Parkinson's Disease Detection from Voice and Speech Data Using Machine Learning

Chapter · May 2021

DOI: 10.1007/978-981-16-0586-4_36

2 authors:

Anik Pramanik
New Jersey Institute of Technology

Amlan Sarker
Daffodil International University

Parkinson's Disease Detection from Voice and Speech Data Using Machine Learning

Anik Pramanik* and Amlan Sarker

Khulna University of Engineering & Technology, Khulna-9203, Bangladesh
{anikpramanikcse,sarkeramlan}@gmail.com

Abstract. Parkinson's Disease (PD) is a growing neurological disease that causes dysfunction in the nervous system. But, there is no specific diagnosis for PD and it can be only be detected through various motor features. More than 90% of the PD patients reported to have exhibited voice impairment. In this paper, we have proposed a model for PD detection using voice and speech signal data. The PD Speech data-set used in this experiment exhibits huge dimensionality with comparatively less data-points. Our proposed model introduced different data preprocessing methods, such as data standardization, multicolinearity diagnosis, dimensionality reduction technique to improve the quality of data. Different Machine Learning (ML) classifiers (k-nearest Neighbour, Support Vector Machine, Random Forest, AdaBoost, Logistic Regression) were used for classification of PD. Hyper-parameter tuning, cross fold validation and grid search were employed in this experiment to maximize the performance of the classifiers and preserve the class distribution of the imbalanced data-set. Our proposed model achieved a highest accuracy of 94.10% which outperformed the previous experiments on the same data-set by roughly 8.00%.

Keywords: Parkinson's Disease detection · Machine Learning · Data preprocessing.

1 Introduction

Parkinson's disease (PD) is a progressive neurological disorder that exhibits several motor and non-motor dysfunctions [11]. It is one of the most prominent neurological disorder, preceded only by Alzheimer's disease [1]. PD is caused by decay of dopaminergic neurons and Lewy bodies' presence in mid-brain [3]. People with the age of 60 or more are more vulnerable to PD [23]. PD exhibits diverse symptoms including postural instability, tremor, rigidity, dysarthria, hypomimia, dysphagia, shuffling gait, decreased arm swing, micrographia, cognitive impairment and several other motor and non-motor disability [5, 7]. PD is a growing health care problem. Statistical results show that the number of PD patients is expected to double from 6.2 million cases as of 2015 to 12.9 million cases by 2040 [8]. However, detection of PD is a challenging task as there is

* Corresponding author

no specific diagnostic test for a patient with PD. Patients must be diagnosed according to the clinical symptoms and criteria [20].

Among different procedures for diagnosis of Parkinson's Disease, PD detection based on phonation and speech data has been proven to be very effective as 90% of patients with Parkinson's have shown voice impairments [13]. As symptoms of Parkinson's Disease includes dysrathia, dysphagia which causes vocal-muscle disorder, swallowing difficulties, inability in controlling salivary secretion, turbidity and reduced degree of facial manifestation [5, 7]. These type of dysfunction causes speech disorder, variations in voice intensity, pitch range, articulation rate. So, vocal and acoustic analysis of audio samples is used as a non-invasive procedure for diagnosis of PD by researchers [2, 19, 13]. Furthermore, phonation and speech data analysis is practised in early diagnosis of PD by many researchers [1, 4].

In this literature, a model is proposed to detect Parkinson's Disease using voice and speech signal data in an efficient and robust manner. The features in the data-set include fundamental frequencies, harmonicity variants, time frequency attributes, wavelet and vocal based features, and many other speech signal data, all totalling no less than 750 attributes of 252 persons. The higher level of dimensions causes exponential increase in feature space and increases the risk of overfitting the classifier model [24]. Variance Inflation Factor (VIF) was extracted from features to diagnose Multicollinearity in data-set and remove highly correlated attributes [22]. Data was standardized to equalize the magnitude and variability of the attributes [12]. Dimension reduction techniques, such as Principal Principal Component Analysis (PCA), Independent Component Analysis (ICA), were used to reduce the number of features [16, 10]. During classification procedure, k-fold cross-validation was performed to maintain the class distribution proportion as close as original data-set [17]. Different Machine Learning classifiers, such as Support Vector Machine (SVM), Logistic Regression (LR), k-Nearest Neighbour (k-NN), AdaBoost (AdB), Random Forrest (RF) were used in the experiment [9, 14, 6, 15]. Grid search was used for hyper-parameter tuning and optimize the classifiers' performance [21]. This paper contains detailed discussion about how variation of classifiers, their parameters and feature representations affect the performance of the model and thus producing a more accurate, robust and efficient model. The main contributions of this research can be summarized as follows:

1. A model is proposed to detect Parkinson's Disease using voice and speech signal data.
2. Introduction of intensive data pre-processing to locate underlying and repeating patterns in the features.
3. Usage of different dimension reduction techniques to create a more relevant feature-space.
4. Comparison of five different ML based model to select the best model for Parkinson's Disease detection.
5. Achievement of highest accuracy in comparison to other research works on same data-set.

Section 2 contains discussion about similar works in this topic. Section 3 gives description of the data-set. Section 4 demonstrates the proposed framework. Details about the used data processing techniques, the ML classifiers and validation techniques are described here. Section 5 discusses the experiment result. The conclusion is drawn in section 6.

2 Literature Review

During recent years, several papers have been published regarding detection of Parkinson's Disease using speech and voice data. An author proposed a model where Neural Networks, Decision Tree, Regression and DMneural were used for detection of PD and comparative analysis was made [4]. A paper used (LOSO) validation technique on MFCC voice recording samples and SVM classifier to discern between PD patients and healthy people [2]. Another paper describes a two-staged attribute selection and classification model for detection of PD [13]. Ali proposed a early predictive model for PD detection using two-dimensional simultaneous sample and feature selection method [1]. Another author used ensemble learning techniques combining different classifiers for PD detection with accuracy of 86% [19]. There are also some other papers having proposed models with very high accuracy on detection of PD. For instance, a paper combined weighted clustering with Complex Valued Artificial Neural Network (CVANN) in their proposed model which achieved an accuracy of 99.5%. However, despite having high accuracy these experiments show biased results. The reported data-sets used in the experiments had small data points and each subject had multiple voice recording samples [18]. Despite having several papers published on PD detection, improvement is still necessary to provide models with better accuracy, robustness and efficiency.

3 Dataset

The data-set, on which the proposed model was tested, was publicly available as PD Speech data-set from Department of Neurology in Cerrahpaşa, Faculty of Medicine, Istanbul University. The data-set contained data from 188 PD patients (81 women and 107 men). Their ages ranged between 33-87 and their average age was 65.1. The control group of the data-set contains data from 64 healthy subjects (41 women and 23 men) and their average age being 61.1.As shown in Table 1, the data-set contains baseline features, time frequency features, Mel Frequency Co-efficient features and many other speech signal data regarding the subjects. The PD Speech data-set consists of a total 753 speech signal attributes of 252 people.

4 Proposed Framework

In this paper, a model is proposed that efficiently classifies Parkinson's Disease based on PD Speech data-set. The visual illustration of the proposed model is

Table 1. Description of speech signal attribute sets, measurement and number of attributes.

Attribute Set	Measurements	Number of features
Baseline Features	Shimmer, Jitter, FFP, RPDE, DFA, PPE	21
Time Frequency Features	Bandwidth, Intensity, Formant	11
Wavelet Based Features	F_0 related Wavelet attributes	182
Vocal Fold Features	GQ, GNE, VFER, EMD	22
Mel-Frequency Cepstral Constants (MFCCs)	MFCCs	84
Tunable Q-Factor Wavelet Features	TQWT	432

displayed in Fig. 1. Different data-processing techniques, machine learning classifiers, validation approaches were used in the proposed model. The description of those classifiers, methods, their algorithms, parameters are defined below.

4.1 Data Pre-processing

The PD Speech data-set has in total 753 attributes of 252 subjects. That leads to huge feature space for a comparatively smaller number of data-points. So, data pre-processing lies at the heart of high performing classifier models. The description of the all used methods are given below.

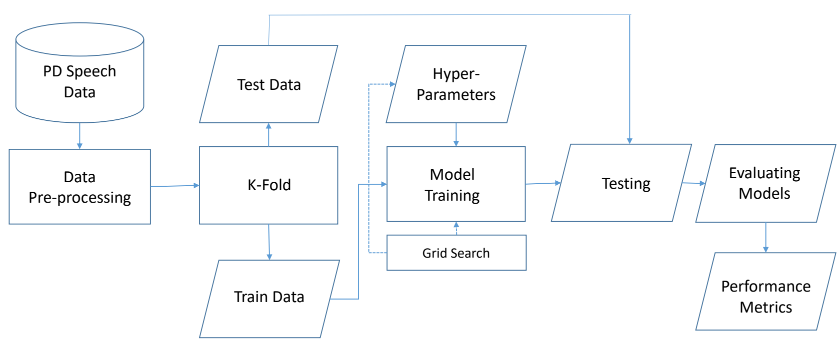

Fig. 1. Block Diagram of Proposed Parkinson's Disease Detection Model.

Data Standardization Standardization is a technique that prevents attributes with large values having more precedence over attributes with smaller values. Z-score is a standardization technique that transforms attribute values to standard scores such that the mean and the variance of the attributes become 0 and 1, respectively. Given n dimensional raw data-set $X = [x_{11}, x_{12}, .., x_{mn}]$ with m data-points, Z-score standardization formula can be depicted as:

$$x_{ij} = Z(x_{ij}) = \frac{x_{ij} - \bar{x}_j}{\sigma_j}. \tag{1}$$

In the equation, $\bar{o}_j$ and σ_j denote the sample mean and standard deviation regarding the jth attribute, accordingly.

Multicollinearity Diagnosis Multicollinearity refers to significant correlation and interdependence among attribute values in the data-set. Multicollinearity causes larger confidence intervals, increases error of estimates in the regression model, and leads to inconsistent outcomes of the classifier algorithm. Variance Inflation Factor (VIF) measurement was used to diagnose multicollinearity of each independent variable. VIF indicates how well a feature can be predicted using the rest of the features. VIF of jth attribute can be computed as:

$$VIF_j = \frac{1}{1 - R_j^2}, \tag{2}$$

where R_j^2 represents the multiple correlation coefficient.

Principal Component Analysis (PCA) PCA is dimension reduction technique such that given an n dimensional standardized data-set Z, PCA outputs a reduced k-dimensional data-set Y.

Steps of the algorithm are as follows:

1. Compute the co-variance matrix of Z as: $COV(Z) = Z^T Z$.

2. Compute the eigen decomposition of $COV(Z)$ as PDP^{-1}, where P denotes the eigenvectors matrix and D denotes the diagonal matrix with eigenvalues in the diagonal.

3. Compute the sorted k-dimensional eigenvectors matrix i.e. the projection matrix P^* based on k largest eigenvalues.

4. Use projection matrix P^* to calculate the output i.e. new k-dimensional space as: $Y = P^* Z$.

Independent Component Analysis (ICA) ICA is a method for differentiating a multivariate signal into its fundamental components. Given an n-dimensional data X, ICA outputs a reduced k-dimensional data Y. Steps of the algorithm are as follows:

1. First, whiten the given X by using formula: $X = ED^{-1/2}E^T X$, where, E denotes orthogonal matrix of eigenvectors of covariance of X and D denotes diagonal matrix of eigenvalues.

2. Initially, choose a weight vector randomly as W.

3. Calculate, $W = Xg(w^T X) - g'(w^T X)W$, where g is a non-quadratic derivative function.

4. Then, compute $W = orthogonal(W)$.

5. If converged, terminate the algorithm, else, go to step 3.

4.2 Cross-Fold Validation

The PD Speech data-set was split into 5 folds or groups. One group was used as test data; the rest of the $k-1$ groups were used as training data. Then, the classifier model was evaluated and scores were recorded. The process was repeated k times with a different fold as the test data each time. The final score of the classier were calculated by combining all scores, as shown equation below:

$$M = \frac{1}{K} \times \sum_{n=1}^{K} C_n \pm \sqrt{\frac{\sum_{n=1}^{K} (C_n - \bar{C})^2}{K-1}}, \tag{3}$$

Here M defines the final performance measurement regarding that classifier and C_n denotes the performance measurement for nth $(1 \leq n \leq K)$ fold.

4.3 Support Vector Machine (SVM)

SVM is a supervised learning technique that classifies data-points using hyperplane. Given, n data-points $(\bar{x}_1, y_1), ..., (\bar{x}_n, y_n)$, we define a hyperplane as a set of points satisfying $\bar{x}$ as $\bar{w}\bar{x} - b = 0$. Objective is to find the maximum-margin hyperplane such that:

$$\begin{aligned} \bar{w}\bar{x}_i - b = 0 \geq 1, \quad \text{if} \quad y_i = +1 \\ \bar{w}\bar{x}_i - b = 0 \leq 1, \quad \text{if} \quad y_i = -1. \end{aligned} \tag{4}$$

Different kernels (such as rbf, polynomial, sigmoid) are used to separate data points in higher dimensional planes.

4.4 Logistic Regression (LR)

Logistic regression computes the probability of existing data-points within a certain class. Given a data-set with n feature set $(x_1, ...x_n)$, binary target class Y, assume $p = P(Y = 1)$. Then, the linear relationship between log-odds, ℓ and features can be described as:

$$\ell = log_b \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n, \tag{5}$$

where the logarithm base is denoted as b and β_i denotes the parameters of the algorithm. Odds can be calculated by using the exponent of log-odds. For an observation, probability that $Y = 1$ can be calculated as:

$$p = \frac{1}{1 + b^{-(\beta_0 + \beta_1 x_1 + ... + \beta_n x_n)}}. \tag{6}$$

We can optimize the model by changing regularization parameters and we can also set different solver algorithms (such as lbfgs, sag) to be used in optimization.

4.5 k-Nearest Neighbour (k-NN)

k-NN is a lazy learning and non-parametric algorithm. The algorithm uses a local network, which generates highly adaptive models. Given n-dimensional input data X and target class data Y the algorithm outputs the probability $P\epsilon[0,1]$ of test data-point, x, where $\sum_i^y P_i = 1$, where y denotes different target classes. Steps of the algorithm are as follows:

1. For the test data-point that is to be classified, compute its distances from each training sample. For euclidean distance, it can be calculated as: $D = \sum_{i=1}^{n} |X_i - x_i|^2$, where, x denotes the sample that needs to be classified and X denotes the training data point.

2. Store the distances in a set and choose k closest data-points based on the distance set.

3. Estimate the class of test data based on the majority class of k closest Neighbours.

4.6 Random Forrest (RF)

Random forest is an ensemble learning classification method that operates by constructing multiple instances of decision trees and combining the results of those trees to estimate output class. It is based on Bootstrap aggregating (Bagging) technique. Given a n dimensional training set X with target class Y steps of RF algorithm can be described as follows:

1. Select k features among the n features such that $k = \sqrt{n}$.
2. Select the best node or split-point among k features.
3. Using best-split, split the node into doughter nodes.
4. Repeat steps (1-3) until node size in minimum.
5. Steps (1-4) is decision tree creation. To build the forest, repeat steps (1-4) n times.
6. For test data, calculate the target class prediction for the data of n decision tree. Assign the majority class to the test data.

4.7 AdaBoost (AB)

AdaBoost is a machine learning algorithm in which outputs of a series of sequential learning algorithms are combined to calculate the final output of the classifier. Given n dimensional training set X and target class $Y\epsilon[-1,+1]$, AdaBoost algorithm can be described as follows:

1. Set $D_l(i) = \frac{1}{m}$, where $i = [1,..,m]$.

2. Using distribution D_t where $t = [1,..,T]$, train weak learners. Assume, a boosted classifier as

$$F_T(x) = \sum_{t=1}^{T} f_t(x), \tag{7}$$

where f_t denotes a weak classifier that outputs the class of object x.

3. At each iteration, assign coefficient α_t to the selected weak learner such that training error E_t is minimized. E_t can be described as:

$$E_t = \sum_i E[F_{t-1}(x_i) + \alpha_t h(x_i)], \tag{8}$$

where $F_{t-1}(x_i)$ is the boosted classifier that was constructed on the previous training stage. $h(x_i)$ denotes the weak learner output hypothesis for samples in the training set.

5 Experiment Result and Discussion

This section discusses different extensive experiments of the proposed Parkinson's Disease detection model, the performances of ML classifiers and how different data pre-processing methods affected the result of the classifiers.

5.1 Experiment and Result of Data Preprocessing and Transformation

The raw PD Speech data-set contained 753 attributes for 252 subjects. The features of the data-set showed multicollinearity among themselves. To view correlation between features, covariance matrix of vocal ford features is shown in Fig. 3. So, to address the high dimensionality of the data different data processing methods (such as scaling, correlation diagnosis, PCA, ICA) were introduced. Feature reduction techniques, such as PCA, ICA further increased the performance of the classifiers. PCA increased classifiers performance more than ICA for most classifiers, except RF. But, ICA required less number of components than PCA to reach its peak potential. For both algorithms, we experimented over different numbers of components to discover the best fit for different classifiers. Fig. 2 visually illustrates the number of features needed to acquire the total variance of the dataset for different classifiers. It can be seen that about (30-60) is enough to get the highest performance for all classifiers. We also combined PCA with correlation diagnosis on data-set which yielded better performances for classifiers such as SVM, LR, AdB.

5.2 Experiment and Result of ML Models

In this experiment, different ML classifiers with different combinations of hyperparameters were used to yield the best performance of the model. Table 3 shows the performance of different classifiers on different data representations. Python's Library and API's were used to implement the model and carry out the experiment. For classification on raw data RF showed the best performance. SMV showed the highest overall performance with accuracy 94.1%. For tree based classifiers (such as : AdB, RF), AdB showed superior highest accuary of 90.4%. k-NN showed the lowest peak performance of 86.3%. The hyperparameters of the classifier models are describe in Table 2. Our model significantly outperformed previous experiments on the same data-set which achieved the highest accuracy of 86% [19].

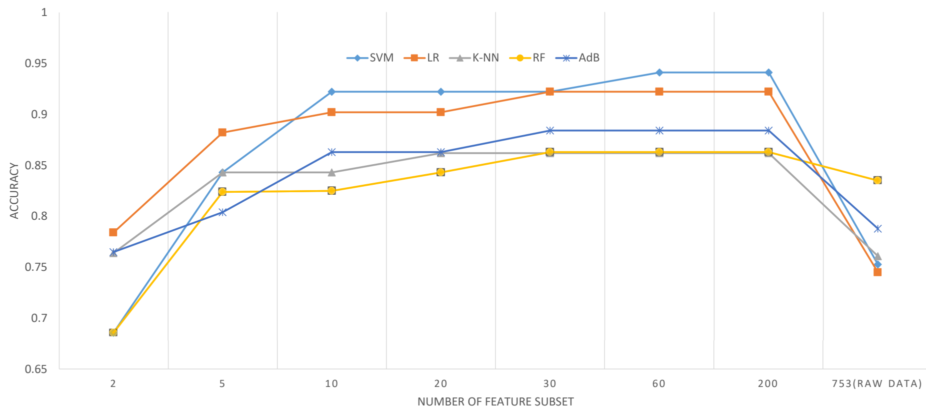

Fig. 2. Accuracy of Different Classifiers for Different Number of PCA Components and Raw Data (at the right most corner).

Table 2. The best performing ML model and preprocessing along with tuned hyperparameters with highest possible accuracy. (Note: the best performing classifier is shown in bold type.)

Classifiers	**Performances**	**Best hyperparameters**
SVM	$\mathbf{0.941 \pm 0.019}$	gamma=scale kernel=sigmoid C=1
LR	0.924 ± 0.087	penalty=l2 solver=lbfgs max_iter=1000
K-NN	0.863 ± 0.075	n_neighbors=6 weights=uniform algorithm=ball_tree
RF	0.88 ± 0.023	n_estimators=1000 criterion=gini min_samples_split=2
Adb	0.90 ± 0.079	n_estimators=50 algorithm=SAMME.R learning_rate=1

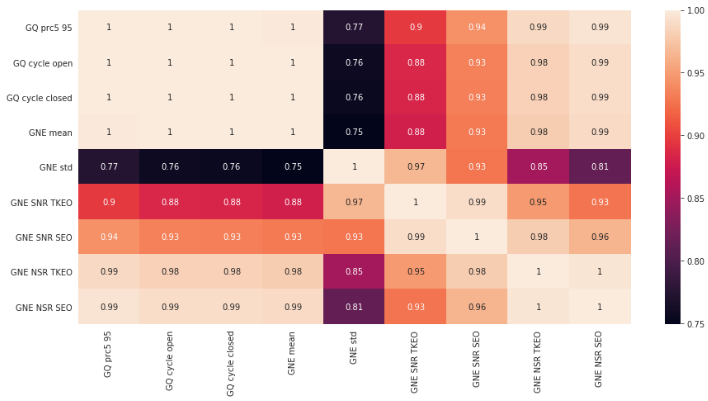

Fig. 3. Heat-map of Correlation Among Vocal Fold Based Features of PD Speech Data-set.

Table 3. Summary of performance of ML classifiers with data pre-processing techniques, feature reduction methods with number of attributes. (Note: the best performing approaches were shown in bold type.)

Classifiers	Pre-processing	Algorithm	n_features	Performance
SVM	Raw Data	N/A	753	0.746 ± 0.146
	Processed Data	Corr	466	0.847 ± 0.120
		PCA	45	0.923 ± 0.07
		ICA	7	0.901 ± 0.023
		Corr+PCA	**55**	$\mathbf{0.941 \pm 0.019}$
LR	Raw Data	N/A	753	0.745 ± 0.116
	Processed Data	Corr	466	0.888 ± 0.084
		PCA	40	0.922 ± 0.07
		ICA	5	0.765 ± 0.091
		Corr+PCA	**55**	$\mathbf{0.924 \pm 0.087}$
K-NN	Raw Data	N/A	753	0.761 ± 0.103
	Processed Data	Corr	466	0.776 ± 0.123
		PCA	**45**	$\mathbf{0.863 \pm 0.075}$
		ICA	17	0.843 ± 0.103
		Corr+PCA	20	0.854 ± 0.059
RF	Raw Data	N/A	753	0.835 ± 0.046
	Processed Data	Corr	466	0.831 ± 0.104
		PCA	40	0.852 ± 0.07
		ICA	**17**	$\mathbf{0.881 \pm 0.023}$
		Corr+PCA	30	0.858 ± 0.019
AdB	Raw Data	N/A	753	0.788 ± 0.134
	Processed Data	Corr	466	0.812 ± 0.107
		PCA	40	0.883 ± 0.07
		ICA	20	0.8063 ± 0.023
		Corr+PCA	**45**	$\mathbf{0.904 \pm 0.079}$

6 Conclusion

In this paper, Parkinson's Disease detection was achieved by employing our proposed model on the PD Speech data-set. Extensive data-processing was introduced to improve the quality of data as the data-set contained huge dimensionality with smaller data-points. Inter-attribute dependency was removed, data was standardized, dimensionality reduction method was introduced. Different machine learning classifiers were fine-tuned using different hyper-parameters for optimization. Cross fold validation were used to diminish the impact of imbalanced data-set. The proposed model significantly outperformed the experiments on the same data-set. For future works, the proposed model can be used as a web application with user-friendly UI. Our proposed model can be used as a widespread tool for prediction of Parkinson's Disease with great efficiency and accuracy.

References

1. Ali, L., Zhu, C., Zhou, M., Liu, Y.: Early diagnosis of parkinson's disease from multiple voice recordings by simultaneous sample and feature selection. Expert Systems with Applications **137**, 22–28 (2019)
2. Benba, A., Jilbab, A., Hammouch, A.: Voice assessments for detecting patients with parkinson's diseases using pca and npca. International Journal of Speech Technology **19**(4), 743–754 (2016)
3. Blauwendraat, C., Nalls, M.A., Singleton, A.B.: The genetic architecture of parkinson's disease. The Lancet Neurology **19**(2), 170–178 (2020)
4. Cantürk, İ., Karabiber, F.: A machine learning system for the diagnosis of parkinson's disease from speech signals and its application to multiple speech signal types. Arabian Journal for Science and Engineering **41**(12), 5049–5059 (2016)
5. Cunningham, L., Mason, S., Nugent, C., Moore, G., Finlay, D., Craig, D.: Home-based monitoring and assessment of parkinson's disease. IEEE Transactions on Information Technology in Biomedicine **15**(1), 47–53 (2010)
6. Cunningham, P., Delany, S.J.: k-nearest neighbour classifiers–. arXiv preprint arXiv:2004.04523 (2020)
7. Dastgheib, Z.A., Lithgow, B., Moussavi, Z.: Diagnosis of parkinson's disease using electrovestibulography. Medical & biological engineering & computing **50**(5), 483–491 (2012)
8. Dorsey, E.R., Bloem, B.R.: The parkinson pandemic—a call to action. JAMA neurology **75**(1), 9–10 (2018)
9. Evgeniou, T., Pontil, M.: Support vector machines: Theory and applications. In: Advanced Course on Artificial Intelligence. pp. 249–257. Springer (1999)
10. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural networks **13**(4-5), 411–430 (2000)
11. Jankovic, J.: Parkinson's disease: clinical features and diagnosis. Journal of neurology, neurosurgery & psychiatry **79**(4), 368–376 (2008)
12. Mohamad, I.B., Usman, D.: Standardization and its effects on k-means clustering algorithm. Research Journal of Applied Sciences, Engineering and Technology **6**(17), 3299–3303 (2013)

13. Naranjo, L., Pérez, C.J., Campos-Roca, Y., Martín, J.: Addressing voice recording replications for parkinson's disease detection. Expert Systems with Applications **46**, 286–292 (2016)
14. Ng, A.Y., Jordan, M.I.: On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In: Advances in neural information processing systems. pp. 841–848 (2002)
15. Pal, M.: Random forest classifier for remote sensing classification. International journal of remote sensing **26**(1), 217–222 (2005)
16. Rao, C.R.: The use and interpretation of principal component analysis in applied research. Sankhyā: The Indian Journal of Statistics, Series A pp. 329–358 (1964)
17. Rodriguez, J.D., Perez, A., Lozano, J.A.: Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE transactions on pattern analysis and machine intelligence **32**(3), 569–575 (2009)
18. Sakar, B.E., Isenkul, M.E., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., Kursun, O.: Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics **17**(4), 828–834 (2013)
19. Sakar, C.O., Serbes, G., Gunduz, A., Tunc, H.C., Nizam, H., Sakar, B.E., Tutuncu, M., Aydin, T., Isenkul, M.E., Apaydin, H.: A comparative analysis of speech signal processing algorithms for parkinson's disease classification and the use of the tunable q-factor wavelet transform. Applied Soft Computing **74**, 255–263 (2019)
20. Schrag, A., Ben-Shlomo, Y., Quinn, N.: How valid is the clinical diagnosis of parkinson's disease in the community? Journal of Neurology, Neurosurgery & Psychiatry **73**(5), 529–534 (2002)
21. Syarif, I., Prugel-Bennett, A., Wills, G.: Svm parameter optimization using grid search and genetic algorithm to improve classification performance. Telkomnika **14**(4), 1502 (2016)
22. Thompson, C.G., Kim, R.S., Aloe, A.M., Becker, B.J.: Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. Basic and Applied Social Psychology **39**(2), 81–90 (2017)
23. Van Den Eeden, S.K., Tanner, C.M., Bernstein, A.L., Fross, R.D., Leimpeter, A., Bloch, D.A., Nelson, L.M.: Incidence of parkinson's disease: variation by age, gender, and race/ethnicity. American journal of epidemiology **157**(11), 1015–1022 (2003)
24. Verleysen, M., François, D.: The curse of dimensionality in data mining and time series prediction. In: International work-conference on artificial neural networks. pp. 758–770. Springer (2005)