band with this sub-feature group, which was much closer to the perfect classification result. The LEE sub-feature group was followed by Std value and TKEO mean with 92.85% and 87.96% ACC rates, respectively.

The same feature sets were also analyzed in detail using the IGKNN system. In this way, the effects of each sub-feature set on both systems could be seen more clearly. Table 3

Table 1 The best results for all of the 12 sub-feature sets of TQWT with KNN and IGKNN

Statistical criteria	TNI	CCI	Kappa	TP rate	FP rate	Prec	F	MCC	ACC (%)	ROC	PRC
KNN (TNS = 432 NNS = 432)	756	686	0.751	0.907	0.169	0.906	0.907	0.752	90.74	0.869	0.874
IGKNN (TNS = 432 NNS = 108)	756	718	0.867	0.95	0.082	0.95	0.95	0.867	94.97	0.937	0.932

TNI total number of instances, CCI correctly classified instances, k-parameter = 1, distance function: Euclidean, TNS total number of sub-bands ($12 \times 36 = 432$), NNS number of sub-bands selected

Table 2 The best results for each of the 12 sub-feature sets of TQWT with KNN (TN) total number of instances, CCI correctly classified instances, k-parameter = 1, distance function: Euclidean)

Statistical criteria	Energy	Shannon entropy	Log energy entropy	TKEO mean	TKEO std	Median value	Mean value	Std value	Min value	Max value	Skewness value	Kurtosis value
TNI	756											
CCI	632	656	724	665	631	546	535	702	641	634	555	601
Kappa	0.565	0.646	0.887	0.68	0.568	0.173	0.107	0.812	0.599	0.585	0.284	0.446
TP rate	0.836	0.868	0.958	0.88	0.835	0.722	0.708	0.929	0.848	0.839	0.734	0.795
FP rate	0.272	0.231	0.076	0.202	0.259	0.569	0.615	0.114	0.248	0.237	0.455	0.358
Prec	0.835	0.866	0.957	0.879	0.837	0.691	0.666	0.929	0.848	0.844	0.729	0.791
F	0.836	0.867	0.958	0.879	0.836	0.701	0.679	0.929	0.848	0.841	0.731	0.793
MCC	0.566	0.646	0.888	0.681	0.569	0.179	0.113	0.812	0.599	0.586	0.285	0.447
ACC (%)	83.59	86.77	95.76	87.96	83.46	72.22	70.76	92.85	84.78	83.86	73.41	79.49
ROC	0.78	0.809	0.944	0.835	0.775	0.574	0.553	0.903	0.788	0.79	0.634	0.706
PRC	0.793	0.819	0.942	0.841	0.789	0.657	0.647	0.901	0.799	0.798	0.69	0.74

shows the results obtained with the number of sub-bands selected for each feature set.

When Table 3 was examined, it could be seen that according to Table 2, an increase was achieved in the performance results of all sub-feature sets with the exception of Energy. Although no increase in the energy feature occurred, the number of sub-bands was reduced from 36 to 12, and almost identical results were achieved. Achieving the same performance with less sub-bands contributed to the reduction of computational complexity.

In the next stage of the study, the IGKNN system was used to perform double and triple analyses of the sub-feature groups with the best classification performance. These groups were created by adding the second (Std value) and the third (TKEO mean) best sub-feature groups next to the LEE from which the best results were obtained in both KNN and IGKNN. Later, these combined feature groups were submitted to the IGKNN system for the necessary evaluation and classification processes. The results obtained as a result of these processes are given in Table 4.

When Tables 2 and 3 are examined in terms of statistical performance results, it is seen that better results was obtained for LEE sub-feature group after using IGKNN system. The ACC performance result increased almost 2% with effective LEE feature of 22 sub-bands. Namely, 737 of 756 input instances were correctly classified. Moreover, the Kappa value was calculated as 0.933, and determining this value above 0.8 showed that perfect agreement between actual and predicted values existed. In addition, the TP rate, Prec, F, and MCC criteria closest to “1” value was statistically supported by the ACC rate. When the FP rate criterion is considered, it reached the lowest value (0.05) among all FP rate calculations in the study. According to the results in Table 4, the LEE sub-feature group was followed by “LEE-Std value and LEE-Std value-TKEO mean” with 96.69% and 96.42% ACC ratios, respectively. As can be seen from Table 4, as the number of features in the groups increases, the performance results decrease slightly. In spite of this, the decrease in the number of NSS contributes to the decrease in the computational complexity of the system as mentioned previously.

Finally, in Table 5, only the information gain rates of the LEE sub-feature set from which the best performance was obtained are given as an example. According to this table, the ratios of 14 out of 36 features were obtained as “0”. The other 22 features were selected for the next step, which was the classification process.

Figure 6 shows the ROC and PRC curves for the values indicated in the LEE sub-feature group in Table 4. The ROC curve area is the most preferred ROC statistic. Additionally, the balance between precision and recall should be created because these metrics are inversely related. The balance between these two metrics is stated by the PRC curve. In

Table 3 The best results for each of the 12 sub-feature sets of TQWT with IGKNN (TNI total number of instances, CCI correctly classified instances, k-parameter = 1, distance function: Euclidean, TNS total number of sub-bands, NSS number of sub-bands selected)

Statistical criteria	Energy	Shannon entropy	Log Energy entropy	TKEO mean	TKEO std	Median value	Mean value	Std value	Min value	Max value	Skewness value	Kurtosis value
TNI	756											
CCI	630	658	737	674	647	554	545	705	659	663	571	610
Kappa	0.56	0.655	0.933	0.711	0.613	0.198	0.161	0.819	0.658	0.669	0.337	0.468
TP rate	0.833	0.87	0.975	0.892	0.856	0.733	0.721	0.933	0.872	0.877	0.755	0.807
FP rate	0.273	0.219	0.05	0.185	0.252	0.558	0.579	0.123	0.219	0.221	0.427	0.358
Prec	0.833	0.87	0.975	0.891	0.854	0.703	0.687	0.932	0.871	0.875	0.749	0.8
F	0.833	0.87	0.975	0.891	0.855	0.711	0.698	0.932	0.871	0.876	0.752	0.802
MCC	0.56	0.656	0.933	0.712	0.614	0.207	0.168	0.82	0.659	0.67	0.338	0.47
ACC (%)	83.33	87.03	97.48	89.15	85.58	73.28	72.08	93.25	87.16	87.69	75.52	80.68
ROC	0.754	0.808	0.959	0.841	0.792	0.591	0.577	0.904	0.807	0.813	0.644	0.727
PRC	0.778	0.82	0.961	0.847	0.807	0.667	0.659	0.906	0.818	0.824	0.699	0.755
TNS	36											
NSS	12	28	22	21	22	28	21	20	25	17	30	11

this study, the ROC and PRC criteria exceeded the value of 0.95 after using the IGKNN system and again demonstrated the high success of classification.

Discussion

The PD is a stealthy brain disorder that progresses slowly. As in any disease, early diagnosis is likely to improve PD patients' quality of life. People diagnosed in the early stage of the disease can shape their new life according to this situation and take the necessary precautions. There are several diagnostic methods of this disease, including the analysis of the audio signals [49]. Expert systems can be used in real-life applications in the areas with medical deficiencies for PD, which can be diagnosed from these signals under specialist supervision. These systems can also help physicians strengthen the diagnosis with high success rates in existing health institutions. In addition, if the proposed system can be made available to people online, this type of system can contribute to directing the individuals suspected of having this disease to a specialist physician in the field. For this purpose, in this study, it was proposed to design an expert analysis system that could work with fast and high accuracy in real-life in addition to the virtual environment and can automatically diagnose PD from the audio signals of the individual.

In the literature, data recording and processing, feature extraction and selection, and also classification processes have been generally used in PD studies. A detailed analysis of PD-related studies is given in Table 6. The studies in this table were compared according to the number of data, experimental methods, and performance results of these systems. Accuracy rates obtained in the studies ranged from 82.5 to ~ 100%. The main goal of the studies in this field is to obtain maximum performance with the available data. Although the number of subjects in this paper and other studies [21, 50–53] using the same data was 252, this number changed between 31 and 50 for other studies.

As seen in Table 6, the data set used in this study has also been examined in other studies so far [21, 50–53], and some comparisons were made with these studies in terms of the process and results. First, the feature groups formed in these relevant studies [21, 50–53] were classified by several classifiers, but the classifier in which the best result was obtained for each feature group usually changed. Such situations are undesirable as they are restrictive for expert systems in terms of time and process intensity. While starting this study, the classifier choice was switched to the KNN. Second, the sub-feature data sets belonging to the TQWT feature group formed in the above-mentioned study [21] were not given separate classification systems and were not analyzed. However, in this study, great emphasis was placed on the

Table 4 The best results for double and triple analysis of the most successful sub-feature groups with IGKNN system (TNI total number of instances, CCI correctly classified instances, k-parameter = 1, distance function: Euclidean, TNS total number of sub-bands, NSS number of sub-bands selected)

Statistical criteria	LEE sub-feature group	“LEE-Std value” combined sub-feature group	“LEE-Std value-TKEO mean” combined sub-feature group
TNI	756		
CCI	737	731	729
Kappa	0.933	0.912	0.904
TP rate	0.975	0.967	0.964
FP rate	0.05	0.059	0.067
Prec	0.975	0.967	0.964
F	0.975	0.967	0.964
MCC	0.933	0.912	0.905
ACC (%)	97.48	96.69	96.42
ROC	0.959	0.948	0.945
PRC	0.961	0.949	0.948
TNS	36	72	108
NSS	22	17	16

sub-feature data sets of this feature group from which the best result was obtained, and efforts were made to decrease the data density of the expert system. Last, in the related studies [21, 50–53], some attribute selection methods were implemented for the whole data set. As a result, an ACC in the range of 86% to 96.83% was achieved with 20–555 features. This situation forces the whole system to search

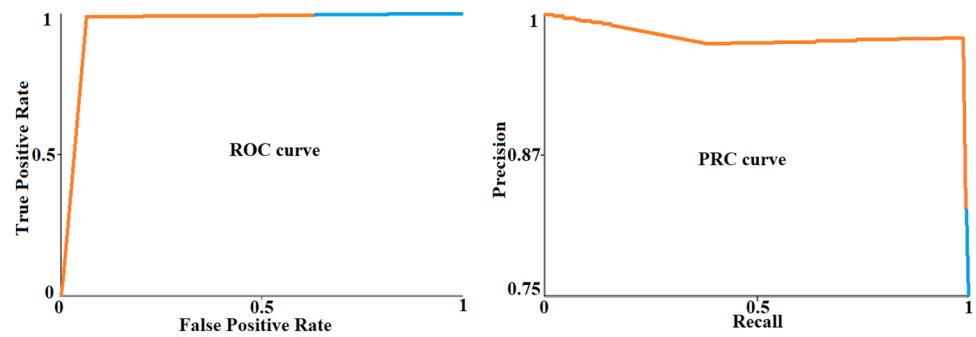
for an effective feature in all data groups. Therefore, the expert system may have difficulty in terms of data density and processing time. Despite this difficulty, in this study, a much higher ACC rate (approx. 98%) was obtained using 22 features with the IGKNN mechanism. Also, the effectiveness of the proposed approach was supported by multiple statistical metrics, such as Kappa, Prec, ROC, and PRC.

Table 5 Gain ratios of LEE sub-feature set for 22 sub-bands

Gain ratio	Attribute name
0.1187	LEE feature of sub-band 12
0.1188	LEE feature of sub-band 11
0.089	LEE feature of sub-band 15
0.088	LEE feature of sub-band 17
0.0849	LEE feature of sub-band 35
0.0798	LEE feature of sub-band 16
0.0667	LEE feature of sub-band 26
0.0664	LEE feature of sub-band 6
0.0629	LEE feature of sub-band 5
0.604	LEE feature of sub-band 10
0.0592	LEE feature of sub-band 9
0.0449	LEE feature of sub-band 28
0.0433	LEE feature of sub-band 19
0.0412	LEE feature of sub-band 4
0.0396	LEE feature of sub-band 3
0.0379	LEE feature of sub-band 23
0.0346	LEE feature of sub-band 21
0.0282	LEE feature of sub-band 36
0.0241	LEE feature of sub-band 32
0.0237	LEE feature of sub-band 1
0.0225	LEE feature of sub-band 29
0.0203	LEE feature of sub-band 2

When examining other studies in the literature that were conducted with different data sets for the diagnosis of PD based on audio signals, no other studies using the Information Gain features analysis approach or IGKNN hybrid system were found. Moreover, the fact that the number of subjects whose audio signals have been obtained in other studies is relatively low compared to this study further reinforces the importance of this study.

Although the hierarchical IGKNN system presented in this study achieves successful results for the early diagnosis of PD, it also contains some limitations. The limitations of the proposed approach should be expressed under two main headings. The first one is related to the IG algorithm used for feature selection. An important disadvantage of this method has a poor performance against features that contain different variables (such as “date: 19_8_1996”). For this reason, it is necessary to be selective against the data to be presented to the proposed sequential system. Otherwise, the possibility of successful results of the feature selection phase will decrease. Another important limitation is related to the KNN classifier, which is particularly powerful against noisy data, such as audio signals. The disadvantage of this algorithm is the requirement for a large amount of memory space, especially for large data, since it stores the results of all situations while calculating the distance. This limitation is proportional to the number of samples (p) in and size (m) of the data set. This situation directly affects the computational

Fig. 6 ROC and PRC curves of LEE sub-feature group after processing with IGKNN method**Table 6** Comparison of the results in this study with those available in the literature

References	Method—system	Classifier	Number of data	Accuracy (%)
Sakar, Kursun [7]	Mutual information + mRMR (maximum relevance–minimum redundancy)	Support vector machine (SVM)	32 (PD: 24 Healthy: 8)	92.75
Gürüler [54]	k-means clustering based feature weighting	Complex-valued artificial neural network	31 (PD: 23 Healthy: 8)	99.52
Little et al. [55]	Pitch period entropy	Kernel-SVM	31 (PD: 23 Healthy: 8)	91.4
Peker et al. [56]	mRMR attribute selection algorithm	CVANN	31 (PD: 23 Healthy: 8)	98.12
Sakar et al. [8]	Unified Parkinson's Disease Rating Scale score method	Kernel-SVM	50 (PD: 42 Healthy: 8)	96.4
Cai et al. [13]	Relief feature selection method	SVM based on bacterial foraging optimization	31 (PD: 23 Healthy: 8)	97.42
Benba et al. [18]	Mel frequency cepstral coefficients (MFCCs) technique	Kernel-SVM	40 (PD: 20 Healthy: 20)	82.5
Sakar et al. [21] ^a	Tunable Q-factor wavelet transform (TQWT) and mRMR-50	SVM (RBF kernels)	252 (PD: 188 Healthy:64)	86 (with 50 features)
Solana-Lavalle et al. [50] ^a	Wrappers feature subset selection	SVM	252 (PD: 188 Healthy:64)	94.7 (with 20 features)
Gunduz [51] ^a	Combining the sets of features	Convolutional neural networks	252 (PD: 188 healthy: 64)	86.9 (TQWT+MFCC+concat combination features) (432+84+39)
Tuncer et al. [52] ^a	A combination of minimum average maximum (MAMa) tree and singular value decomposition (SVD)	KNN	252 (PD: 188 Healthy:64)	92.46 and 96.83 (with 50 features)
Yücelbaş [53] ^a	Simple Logistic hybrid system with Greedy Stepwise algorithm (SLGS)		252 (PD: 188 Healthy:64)	for males 88.71 (with 11 features) for females 87.15 (with 9 features)
This study ^a	A New Approach: Information Gain Algorithm-based KNN Hybrid Model (IGKNN)		252 (PD: 188 Healthy:64)	~98 (with 22 features)

^aIn these studies, the same data set was used

complexity ($O(pm)$) of the proposed system. The value was reduced by eliminating the ineffective attributes of the samples used in the study as a result of the necessary procedures. Other limitations of the KNN classifier used in the system are hyper-parameters, such as the number of k-neighbors that affect the system performance and the distance calculation criterion. When it comes to the general limitations of the study, the major deficiency of this and similar studies is the lack of testing of the proposed systems due to the limited amounts of data. In addition, noises in audio signals can adversely affect system performance. Another common deficiency of the studies in this area is the lack of raw signal states in the data in the ready-made data banks.

In the literature, besides classifiers (such as ANN, SVM), data segmentation methods (such as k-fold CV, leave-one-out CV) and signal processing techniques (such as Fourier and wavelet) were used. To summarize, a few studies that incorporate a new approach are available in this field. Thus, in this study, the new proposed approach system makes a significant contribution (in terms of both the data segmentation method and the feature selection method) to the literature. In addition, a detailed analysis of attributes extracted from speech signals with the proposed IGKNN system has not been previously reported in the literature. Furthermore, the large amount of data in our study compared to similar studies in this field is another factor that makes this study important.

Conclusions

In artificial intelligence-based automatic diagnosis systems, data preparation, method implementation, feature extraction–selection–reduction, and dimension change methods are important and necessary steps. As a result of all of these stages, development the performance results of the selected classification system was the aim of this study. In this study, a novel approach (IGKNN approach) was recommended for diagnosing PD with high accuracy based on audio signals. For this system, the attributes extracted from the previously recorded speech signals of 252 people [21] were used as a data set; data were taken from the UCI. These recordings were separated as a training-test by virtue of the tenfold Stratified CV method. The KNN algorithm, which is effective against noisy data such as audio signals, was used for the automatic PD diagnostic system. This study purposed to examine the effect of the Information Gain approach, which to my knowledge, has not been previously used in PD diagnosis. Also, as in this IGKNN approach, an expert system that can diagnose PD and achieve maximum performance with fewer features from the audio signals has not been encountered previously. Considering the low number of subjects used in the

studies so far, another goal of this study was to define all of the details of this success rate obtained on 252 subjects. As a result of the proposed system performance, the ACC ratio was obtained as 97.48% with 22 features determined. Also, Kappa coefficient was achieved as 0.933, and calculating this value above 0.8 showed that there was a perfect reliability between actual and predicted values. Moreover, the ROC and PRC areas criteria exceeded the value of 0.95 and demonstrated the high success of classification. In a nutshell, a maximum performance result was obtained with a minimum number of attributes thanks to the IGKNN approach. Furthermore, the number of data in this paper was higher than the other studies.

Gender difference between subjects is a factor emphasized in some voice processing studies. In addition to differences in tone of voice due to gender or age factors, many situations such as accent, mouth, tooth structure, hormonal, race/ethnic differences, and environmental factors (smoking, other habits) can affect the success of voice processing studies [57–60]. However, the main purpose of the studies in this area is to achieve high classification success under optimum system parameters in spite of all of these differences/negativities. In the source from which the data set used in this study was taken [21], differences in voice tone of the subjects or any other negative factors were not mentioned. However, thanks to the presented approach, the classification success was achieved as almost 98%, and this situation demonstrated the success of the study. As a result, it is understood that the selection of 36 sub-bands and the extracted features for both female and male subjects are effective in minimizing the disadvantage that may arise from the stated possible differences. According to this information, in the future, detailed studies can be carried out on the effects of other factors in addition to differences in tone of voice between subjects. Besides, the IGKNN system can be applied to handwriting, gait, and other medical parameters of people with PD. The results of this proposed approach can be developed with larger PD data sets and more significant properties obtained by various methods. In addition, system performance can also be assessed on this existing 12-dimensional data set by using dimension reduction methods, such as a principal component analysis (PCA). In the implementation phase, a new dimensional data matrix obtained by changing the size parameter between 1 and 12 is presented to the system. The most appropriate size is decided according to the performance values recorded for each dimension parameter. Thanks to the reduction of the size in the property space, a reduction in the computational complexity of the algorithm is also found.

Acknowledgements As stated in Ref. [21]; the data from volunteers were obtained in accordance with the approval of Clinical Research Ethics Committee, Bahçeşehir University, İstanbul, Turkey.

Funding No funding was received for conducting this study.

Declarations

Conflict of interest The author declares that he has no conflict of interest.

Ethical standards According to Ref. [21]; all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national Non-invasive Clinical Research Medical Ethics Review Board and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent As stated in Ref. [21]; signed informed consent was obtained from all individual participants.

References

- Shulman LM (2007) Gender differences in Parkinson's disease. *Gend Med* 4(1):8–18
- Jankovic J (2008) Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry* 79(4):368–376
- Foulds PG, Mitchell JD, Parker A, Turner R, Green G, Diggle P, Hasegawa M, Taylor M, Mann D, Allsop D (2011) Phosphorylated α -synuclein can be detected in blood plasma and is potentially a useful biomarker for Parkinson's disease. *FASEB J* 25(12):4127–4137
- Sekine M, Akay M, Tamura T, Higashi Y, Fujimoto T (2004) Investigating body motion patterns in patients with Parkinson's disease using matching pursuit algorithm. *Med Biol Eng Comput* 42(1):30–36
- 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(1):24–29
- Tsanas A, Little MA, McSharry PE, Ramig LO (2010) Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng* 57(4):884–893
- Sakar CO, Kursun O (2010) Telediagnosis of Parkinson's disease using measurements of dysphonia. *J Med Syst* 34(4):591–599
- Sakar BE, Serbes G, Sakar CO (2017) Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease. *PLoS ONE* 12(8):e0182428
- Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S, Apaydin H, Kursun O (2013) Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE J Biomed Health Inform* 17(4):828–834
- Braga D, Madureira AM, Coelho L, Ajith R (2019) Automatic detection of Parkinson's disease based on acoustic analysis of speech. *Eng Appl Artif Intell* 77:148–158
- Chen H-L, Huang C-C, Yu X-G, Xu X, Sun X, Wang G, Wang S-J (2013) An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Syst Appl* 40(1):263–271
- Parisi L, RaviChandran N, Manaog ML (2018) Feature-driven machine learning to improve early diagnosis of Parkinson's disease. *Expert Syst Appl* 110:182
- Cai Z, Gu J, Chen H-L (2017) A new hybrid intelligent framework for predicting Parkinson's disease. *IEEE Access* 5:17188–17200
- Das R (2010) A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Syst Appl* 37(2):1568–1572
- Lahmiri S, Dawson DA, Shmuel A (2018) Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures. *Biomed Eng Lett* 8(1):29–39
- Åström F, Koker R (2011) A parallel neural network approach to prediction of Parkinson's Disease. *Expert Syst Appl* 38(10):12470–12474
- Eskidere Ö, Ertaş F, Hanilçı C (2012) A comparison of regression methods for remote tracking of Parkinson's disease progression. *Expert Syst Appl* 39(5):5523–5528
- Benba A, Jilbab A, Hammouch A (2016) Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with Parkinson's disease and healthy people. *Int J Speech Technol* 19(3):449–456
- Al Mamun KA, Alhussein M, Sailunaz K, Islam MS (2017) Cloud based framework for Parkinson's disease diagnosis and monitoring system for remote healthcare applications. *Futur Gener Comput Syst* 66:36–47
- Zuo W-L, Wang Z-Y, Liu T, Chen H-L (2013) Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach. *Biomed Signal Process Control* 8(4):364–373
- Sakar CO, Serbes G, Gunduz A, Tunc HC, Nizam H, Sakar BE, Tutuncu M, Aydin T, Isenkul ME, Apaydin H (2019) 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* 74:255–263
- Eskidere Ö (2012) A comparison of feature selection methods for diagnosis of Parkinson's disease from vocal measurements. *Sigma* 30:402–414
- Umapathy K, Krishnan S, Parsa V, Jamieson DG (2005) Discrimination of pathological voices using a time-frequency approach. *IEEE Trans Biomed Eng* 52(3):421–430
- Yumoto E, Gould WJ, Baer T (1982) Harmonics-to-noise ratio as an index of the degree of hoarseness. *J Acoust Soc Am* 71(6):1544–1550
- Heman-Ackah YD, Michael DD, Baroody MM, Ostrowski R, Hillenbrand J, Heuer RJ, Horman M, Sataloff RT (2003) Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol* 112(4):324–333
- Parsa V, Jamieson DG (2001) Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Lang Hear Res* 44(2):327–339
- Eskenazi L, Childers DG, Hicks DM (1990) Acoustic correlates of vocal quality. *J Speech Lang Hear Res* 33(2):298–306
- Shrivastav R (2003) The use of an auditory model in predicting perceptual ratings of breathy voice quality. *J Voice* 17(4):502–512
- Godino-Llorente JJ, Gomez-Vilda P (2004) Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Trans Biomed Eng* 51(2):380–384
- Selesnick IW (2011) Wavelet transform with tunable Q-factor. *IEEE Trans Signal Process* 59(8):3560–3575
- Hagen J, Dhaval T, Michael S (2019) Mobile device voice recordings at King's College London (MDVR-KCL) from both early and advanced Parkinson's disease patients and healthy controls. Zenodo. <https://doi.org/10.5281/zenodo.2867216>
- Gray RM (1990) Entropy and information. Entropy and Information Theory, Springer, New York
- Aydn S, Saraoğlu HM, Kara S (2009) Log energy entropy-based EEG classification with multilayer neural networks in seizure. *Ann Biomed Eng* 37(12):2626
- Kaiser JF (1990) On a simple algorithm to calculate the 'energy' of a signal. In: International conference on acoustics, speech, and signal processing, ICASSP-90, IEEE, pp 381–384
- Kaiser JF (1993) Some useful properties of Teager's energy operators. In: IEEE international conference on acoustics, speech, and signal processing, 1993. ICASSP-93, IEEE, pp 149–152

36. Maragos P, Kaiser JF, Quatieri TF (1993) On amplitude and frequency demodulation using energy operators. *IEEE Trans Signal Process* 41(4):1532–1550
37. Solnik S, Rider P, Steinweg K, DeVita P, Hortobágyi T (2010) Teager-Kaiser energy operator signal conditioning improves EMG onset detection. *Eur J Appl Physiol* 110(3):489–498
38. Randall RB, Smith WA (2017) Application of the Teager Kaiser energy operator to machine diagnostics. In: Tenth Dst group international conference on health and usage monitoring systems
39. Karegowda AG, Manjunath A, Jayaram M (2010) Comparative study of attribute selection using gain ratio and correlation based feature selection. *Int J Inf Technol Knowl Manag* 2(2):271–277
40. Al Janabi KB, Kadhim R (2018) Data reduction techniques: a comparative study for attribute selection methods. *Int J Adv Comput Sci Technol* 8(1):1–13
41. Shirvan RA, Tahami E (2011) Voice analysis for detecting Parkinson's disease using genetic algorithm and KNN classification method. In: 2011 18th Iranian Conference of Biomedical Engineering (ICBME), IEEE, pp 278–283
42. Raschka S (2018) STAT 479: machine learning lecture notes. https://github.com/rasbt/stat479-machine-learning-fs18/blob/master/02_knn/02_knn_notes.pdf Accessed 21 March 2021
43. Ma Y, Guo L, Cukic B (2006) A statistical framework for the prediction of fault-proneness. *Advances in Machine Learning Application in Software Engineering*, Idea Group Inc, pp 237–265
44. Powers DM (2011) Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation
45. Nicolov N (2012) Machine learning with applications in categorization, popularity and sequence labeling: 57th and 58nd slides. http://www.slideshare.net/Nicolas_Nicolov/machine-learning-14528792. Accessed 10 April 2016
46. Yücelbaş Ş, Yücelbaş C, Tezel G, Özşen S, Yosunkaya Ş (2018) Automatic sleep staging based on SVD, VMD, HHT and morphological features of single-lead ECG signal. *Expert Syst Appl* 102:193–206
47. Landis JR, Koch GG (1977) An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 33:363–374
48. Saito T, Rehmsmeier M (2015) The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* 10(3):e0118432
49. Wroge TJ, Ozkanca Y, Demiroglu C, Si D, Atkins DC, Ghomi RH (2018) Parkinson's disease diagnosis using machine learning and voice. In: 2018 IEEE signal processing in medicine and biology symposium, at Philadelphia
50. Solana-Lavalle G, Galán-Hernández J-C, Rosas-Romero R (2020) Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features. *Biocybern Biomed Eng* 40(1):505–516
51. Gunduz H (2019) Deep learning-based Parkinson's disease classification using vocal feature sets. *IEEE Access* 7:115540–115551
52. Tuncer T, Dogan S, Acharya UR (2020) Automated detection of Parkinson's disease using minimum average maximum tree and singular value decomposition method with vowels. *Biocybern Biomed Eng* 40(1):211–220
53. Yücelbaş Ş (2020) Simple logistic hybrid system based on greedy stepwise algorithm for feature analysis to diagnose Parkinson's disease according to gender. *Arab J Sci Eng* 45(3):2001–2016
54. Gürüler H (2017) 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
55. 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):1015–1022
56. Peker M, Sen B, Delen D (2015) Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm. *J Healthc Eng* 6(3):281–302
57. Miller IN, Cronin-Golomb A (2010) Gender differences in Parkinson's disease: clinical characteristics and cognition. *Mov Disord* 25(16):2695–2703
58. Dluzen D, McDermott J (2000) Gender differences in neurotoxicity of the nigrostriatal dopaminergic system: implications for Parkinson's disease. *JGSM* 3(6):36–42
59. Van Den Eeden SK, Tanner CM, Bernstein AL, Fross RD, Leim-peter A, Bloch DA, Nelson LM (2003) Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. *Am J Epidemiol* 157(11):1015–1022
60. Haaxma CA, Bloem BR, Borm GF, Oyen WJ, Leenders KL, Eshuis S, Booij J, Dluzen DE, Horstink MW (2007) Gender differences in Parkinson's disease. *J Neurol Neurosurg Psychiatry* 78(8):819–824

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