GAZİOSMANPAŞA BİLİMSEL ARAŞTIRMA DERGİSİ (GBAD)
Gaziosmanpasa Journal of Scientific Research
ISSN: 2146-8168
http://dergipark.gov.tr/gbad
Araştırma Makalesi (Research Article)

Cilt/Volume : 13
Sayı/Number: 2
Yıl/Year: 2024
Sayfa/Pages: 101-113

Alınış tarihi (Received): 13.11.2024
Kabul tarihi (Accepted): 18.11.2024

Detection of Parkinson Disease Using Frequency Sub-bands of Vocal Cords Vibration Signals and Machine Learning Techniques

Büşra Zeynep GÜREL [1], Kübra TANCI [1,*], Mahmut HEKİM [1], Cem EMEKSİZ [1]

[1] *Tokat Gaziosmanpasa University, Electrical and Electronics Engineering, Tokat, Turkey*
**Corresponding author: kubra.tanci@gop.edu.tr*

ABSTRACT: In this study, we proposed a new approach for diagnosing Parkinson's disease (PD) based on the slope values between neighboring amplitudes of vocal cord vibration signals. The inter-amplitude slope signals were obtained by computing the slopes between adjacent amplitudes in the vocal cord vibration signals. Feature vectors were extracted using common statistical parameters and applied to widely used machine learning classifiers such as Naive Bayes (NB), Generalized Logistic Regression (GLR), Logistic Regression (LR), Decision Tree (DT), and Random Forest (RFs). Different experiments were conducted to evaluate the contribution of the inter-amplitude slope approach and the performance of the classifiers in distinguishing healthy and PD segments. The experiments were carried out on original signals, inter-amplitude slope signals, and sub-band decompositions of both original and slope signals. The results showed satisfactory classification accuracy for all feature extraction methods, with the highest accuracy achieved using inter-amplitude slope signals. The GLR and Random Forest (RFs)-based classifiers outperformed others, achieving 100% accuracy, while the LR classifier reached 91%, and the DT and NB classifiers achieved 95%. Finally, the inter-amplitude slope approach, used for the first time in this study, enhanced classifier performance in PD diagnosis.

Keywords- *Vocal cords, Parkinson's Disease, inter-features slope.*

1. Introduction

Parkinson's disease (PD) first introduced by Doctor James Parkinson in 1817 is a neurodegenerative disorder with major symptoms such as speech disorders, gait disturbances, dementia and tremors occurring due to damaging of neurons that are responsible for the emitting of dopamine in the brain. (Langston, 2002; Jankovic, 2007). It is estimated that PD affects 5 million people worldwide and this disease will unfortunately increase significantly in the next 10 years (Yigit et al., 2018). PD occurs in one person in every 100 people aged 65 and over, which is the most common disease after Alzheimer's disease in a particular age group (Parkinson's Disease Association, 2011). Although some of the symptoms of PD may be reduced by means of drug treatment and/or surgical intervention, there is still no definitive diagnostic method. However, the scales, which was introduced by International Parkinson and Movement Disorder Society (MDS), such as the PD Questionnaire 39 (PDQ-39) (Peto et al., 1995), the Unified PD Rating Scale (MDS-UPDRS) (International Parkinson and Movement Disorder Society, 2008), Hoehn and Yahr Scale (Hoehn et al., 1967), and Brief Psychiatric Rating Scale (Overall et al., 1962), are helpful for the analysis of disease, but they are not seen enough. Therefore, in order to make a significant contribution to early diagnosis of the disease and support the results of the scale, many diagnosis models based on artificial neural networks (ANNs) have recently been suggested by researchers (Guruler and Huseyin, 2017).

The feature vectors extracted from different datasets such as vocal cords signals, speech signals, pre-processed walking signals and Fourier-transformed infrared micro-spectroscopy (FTIR) datasets have been used as the inputs of the ANN-based classifier models in the diagnosis of PD. These include studies on the use of artificial intelligence and machine learning techniques in the diagnosis of PD (Kwon, et al., 2020), studies examining the relationship between audio signals and audio analysis and Parkinson's disease (Muro et al., 2018), studies on PD diagnosis with pre-processed gait signals (Mirelman et al., 2011), and the effects of data sets obtained using FTIR techniques on Parkinson's disease (Kavanagh, et al., 2015). FTIR is a powerful technique in biochemistry to examine the secondary structure of protein molecules, nucleic acids, carbohydrates and lipids (Hamilton, 2016). This study of Hamilton in 2016 allowed the analysis of PD progress on the semi-numerical evaluation of DaTSCAN images by using the ratio of audience accumulation in putamen to the audience accumulation in the caudate core. In a study of Baby in 2017, the pre-processed walking dataset was also used for the classification of PD. Muniz et al. used the basic components derived from the vertical component of ground reaction force (vGRF) as the inputs of a probabilistic neural network (PNN) in order to determine the difference between normal and PD subjects (Muniz et al., 2009). Ai et al. (2008) proposed a novel method to determine three types of common tremor (ET), PD tremor and physiological tremor (PT) which were frequently incorrectly diagnosed clinically. To this end, they proposed a novel synthetic multifaceted diagnostic method for hand tremendous acceleration signals by combining the output results of the independent back-propagated neural network (BPNN) with the D-S evidence theory. In another study, the dataset consisting of PD and healthy subjects were weighted the by using the fuzzy C-means (FCM) clustering-based feature weighting (FCMFW) model, and then classified them by using the k-NN classifier (Polat, 2011). While *k* was selected as 2, while total correct classification (TCC) ratio was 72.16% for the raw dataset, it was 96.96% for the dataset weighted by using FCMFW approach. In the study of Ozcift in 2012, the integration of feature selection and rotation forest classifiers in computer-aided diagnosis of (PD) is investigated. In the study, support vector machines (SVM) are used to select the most important features and rotation forest classifiers are used on these selected features to improve the diagnostic performance. PD was detected by using SVM classifier in TCC ratio of 97% (Ozcift, 2012). Chen et al. proposed a PD diagnostic system with the fuzzy k-Nearest Neighbor (FKNN) method, and compared its experimental results with them of SVM-based approaches (Chen et al., 2013). They also used the Principal Component Analysis (PCA) approach to reduce the size of the dataset in the analysis stage to improve PD diagnostic accuracy. Their experimental results showed that the FKNN-based system significantly improved the classification accuracy with 96.07% and left behind SVM-based approaches and other methods in the literature. In 2014, Chandrashekar developed a new method for the diagnosis of PD by using an ANN-based classifier and speech measurements. For this method, PDs were spoken into a microphone, and these speech signals were analyzed by using the Praat (a free software package for the scientific analysis of phonetics). ANN-based classifier trained by the actual PD dataset could accurately classified the dataset in the TCC ratio of 96.55%. In another study, the feature vectors extracted by using two computational models that match speech signal measurements to clinical results were applied into MLPNN and Adaptive Neuro Fuzzy Inference System (ANFIS) models for the diagnosis of PD (Hlavica et al., 2016). It was seen in this study that the MLPNN model trained by using the Levenberg-Marquardt (LM) optimization algorithm was more successful than the MLPNN model trained by using flexible back propagation and conjugate gradient methods. The artificial neural network model turned out to be 92.6% accurate and the ANFIS model was 90.5% accurate. The findings show that artificial neural networks are better at predicting progression of PD. The models' loss functions were also analyzed and the artificial neural network model's loss value was less than that of the ANFIS model. The latter was more sensitive as the power plant operational conditions affected the models. It shows that the artificial neural network has a greater predictive power. Khan et al. suggested an evolutionary method to train both distinguishable and indistinguishable parameters using the Cartesian Genetic Programming (CGP) (Khan et al., 2017). When real-world datasets for the detection of PD were used, the performance of their method reached to remarkable levels obtained by using many standard classification methods. Nilashi et al. focused on a new hybrid intelligent system to be used for estimating PD progression (Nilashi et al., 2016). In their system, PCA and Expectation Maximization (EM) were used to address multiple common problems in experimental datasets and to cluster the dataset. Then, PD progression was estimated by using ANFIS and Support Vector Regression (SVR). The experimental results performed by using the datasets of public Parkinson's showed that their method increased the prediction accuracy of PD progression, significantly. Guruler in 2017 proposed a new diagnostic system using a combination of K-means clustering-based feature weighting (KMCFW) method and complex valued ANN (CVANN) model. In the first stage of the model, the features in the PD dataset were weighted by using the KMCFW method and these new features were converted into a complex format. The obtained complex feature values are applied into the input of CVANN model. The performance of the proposed system was evaluated by five different evaluation methods. Their experimental results showed that this hybrid system called as KMCFW-CVANN was significantly superior to the real valued ANN and reached the highest classification results reported so far with a classification accuracy of 99.52%. In 2017, Yilancioglu focused on PD diagnosis based on the linear regression (LR) model and used the features which are motor, total United PD Rating Scale (UPDRS) clinical results and vocal cord signals. In addition to this, some features were removed from the training datasets used in the LR model in order to show the importance of different characteristics and correlations between the results and predicted UPDRS scores. The total UPDRS features were excluded from the training dataset and the experiments were repeated. The correlation between the estimated motor UPDRS and clinical motor UPDRS decreased to 72%. When the motor UPDRS features were removed from the training dataset, the total estimation of UPDRS decreased by up to 76%. Next, all Jitter values were excluded from the training dataset, and then the motor UPDRS scores were estimated. The correlation between predicted motor UPDRS score and clinical motor UPDRS score was found as 97%. Jitter values and total UPDRS scores were then removed and motor UPDRS scores were estimated. The correlation between the estimated and clinical motor UPDRS scores was 79%.

In this paper, a new approach is proposed for the diagnosis of PD based on the feature vectors obtained by using the basic statistical parameters on the frequency sub-bands of vocal cord signals. First, the vocal cord signals were decomposed into the sub-bands by using the discrete wavelet transform (DWT). Second, the feature vectors were computed by using the basic statistical parameters that are mean, standard variation, geometric mean and harmonic mean. The obtained feature vectors were used as the inputs of NB, GLR, LR, DT and RFs based classification models with and without normalization in order to illustrate the contributions of the basic parameters to PD learning of the classifier models.

2. Material and Methods

2.1. Dataset

In this study, the dataset was consisted of 54 healthy subjects and 54 PD subjects. 54 healthy subjects were collected by means of the sound recordings from healthy volunteers aged 16-50 years. For this aim, they were asked to say *a*, *o* and *u* for 5 seconds, and the obtained sounds were recorded in *.m4a* format. The sound samples of 20 PD subjects were taken from University of California repository, and they were rearranged by dividing with 5 seconds. Both of healthy and PD sounds were sampled at 48000 Hz (Parkinsons Data Set). The examples of signals taken from the PD and healthy segments are given in Figure 1.

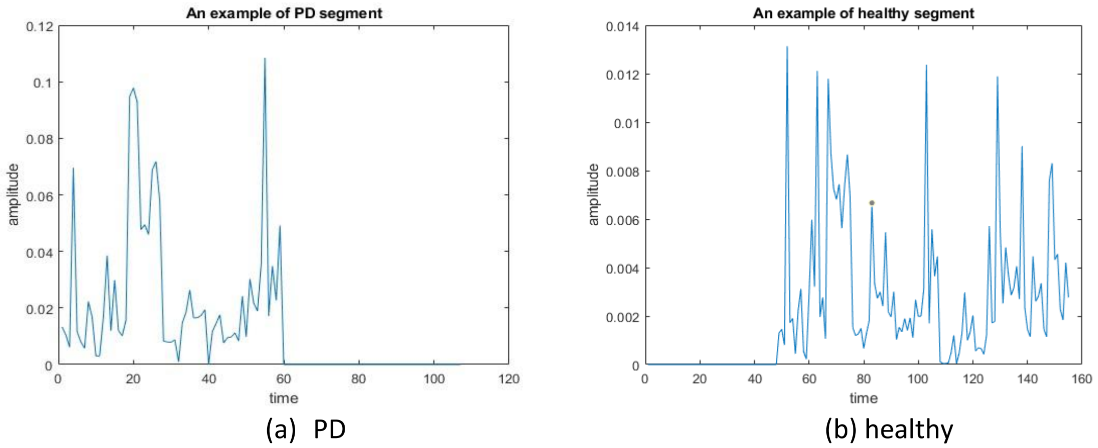

Figure 1. The examples of signals taken from the PD and healthy segments

2.2. Discrete Wavelet Transform (DWT)

The Discrete Wavelet Transform (DWT) is a spectral analysis technique used for analyzing non-stationary signals, providing a time-frequency representation. It has been widely applied in the analysis of biomedical signals and images, as these signals often exhibit non-stationary characteristics (Hariharan et al., 2022). DWT employs long time windows at low frequencies and short time windows at high frequencies, which ensures effective time-frequency localization (Gonzalez and Woods, 2002). This approach allows for the best time-frequency resolution across all frequency ranges. DWT is also capable of analyzing time series signals with non-stationary power across different frequencies, using both the time-frequency and time-scale domains (Coskun and Istanbullu, 2012).

DWT decomposes a signal into sub-bands through successive high-pass and low-pass filtering of the time-domain signal. The high-pass filter (g) is the discrete mother wavelet, while the low-pass filter (h) is its mirror version. The down-sampled signals from the first filtering step are called the first-level approximation (A1) and detail coefficients (D1). The approximation and detail coefficients for subsequent levels are obtained using the approximation coefficients from the previous level. The number of decomposition levels is determined based on the dominant frequency components of the signal (Zayrit et al., 2020).

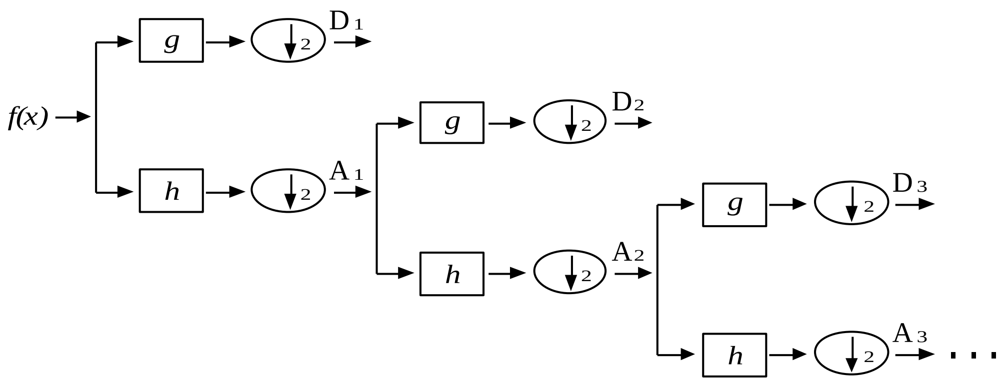

Figure 2. Sub-band decomposition of a signal by using DWT

Scaling function, $\varphi_{j,k}(x)$ based on low pass filter and wavelet function, $\psi_{j,k}(x)$ based on high pass filter are defined as

$$\varphi_{j,k}(x) = 2^{j/2} h(2^j x - k) \tag{1}$$

$$\psi_{j,k}(x) = 2^{j/2} g(2^j x - k) \tag{2}$$

where x=0,1, 2,…, M-1, j=0,1,2,…,J-1, k=0,1,2,…,2^j-1, J equals to $log_2(M)$ and M is the length of the signal and chosen as 2^J (Gonzalez and Woods, 2002).

Sampling rate k and the resolution j specify the positions and the widths on the x axis of functions, respectively. The amplitudes of functions depend on $2^{j/2}$ value (Gonzalez and Woods, 2002). Approximation coefficients $A_i(k)$ and detail coefficients $D_i(k)$ in ith level are described as

$$A_i = \left\{ \frac{1}{\sqrt{M}} \sum_x f(x).\varphi_{j,k}(x) \right\} \text{ and } D_i = \left\{ \frac{1}{\sqrt{M}} \sum_x f(x).\psi_{j,k}(x) \right\} \text{ for } k=0,1,2,\ldots,\ 2^j - 1 \tag{3}$$

Figure 3 and 4 show approximate and detailed coefficients of vocal cord segments taken from the healthy and PD subjects, respectively.

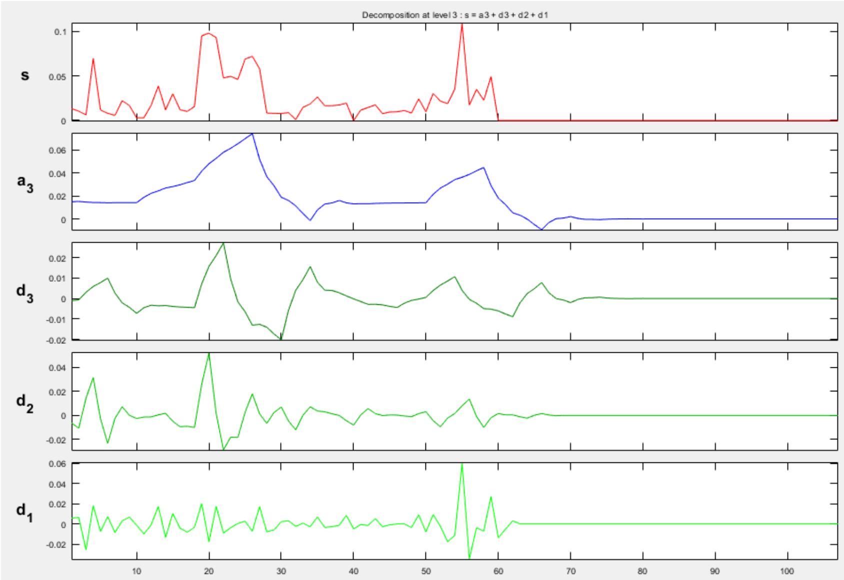

Figure 3. Approximate and detailed coefficients of a sample vocal cord segment taken from healthy subject

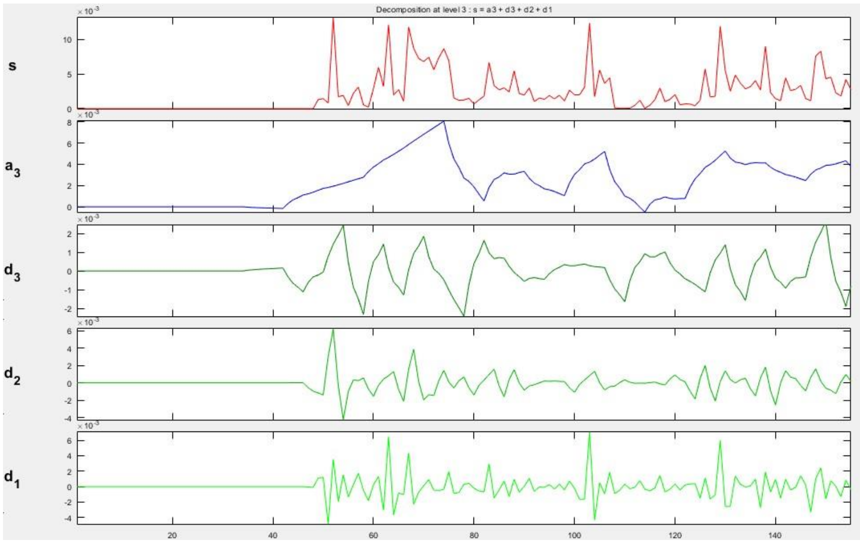

Figure 4. Approximate and detailed coefficients of a sample vocal cord segment taken from the PD subject

In this study, the decomposition level of the DWT applied to both normalized and non-normalized vocal cord signal was determined by the correct classification success rates. Since level 3 provided the highest TCC rates in the experimental studies, this level of transform was used. Therefore, in this study, the feature vectors calculated from the A3, D1, D2, D3 coefficients obtained by wavelet transform at level 3 were applied as input to machine learning (ML)-based classifier models for PD diagnosis.

2.3. Feature extraction

For comparing the signals, they need to be normalized in power level by scaling in identical level, therefore we normalized both of PD and healthy vocal cord segments between [0,1] using min-max normalization formula as below

$$x_i = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4}$$

The obtained normalized signals were decomposed into frequency sub-bands using DWT, and the feature vectors were obtained from the sub-band coefficients using basic statistical parameters in following Table 1.

Table 1. The used statistical parameters

Statistical parameters	***Formulas***	***Explain***
Mean	$X_{mean} = \frac{\sum_{i=1}^{n} x_i}{n}$	The mean is the ratio of the sum of the amplitudes in a segment to the number of data points.
Standard variation	$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - X_{mean})^2}{n}}$	The standard deviation is the ratio of the sum of the squares of the distance to the mean of the amplitudes in a segment to the number of data points.
Geometric mean	$G = \sqrt[n]{\prod_{i=1}^{n} x_i}$	Geometric mean is the multiplication of the amplitudes in a segment by the root of the number of data points in that segment.
Harmonic mean	$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$	The harmonic average is multiplication of the inverse of the amplitudes in a segment by the number of data points in that segment.

where, x_i is the amplitude of i_{th} data point, n is the number of data points.

2.4. Detection of PD using the sub-bands of vocal cords

In this study, we used the widely known classifiers for the detection of PD which are Naïve Bayes (NB), General Logistic Regression (GLR), Logistic Regression (LR), Decision Tree (DT) and Random Forest (RFs). These classifiers are briefly described in the following subsections:

2.4.1. NB classifier

In this approach, data with a certain proportion of pre-classified data is presented and learned. The new data is compared with the taught data to predict which category the new data belongs to (Mitchel, T. ,1997). The more pre-classified data, the closer the prediction is to the truth. What matters in NB is the probability of an event happening (Bishop, 2006).

$$PP\left(\frac{Y_i}{X}\right) = \frac{P\left(\frac{X}{Y_i}\right)P(Y_i)}{P(X)} \tag{5}$$

where, Y_i is one of the classes previously labeled as 1 or -1. NB is eager learning, meaning that it learns before the data to be predicted arrives, calculating all the probabilities. After learning, it is memory efficient as it deletes the data used for training and keeps only the probabilities (Bhattacharyya et al., 2011).

2.4.2. Logistic Regression (LR) classifier

In logistic regression, the estimated values of the dependent variable are calculated probabilistically and classified accordingly (Ayhan 2006). Binary logistic regression method was used in the experimental studies. Binary logistic regression model is expressed by Equation 6 (Ozdamar, 2009).

$$P(Y)=\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0}+e^{\beta_1 x}} \quad (6)$$

2.4.3. DT classifier

Classification based on this approach is predicated on learning the dataset by partitioning it into certain groups. The decision tree starts with a node, which is the most discriminative category in the dataset. The most discriminative category is the one with the highest gain. Equation 7 calculates the entropy of the class to be predicted (Ghanbari, et al., 2023).

$$E(D) = -\sum_{i=1}^{N} P_i log_2(P_i) \quad (7)$$

where, P_i is the class to be predicted. After calculating the information gain of attribute, A in dataset D with Equation 8, the gain is calculated with Equation 9 and the node with the highest gain is selected (Balaji, et al., 2023).

$$info_A(D) = \sum_{j=1}^{N} (\frac{|D_j|}{|D|} E(D_j)) \quad (8)$$

$$Gain(A)=E(D)-info_A(D) \quad (9)$$

where, $Gain(A)$ is the gain of attribute A, D_j is the jth element in the dataset. N is the length of the dataset (Ali et al., 2023).

2.4.4. Random Forests (RFs) classifier

A random forest (RF) multi-class classifier consists of multiple trees, each grown with some form of randomness. The leaf nodes of each tree are labeled with estimates of the posterior distribution over the image classes. Each internal node contains a test that optimally splits the data space to be classified. A pattern is classified by passing it down each tree and aggregating the leaf distributions that it reaches. Randomness can be introduced at two stages during training: by subsampling the training data so that each tree is grown with a different subset, and by selecting the node tests randomly (Bosch et al., 2007). The RF classifier approach aims to achieve higher accuracy by using multiple decision trees simultaneously. The key difference from bagging is how the decision trees are constructed. The feature to split at each node is chosen as the best from a randomly selected set of features.

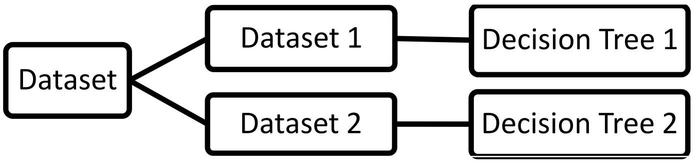

Figure 5. The schematic structure of RFs

As it can be seen in Figure 5, the majority vote can be taken as the decision (averaged if a prediction is to be made). In RF, unlike NB, as the data increases, the results may be wrong. This is due to overlearning. Also, the computation time gets longer as the branches increase. Therefore, it is useful for small and medium sized data sets.

3. Results and Discussion

In this paper, the data normalization and feature extraction processes, and the classification processes based on machine learning algorithms was implemented in MATLAB and Rapid Miner program, respectively. In first, dataset was normalized by using Equation 4 in order to have the same range of values for each of the inputs to the classifier models. This can guarantee stable convergence of weight and biases. In second, the normalized vocal cords were decomposed by using DWT into sub-bands, and the feature vectors were extracted from each sub-band of vocal cords by using widely used statistical parameters which are mean, geometric mean, harmonic mean, variance, inter-point slope, Chi-Square and its entropy. In order to detect PD, the feature vectors were applied into the classifier models which are NB, GLR, LR, DT, RFs.

Data preprocessing algorithms were written in MATLAB. Machine learning methods were tested in Rapid Miner. According to the tests performed in Rapid-Miner, the results (discarding those with stability=100%) are given in the figures below. According to the calculations made in Rapid Miner, the stability of the geometric and harmonic averages of the virtual parts of the data and the averages of the coefficients A of the 2nd wavelet transform and D of the 3rd wavelet transform of the real parts were found to be 100%. They were removed from the dataset as they may lead to overlearning (memorization). The correlations of the standard deviations of the D coefficients of the real and imaginary parts of the 1st wavelet transforms and the deviations of the harmonic means of the real parts of the 4th wavelet transform were found to be close to 0%. The results for non-normalized signals are given in the Table 2.

Table 2. Results for non-normalized signals

	ACC (%)	F1 Score (%)	Precision (%)	AUC
NB	73	77	67	0.905
GLR	91	92	85	0.917
LR	95	95	100	0.934
DT	91	92	85	0.909
RFs	100	100	100	1

As can be seen in Table 2, NB shows a lower performance compared to other models. Especially in terms of Precision, it gave a low result with 67%. In addition, although the area under the curve (AUC) value is 0.905, it is weaker than the other models. The GLR model shows a strong performance in terms of accuracy and F1 score. The AUC value is also quite high at 0.917, indicating that the model can successfully separate the positive class. Logistic Regression performs the best in terms of Precision with 100%. The AUC value is also quite high at 0.934, indicating that the model is a strong classifier. Overall, it can be considered as a model that can produce accurate and balanced predictions. The Decision Tree performs similarly to GLR. Although it has high accuracy, F1 score and AUC values, it falls below other strong models. The Random Forests model performs 100% on all metrics and has an excellent AUC of 1. It is possible that this model overfits the data, but it appears to be the strongest model on the available data. The ROC curves for these classifiers are given in Figure 6.

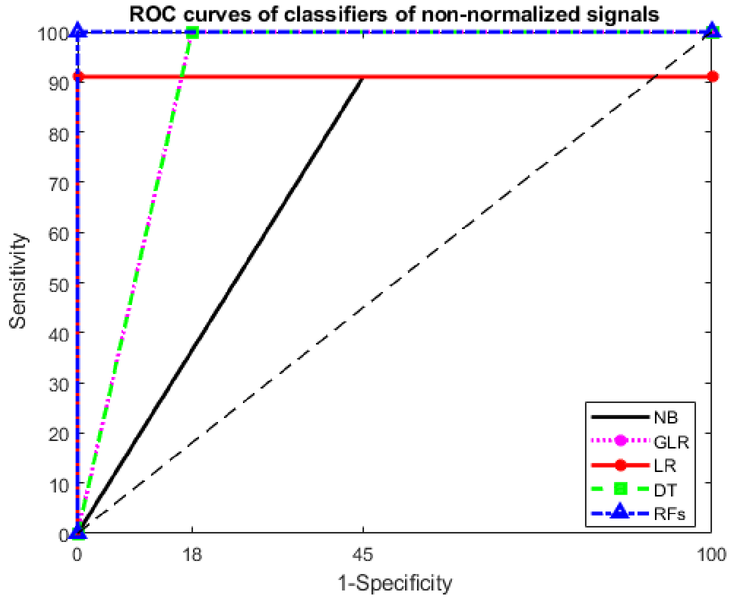

Figure 6. ROC curves for non-normalized signals

As can be seen in Figure 6, the RFs model shows the highest performance in all metrics, but may pose a risk of overfitting. Logistic Regression (LR) and GLR stand out with high accuracy and AUC values and can be considered as reliable classifiers.

The dataset was normalized in MATLAB and then fed back to Rapid Miner Auto Model for machine learning methods. In the dataset, the collinearities of the averages, the means and harmonic means of the 1st Wavelet A coefficients, the real coefficients and the standard deviations of both the A and D coefficients of the 1st, 2nd, 3rd, 4th Wavelets are greater than 95%. Also, the stability of the geometric and harmonic means of the virtual parts is greater than 90%. Since these have a bad effect on learning, they were removed from the dataset and a new dataset was given to the input of the machine learning models. Accordingly, the results are given in Table 3.

Table 3. Results for non-normalized signals

	ACC	F1 Score	Precision	AUC
NB	95	95	100	0.917
GLR	100	100	100	1
LR	91	92	85	1
DT	95	96	92	0.5
RFs	100	100	100	1

As seen in Table 3, the NB model performs quite well, especially with a Precision of 100%, minimizing the false positive rate. The AUC value of 0.917 shows that the model separates the positive and negative classes well, but is slightly weaker than the other models. The GLR model turns out to be an excellent classifier, performing 100% on all metrics. The AUC value is also 1, indicating that the model successfully distinguishes classes. However, the 100% performance may carry a risk of overfitting the model. Although the Logistic Regression model performs well in terms of accuracy and F1 score, it may tend to misclassify some positive examples with a Precision of 85%. The AUC value is quite high at 1, indicating that it can distinguish between positive and negative classes very well. While the accuracy and F1 score of the Decision Tree model is high, the AUC value is given as 0.5. This indicates that the model cannot distinguish between positive and negative classes and makes random predictions. Such a low AUC may suggest that the model's classification success is not independent of the data distribution and only performs well in certain parts of the data. The Random Forests model performs 100% on all metrics. An AUC of 1 indicates that the model can perfectly separate classes. The ROC curves for these classifiers are given in Figure 7.

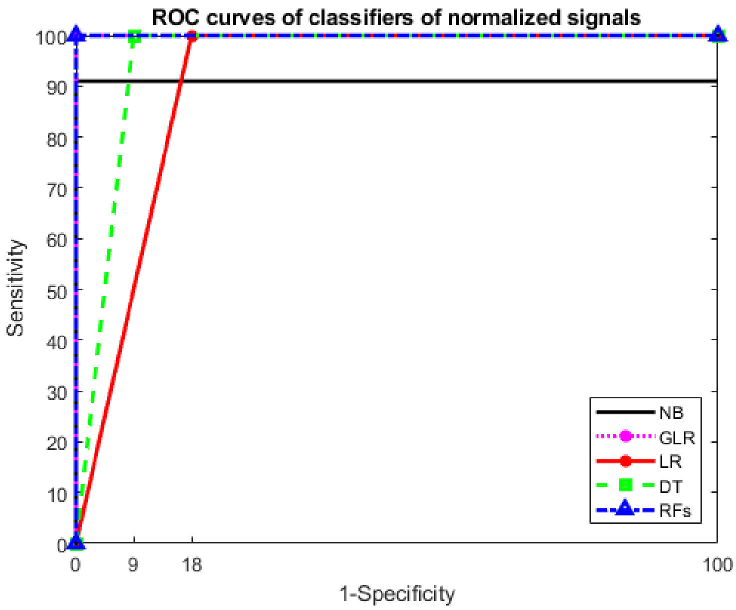

Figure 7. ROC curves for normalized signals

As shown in Figure 7, the GLR and RFs models show the highest performance, but this can carry a risk of overfitting. The NB model also provides very reliable results, while LR shows a balanced performance. The low AUC value for the DT model indicates that this model has difficulty distinguishing between classes and is not as reliable as the other models.

Normalization provided a significant improvement in classifier performance. With non-normalized data, the RF classifier showed the highest accuracy and F1 score, while there was a significant difference between the other classifiers in terms of accuracy and F1 score. However, GLR and DT classifiers showed limited success in some performance metrics. In particular, the AUC value for the NB model remained at 0.905, indicating that the classification capacity of the model is relatively low.

After normalization, the performance of all classifiers improved significantly. For example, the NB classifier reached 95% in ACC and F1 score, and its AUC value increased to 0.917. The GLR and RF classifiers achieved 100% accuracy, 100% F1 score and 100% AUC, providing an optimal classification. Although the AUC of the DT classifier decreased to 50%, it performed well in other metrics (ACC: 95%, F1: 96%). This shows that normalization positively affects the feature extraction process of certain algorithms, but may have different effects on algorithms such as DT. These analyses show that normalization plays a critical role in improving classification accuracy, especially in GLR, LR and RF classifiers.

4. Conclusions

In this study, we investigate learning methods and parameters to help early detection of PD. The studies show that normalization positively affects the feature extraction process of certain algorithms, but may have different effects on algorithms such as DT. These analyses show that normalization plays a critical role in improving classification accuracy, especially in GLR, LR and RF classifiers. In conclusion, these findings suggest that normalization is an important preprocessing step, especially for classifiers such as NB and GLR, and plays a critical role in improving performance. Moreover, the RF classifier maintains its high performance regardless of the normalization status and can be an effective choice for early detection of PD. Future work could explore more classifiers and datasets to generalize these findings and examine normalization techniques' effects on performance. Studies with clinical data may also enhance practical applicability, aiding the development of robust tools for early PD diagnosis.

5. References

Ai, L., Wang, J. ve Wang, X., 2008. Multi-features fusion diagnosis of tremor based on artificial neural network and d–s evidence theory. Signal Processing, 88(12), 2927-2935.

Ali, A. M., Salim, F., Saeed, F., 2023. A Comprehensive Review on Parkinson's Disease Detection Using Machine Learning Techniques. IEEE Access, 11, 78945-78968. https://doi.org/10.1109/ACCESS.2023.3255672.

Anonim, Parkinsons Data Set, https://archive.ics.uci.edu/ml/datasets/parkinsons (21.01.2018).

Ayhan, S. (2006). Determining the Factors Affecting Nurses' Intention to Quit in Turkey Using Ordered Logistic Regression Analysis. Master's Thesis, Institute of Science, Eskişehir Osmangazi University.

Baby, M.S., Saji, A.J., Kumar, C.S., 2017. Parkinsons disease classification using wavelet transform based feature extraction of gait data. International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam India.

Balaji, R., Alam, M. N., et al., 2023. An ensemble classifier for assessing Parkinson's disease severity using decision trees and gait analysis. Journal of Biomedical Informatics. https://doi.org/10.1016/j.jbi.2023.103647.

Bhattacharyya, S., Gupta, D., & Pradhan, R. (2011). Performance analysis of left/right hand movement classification from EEG signal by intelligent algorithms. In 2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB) (pp. 1-6). IEEE. https://doi.org/10.1109/CCMB.2011.5952133

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

Bosch, A., Zisserman, A., & Munoz, X. (2007). Image classification using random forests and ferns. In 2007 IEEE 11th International Conference on Computer Vision (pp. 1-8). IEEE. https://doi.org/10.1109/ICCV.2007.4409066.

Chandrashekar, P., 2014. Mathematically modeling the GPE/STN neuronal cluster to account for Parkinsonian tremor and developing a novel method to accurately diagnose Parkinson's disease using speech measurements and an artificial neural network. Quantitative Biology, Neurons and Cognitio, https://arxiv.org/abs/1405.1314

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 Systems with Applications, 40(1), 263-271.

Coşkun, M., & Ayhan Istanbullu. (2012). Analysis of EEG Signals with FFT and Wavelet Transform. XIV. Academic Computing Conference, Usak, Turkey.

Ghanbari, A., Hossain, M. S., Saba, T., Yaqoob, I., 2023. Information Set-Based Decision Tree for Parkinson's Disease Severity Assessment Using Multidimensional Gait Dataset. IEEE Access, 11, 4010-4022. https://doi.org/10.1109/ACCESS.2023.3238656.

Gonzalez, R. C., & Woods, R. E. (2002). Digital image processing (2nd ed.). Prentice Hall.

Guruler, H., 2017. A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. Neural Computing & Applications, 28(7), 1657.

Hamilton, D., List, A., Butler, T., Hogg, S., Cawley, M., 2016. Discrimination between parkinsonian syndrome and essential tremor using artificial neural network classification of quantified DaTSCAN data. Nuclear Medicine Communications, 27(12), 939-944.

Hariharan, M., Balasubramanian, A., Kumar, P. G. P., 2022. Early Detection of Parkinson's Disease Using Fusion of Discrete Wavelet Transformation and Histograms of Oriented Gradients. Mathematics, 10(4218).

Hlavica J., Prauzek M., Peterek, T., Musilek, P., 2016. Assessment of Parkinson's disease progression using neural network and ANFIS models. Neural Network World, 26(2), 111-128.

Hoehn, M. M., Yahr, M. D., 1967. Parkinsonism: onset, progression, and mortality. Neurology, 17(5), 427-442.

Jankovic, J., 2007. Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry, 79(4), 368–376.

Kavanagh, E. R., Diamond, D., 2015. Fourier transform infrared spectroscopy for the diagnosis of Parkinson's disease. Analytical Chemistry, 87(14), 7081-7088. https://doi.org/10.1021/acs.analchem.5b01434.

Khan, M.M., Mendes, A., Zhang, P., Chalup, S., 2017. Evolving multi-dimensional wavelet neural networks for classification using Cartesian Genetic Programming. Neurocomputing, 247, 39-58.

Kwon, H., Kim, S., 2020. Machine learning techniques for diagnosis of Parkinson's disease: A review. Artificial Intelligence in Medicine, 105, 101853. https://doi.org/10.1016/j.artmed.2020.101853.

Langston, J.W., 2002. Parkinson's disease: current and future challenges. Neurotoxicology, 23(4–5), 443–450.

Mirelman, A., Maidan, I., Herman, T., 2011. Gait analysis in Parkinson's disease: A review. Journal of Neurology, 258(1), 90-99. https://doi.org/10.1007/s00415-010-5658-2.

Mitchel, T. "Machine Learning". McGraw-Hill Science, 1997.

Muniz, A., Liu, W., Liu, H., Lyons, K.E., Pahwa, R., Nobre, F.F., Nadal, J., 2009. Assessment of the effects of subthalamic stimulation in Parkinson disease patients by artificial neural network. Annual International Conference of the IEEE Engineering in Medicine and Biology Society Engineering in Medicine and Biology Society, 5673-5676.

Muro, A., Orlik, P., 2018. Vocal characteristics in Parkinson's disease: A review of the literature. Journal of Voice, 32(6), 748-755. https://doi.org/10.1016/j.jvoice.2017.09.001.

Nilashi, M., Ibrahim, O., Ahani, A., 2016. Accuracy Improvement for Predicting Parkinson's Disease Progression.Scientific Reports, 6(34181). Doi:10.1038/Srep34181.

Overall, J. E., Gorham, D. R., 1962. The brief psychiatric rating scale. Psychological Reports, 10(3), 799-812.

Ozcift, A., 2012. SVM Feature Selection Based Rotation Forest Ensemble Classifiers to Improve Computer-Aided Diagnosis of Parkinson Disease. Journal of Medical Systems, 36(4), 2141-2147.

Özdamar, K. (2009). Statistical Data Analysis with Software. Kaan Publishing, Eskisehir.

Peto, V., Jenkinson, C., Fitzpatrick, R., Greenhall, R., 1995. The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. Quality of Life Research, 4(3), 241-248.

Polat, K. (2011). Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering. *International Journal of Systems Science*, *43*(4), 597–609. https://doi.org/10.1080/00207721.2011.581395.

S. Haykin and E. Frank, Neural Networks, 2nd ed. New Jersey: Prentice Hall, 1999.
https://archive.ics.uci.edu/ml/datasets/parkinsons+telemonitoring (21.01.2018).

Yılancıoğlu K. 2017. Vocal Cord Measure Based Ann Approach for Prediction of Parkınson's Diseases Status. SDU Journal of Health Science Institute / SDÜ Sağlık Bilimleri Enstitüsü Dergisi, 1, 139-142.

Yiğit, E.N.; Sönmez, E.; Söğüt, M.S.; Çakır, T.; Kurnaz, I.A. Validation of an In-Vitro Parkinson's Disease Model for the Study of Neuroprotection. *Proceedings* 2018, *2*, 1559. https://doi.org/10.3390/proceedings2251559.

Zayrit, S., Drissi, T. B., Ammoumou, A., Nsiri, B., 2020. Daubechies Wavelet Cepstral Coefficients for Parkinson's Disease Detection. Complex Systems, 29(3), 729-734. DOI: 10.25088/COMPLEXSYSTEMS.29.3.729.