



# Article Weighted Hybrid Feature Reduction Embedded with Ensemble Learning for Speech Data of Parkinson's Disease

Zeeshan Hameed <sup>1</sup>, Waheed Ur Rehman <sup>2,3</sup>, Wakeel Khan <sup>4</sup>, Nasim Ullah <sup>5,\*</sup> and Fahad R. Albogamy <sup>6</sup>

- <sup>1</sup> Faculty of Information Technology, College of Computer Science, Beijing University of Technology, Beijing 100124, China; zeeshanhameed.zh@gmail.com
- <sup>2</sup> College of Mechanical Engineering and Applied Electronics Technologies, Beijing University of Technology, Beijing 100124, China; wrehman87@bjut.edu.cn
- <sup>3</sup> Swedish College of Engineering and Technology, Rahim Yar Khan 64200, Pakistan
- <sup>4</sup> Department of Electrical Engineering, Foundation University Islamabad, Islamabad 44000, Pakistan; wakeel.khan@fui.edu.pk
- <sup>5</sup> Department of Electrical Engineering, College of Engineering, Taif University KSA, P.O. Box 11099, Taif 21944, Saudi Arabia
- <sup>6</sup> Computer Sciences Program, Turabah University College, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia; f.alhammdani@tu.edu.sa
- \* Correspondence: nasimullah@tu.edu.sa

Abstract: Parkinson's disease (PD) is a progressive and long-term neurodegenerative disorder of the central nervous system. It has been studied that 90% of the PD subjects have voice impairments which are some of the vital characteristics of PD patients and have been widely used for diagnostic purposes. However, the curse of dimensionality, high aliasing, redundancy, and small sample size in PD speech data bring great challenges to classify PD objects. Feature reduction can efficiently solve these issues. However, existing feature reduction algorithms ignore high aliasing, noise, and the stability of algorithms, and thus fail to give substantial classification accuracy. To mitigate these problems, this study proposes a weighted hybrid feature reduction embedded with ensemble learning technique which comprises (1) hybrid feature reduction technique that increases inter-class variance, reduces intra-class variance, preserves the neighborhood structure of data, and remove co-related features that causes high aliasing and noise in classification. (2) Weighted-boosting method to train the model precisely. (3) Furthermore, the stability of the algorithm is enhanced by introducing a bagging strategy. The experiments were performed on three different datasets including two widely used datasets and a dataset provided by Southwest Hospital (Army Military Medical University) Chongqing, China. The experimental results indicated that compared with existing feature reduction methods, the proposed algorithm always shows the highest accuracy, precision, recall, and G-mean for speech data of PD. Moreover, the proposed algorithm not only shows excellent performance for classification but also deals with imbalanced data precisely and achieved the highest AUC in most of the cases. In addition, compared with state-of-the-art algorithms, the proposed method shows improvement up to 4.53%. In the future, this algorithm can be used for early and differential diagnoses, which are rated as challenging tasks.

Keywords: Parkinson's disease; dimensionality reduction; ensemble learning; hybrid feature learning

## 1. Introduction

The use of machine learning techniques to control diseases is becoming popular nowadays [1–3]. Parkinson's disease damages the nerve cells that are responsible for body movement [4]. As a symptom of Parkinson's disease, speech plays an informative role in the pathogenesis of Parkinson's disease. The convenience of voice acquisition makes remote monitoring of Parkinson's disease possible. However, speech datasets often have noise and high aliasing characteristics. This brings troublesomeness in the processing of speech



**Citation:** Hameed, Z.; Rehman, W.U.; Khan, W.; Ullah, N.; Albogamy, F.R. Weighted Hybrid Feature Reduction Embedded with Ensemble Learning for Speech Data of Parkinson's Disease. *Mathematics* **2021**, *9*, 3172. https://doi.org/10.3390/ math9243172

Academic Editors: Cornelio Yáñez Márquez, Yenny Villuendas-Rey and Miltiadis D. Lytras

Received: 7 October 2021 Accepted: 7 December 2021 Published: 9 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). data. How to extract efficient representational features from Parkinson's speech dataset has received much attention from researchers. Dimensionality reduction is a technique wherein some of the features are alleviated from the original high dimensional data space in such a way that new lower-dimensional data space can effectively represent the original data space. Dimensionality reduction procedures are mainly divided into feature selection and feature transformation [5].

In feature selection, a low dimensional subset of the features is searched from the original high dimensional data space so as a selected subset of features can efficiently describe the original data. Feature selection methods are divided into three groups: wrapper, filter, and embedded [6]. Some of the typical feature selection methods are Relief, SVM-RFE (support vector machines recursive feature elimination), mRMR (minimum redundancy maximum relevancy), p-value, and LASSO (least absolute shrinkage and selection operator). Feature selection methods are widely used for dimensionality reduction of the PD dataset. Rovini E et al. [7] used *p*-value to select the subset of features. Sakar C and Kurun O [8] developed a hybrid feature selection method by combining a feature selection algorithm with an SVM-based classifier and the accuracy of 92% was calculated. Peker M et al. [9] chose minimum redundancy maximum relevance for feature selection and fed these subsets of attributes to the CVANN (complex-valued artificial neural networks). Benba A et al. [10] used a threshold for selecting the subset of features defined by the MDVP (multi-dimensional voice program). Furthermore, selected features were fed to KNN (k-nearest neighbors) and SVM for classification between pathological and normal voices. Shirvan R et al. [11] combined the effects of generic algorithms and KNN for feature selection. The key benefit of the feature selection method is that it retains only those features that are useful for classification. However, feature selection methods are unable to generate new dimensions with high quality. In addition, feature selection-based algorithms lose some information, thereby eliminating some features from a dataset [12].

Feature transformation maps high dimensional data to low dimensional data without eliminating features and retains the sample's information as much as possible [13]. Some of the typical representations of feature extraction methods are LDA (linear discriminant analysis), PCA (principal component analysis), LPP (locality preserving projection), LPDP (locality preserving discriminant projection), and LDPP (local discriminant preservation projection). Chen H et al. [14] proposed a method using a fuzzy KNN approach combined with PCA for PD classification and the accuracy of 96% was calculated. Hariharan M et al. [15] developed an algorithm using LDA and PCA for the recognition of PD subjects. Although PCA and LDA show good performance, they are not reliable for real-world problems. Most of the real-world problems are non-linear having complex tendencies. Both the PCA and LDA are linear feature extraction methods that assume data are in a linear subspace of high dimensional space, as a consequence, they are unable to classify non-linear datasets appropriately [16]. To remove non-linearity problems, kernel forms of PCA and LDA were developed where KPCA (kernelized principal component analysis) and KLDA (kernelized linear discriminant analysis) are typical non-linear feature extraction methods. However, during the process of data mapping, it is quite difficult to find the appropriate kernel function. Manifold learning is another form of feature extraction method which is more adaptable and does not have the limitation of choosing an appropriate kernel function [17]. LPP is a particular example of manifold learning. LPP optimally preserves the neighborhood structure of data and objective function of LPP and minimizes the distance between data points that have a neighborhood relationship in data space [18]. LPP also has some deficiencies. LPP is not only quite sensitive to the number of neighborhood samples but also undergoes the problem of a small sample size. To deal with these shortcomings, some improved versions of LPP were developed. However, developed LPP approaches still have some issues. Most of the LPP based approaches pay more attention to the variance between the classes without considering the large variance within the classes. There is also a problem of instability while mapping the high-dimensional data. It has been observed that if we do the partition of a dataset with a small sample size, there

will be great randomness. It is only due to the difference in data distribution between test datasets and training datasets [19].

In a recent study, researchers are trying to build hybrid systems by combining the benefits of feature selection and feature extraction for the automatic assessment of Parkinson's disease. Uzer M et al. [20] combined the effects of principal component analysis (PCA) and sequential forward selection (SFS) with an artificial neural network classifier for developing a hybrid system. A. Ul Haq et al. [21] developed a hybrid system by using ant-colony optimization algorithms, and the relief filtering method. They used support vector machine (SVM) as a classifier with the K-fold method for cross-validation.

From the above studies, it can be realized that the methods of PD speech feature extraction can be divided into three categories: feature selection, feature extraction, and feature selection combined with feature extraction (FSFE) or feature extraction combined with feature selection (FEFS). Feature selection and feature extraction have distinct strengths. The former focuses on the relevance (interpretation) of features to PD, while the latter focuses on the retention of PD data information. The FSFE or FEFS method combines the advantages of FS and FE, compared to using the feature selection or feature extraction algorithm alone, it can often achieve better classification results.

Besides, the methods that improve the remote diagnosis of PD not only include highquality features, but also the design of a good classification strategy. Ensemble learning is an effective classification performance improvement strategy and has been widely used in various fields of bioinformatics, including the diagnosis of PD. Kadam V et al. [22] selected the attributes by using a genetic algorithm with 10-fold CV SVM and bagging approach using a bootstrap aggregating method for selecting the ensembles of polynomial kernel SVM classifier. Abuhasel K et al. [23] developed a hybrid algorithm that uses NEWFM as a base classifier and integrates it with a standard adaptive boosting method to improve the diagnosing accuracy of PD. Yongming L et al. [24] proposed an algorithm comprised of decision tree-based instance selection and ensemble learning. The CART algorithm was used for selecting the optimal speech samples and ensemble learning combined with random forest (RF), ELM (extreme learning machine), and SVM for training the optimized training samples. Finally, the trained method was applied to test samples. Lauraitis, A et al. [25] investigated the speech impairments to the CNDS (central nervous system disorders) patients. The dataset used for experiments was collected by a neural impairment test suite mobile app. Three domains of feature extraction methods including auditory spectrograms, cestrum domain, and WST (wavelet time scattering, analytic Gabor) were used in this study. For classification, BiLSTM (bidirectional recurrent neural network (RNN) with long short-term memory) and support vector machine (SVM) with polynomial kernel methods were used. They achieved an accuracy of 96.3% with WST-SVM and 94.50% with BiLSTM. Matheus T et al. [26] developed an algorithm to detect Huntington's disease from voice recordings of patients reading Lithuanian poems. They estimated twelve new signal feature extractors by open-SMILE (open source media interpretation by large feature-space extraction) and integrated with KNN (K-nearest neighbours), SVM, MLP (multilayer perceptron), LDA, and QDA (quadratic discriminant analysis) models. Zhang H et al. [27] optimized the samples by using MENN (multi-edit-nearest-neighbor) algorithm and applied a DENN (deco-related-neural-network) ensemble to train those samples. Lastly, the trained model was applied to test samples. However, another combination of hybrid feature learning and hybrid ensemble learning is often overlooked by scholars. Therefore, if the ensemble learning model is designed as a combination of weighted-boosting and bagging models, and used with hybrid feature learning, it can significantly improve detection performance.

To the best of our knowledge, there is no public report about the hybrid feature learning embedded with the ensemble-learning approach for remote monitoring of PD. To develop a convenient and precise remote monitoring system for the identification of PD subjects, this paper designs a weighted hybrid feature reduction embedded with ensemblelearning model. The proposed algorithm first builds a hybrid feature learning method by integrating feature extraction and feature selection techniques; furthermore, uses a weighted-boosting technique to train the base classifier effectively; and then, repeats the above operations to obtain a series of base classifiers; finally, bagging is used to combine all the output of the base classifier to obtain the final classifier. Therefore, the primary contributions and innovations of this paper are as follows:

Proposed a hybrid system effectively integrates the benefits of feature extraction and feature selection. Moreover, the proposed algorithm precisely reduces the intra-class variance, increases the inter-class variance, preserves the neighborhood structure of data, and eliminates the noisy features concurrently that helps to reduce the high aliasing characteristics of PD speech data.

Furthermore, to improve the performance of PD remote diagnosis, the proposed method builds and implements the projection matrix by weighted-boosting and bagging strategies that not only train the model precisely but enhance the stability of the model as well.

### 2. Material and Methods

2.1. Data

In this paper, three different representative datasets including Parkinson Speech Dataset with Multiple Types of Sound Recording (PSDMTSR), a self-collected dataset (named as SelfData), and PARKINSON were used to authenticate the effectiveness of the proposed method.

The PSDMTSR dataset contains 20 healthy people and 20 PD patients. Among the healthy people, half of them were male and the average age of healthy people was 62.55 years. Among the PD patients, 6 of them were female and the rest were male patients, and the average age of PD patients was 64.86 years. More information can be found in [28].

The PARKINSON dataset contains 8 healthy people and 23 PD patients. Among the healthy people, 5 were female and 3 were male with an average age of 58 and 64 years respectively. Among the PD patients, 7 were female and 16 were male patients with an average age of 68.71 and 67.38 years respectively. More information can be found in [29].

SelfData w collected by Army Medical University, Chongqing, China. The dataset contains 21 (9 female, 12 male) healthy people (patients after medication) and 10 PD (5 female, 5 male) patients (without medication). For each of the participants, 13 data samples were recorded, and each sample consisted of 26 features that ended up a design matrix of 1170\*26. More information can be found in [30].

Brief detail about these datasets is shown in Table 1.

Datasets	Attributes											
	Patients	Healthy People	Instances	Features	Classes	References						
PARKINSON	23	8	195	23	2	[29]						
PSDMTSR	20	20	1040	26	2	[28]						
SelfData	36	54	1170	26	2	_						

Table 1. Information about datasets.

For SelfData, patients indicate the patients before treatment, and healthy people represent the patients after treatment.

#### 2.2. The Proposed Method

The objective function of the proposed method minimizes the trace of the local withinclass scatter matrix and maximizes the trace of the between-class scatter matrix while preserving the locality of the sample. Moreover, the proposed algorithm ranks the features and selects top-ranked features, performs weighted-boosting to characterize the wrongly categorized samples, and finally, ensemble mapping for constructing the final output by using weight coefficients. In order to tune the parameters for best performance, hold-outcross-validation is used to validate the method. In this study, the data matrix is represented as  $X = [x_1, x_2, x_3, ..., x_N]^T = [X^1, X^2, X^3, ..., X^C]^T \in \mathbb{R}^{N \times F}$ , where  $N = \sum_{i=1}^{C} N_i$  denotes the number of data samples and  $N_i$  is the number of data samples in the dataset  $X^i$ , F represents the number of dimensions of dataset X, C denotes the number of classes in dataset X and  $y = [y_1, y_2, y_3, ..., y_N]^T \in \mathbb{R}^N$  represents the labels of each data sample.  $P = (p_1, p_2, p_3, ..., p_f) \in \mathbb{R}^{F \times f}$  is the projection matrix that is used for mapping the high dimensional data from  $\mathbb{R}^{N \times F}$  to a low dimensional space  $\mathbb{R}^{N \times f}$ , where F > f.

To make data samples with the same class label as close as possible, this algorithm reduces the intra-class variance matrix as follows:

$$\min_{P} \sum_{c=1}^{C} \left[ \|P^{T} x'^{(c)} - P^{T} \overline{x'^{(c)}}\|^{2} \right]_{x'^{(c)} \in X_{wc}} = \min_{P} \sum_{c=1}^{C} \left[ P^{T} \left( x'^{(c)} - \overline{x'^{(c)}_{w}} \right) \left( x'^{(c)} - \overline{x'^{(c)}_{w}} \right)^{T} P \right]_{x'^{(c)} \in X_{wc}} = \min_{P} P^{T} S_{W} P^{n} \quad (1)$$

where  $S_W = \sum_{c=1}^{C} \sum_{x'^{(c)} \in X_{wc}} \left( x'^{(c)} - \overline{x'^{(c)}_{w}} \right)^T \left( x'^{(c)} - \overline{x'^{(c)}_{w}} \right)$  is the matrix of intra-class variance,  $\overline{x'^{(c)}_{w}} = \frac{1}{N_c} \sum_{i=1}^{N_c} x'^{(c)}_i$  is the center of *cth* class and  $x'^{(c)}_i$  is the *i*th sample of class  $x'^{(c)}$ .

$$\max_{P} \sum_{c=1}^{C} \left[ \|P^{T} \overline{x'_{w}^{(c)}} - P^{T} \overline{X'}\|^{2} \right]_{\overline{X'} \in X} = \max_{P} \sum_{c=1}^{C} \left[ P^{T} \left( \overline{x'_{w}^{(c)}} - \overline{X'} \right) \left( \overline{x'_{w}^{(c)}} - \overline{X'} \right)^{T} P \right]_{\overline{X'} \in X}$$

$$(2)$$

where  $S_B = \sum_{c=1}^{C} \left( \overline{x'_w^{(c)}} - \overline{X'} \right) \left( \overline{x'_w^{(c)}} - \overline{X'} \right)^T$  means the matrix of inter-class variance,  $\overline{X'} = \frac{1}{N} \sum_{i=1}^{N} x_i$  is the center of *X*, and  $x_i$  is the *i*th sample in *X*.

To preserve the neighborhood relationship means the sample structure would be preserved after mapping which can be described as

$$\sum_{c=1}^{C} \sum_{j=i}^{N} \sum_{i=1}^{N_c} A_{ij}^c \Big[ \|P^T x_i^{(c)} - P^T x_j^{\prime}\|^2 \Big]_{x_i^{(c)}, x_j^{\prime} \in X_{train}} \\ = \sum_{c=1}^{C} \Big[ P^T \left( \sum_{t=1}^{N_c} x_i^{(c)} D_c x_i^{(c)^T} - \sum_{j=1}^{N} \sum_{i=1}^{N_c} x_i^{\prime(c)} A_{ij}^c x_j^{\prime T} \right) P \Big]_{x_j^{\prime}, x_i^{\prime(c)} \in X_{train}} \\ = \sum_{c=1}^{C} \Big[ P^T X_i^{\prime(c)} (D^c_{ii} - A^c) X_{train}^{\prime} P \Big]_{X_i^{\prime(c)} \in X_{train}} \\ = \sum_{c=1}^{C} \Big[ P^T X_i^{\prime(c)} (D - A) X_{train}^{\prime} P \Big]_{X_i^{\prime(c)} \in X_{train}}$$
(3)

where Z = D - A is a Laplacian matrix,  $D^c_{ii} = \sum_j A^c_{ij}$  is a diagonal matrix, and  $A_{ij}$  is an affinity matrix that can be calculated in two different ways

Heat kernel: 
$$A_{ij}^c = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}}, & \text{if } x_i^c N_k(x_j) \parallel x_j \in N_k(x_j^c) \\ 0, & \text{otherwise} \end{cases}$$
 (4)

Simple – minded : 
$$A_{ij}^c = \begin{cases} 1, if x_i^c N_k(x_j) \parallel x_j \in N_k(x_j^c) \\ 0, otherwise \end{cases}$$
 (5)

where  $N_k$  is the set of samples nearest to sample x, and t is a heat kernel parameter. Therefore, the objective function of the proposed algorithm can be expressed as follows

$$\min_{P} Tr\left[P^{T}S_{W}P - P^{T}S_{B}P + X_{train}Z X_{train}^{T}P\right]$$

This equation is equivalent to the Lagrange function

$$L(P,\lambda) = P^{T} \left( S_{W} - \lambda \left[ \mu S_{B} - \gamma \left( X_{train} Z X_{train}^{T} \right) \right] \right) P$$
(6)

Taking the derivative with respect to *P* 

$$\frac{\partial}{\partial P^{n}}[L(P,\lambda)] = 0 \quad \Rightarrow \frac{\partial \left\{ P^{T} \left[ S_{W} - \lambda \left[ \mu S_{B} - \gamma \left( X_{train} Z X_{train}^{T} \right) \right] \right] P \right\}}{\partial P^{n}} = 0$$
  
$$\Rightarrow 2S_{W}P - 2\lambda \left[ \mu S_{B}P - \gamma \left( X_{train} Z X_{train}^{T} \right) P \right] = 0 \Leftrightarrow S_{W}P$$
  
$$= \lambda \left[ \mu S_{B} - \gamma \left( X_{train} Z X_{train}^{T} \right) \right] P \Rightarrow \lambda P$$
  
$$= PS_{W} \left[ \mu S_{B} - \gamma \left( X_{train} Z X_{train}^{T} \right) \right]^{-1}$$
(7)

where the penalty factors  $\mu$  and  $\gamma$  give a trade-off between the manifold preservation term and the local discrimination term that can be simply tuned via a substitute strategy.

After mapping data on *P*, the algorithm filters noisy and redundant features. The proposed algorithm ranks each feature according to its weight. This weight would be the approximation of the probability that a feature will be able to separate two classes in local neighbors. The objective function for feature selection can be described as follows

# $W(f_i) = P(different \ value \ of \ f_i | different \ class) - P(different \ value \ of \ f_i | same \ class)$

$$W(f_{i}) = \sum_{c \neq class(x')} \left[ \frac{P(x')}{1 - P[class(x')]} \sum_{j=1}^{k} \frac{diff[f_{i}, x', M_{j}(x')]}{mk} \right]_{x' \in} -\sum_{j=1}^{k} \frac{diff[f_{i}, x', H_{j}(x')]}{mk}$$
(8)

where  $diff[f_i, x', M_j]$  represents the difference of data sample x' and  $M_j$  on feature  $f_i$ ,  $H_j(x')$  is the neighbor sample of the same class, and  $M_j(x')$  is the neighbor data sample of opposite classes. Initially, the weight for each feature is set to 0. After that sample x' is randomly taken from the training set  $X_{train}$ , and then k nearest hits of x' are found from the sample set similar to x', and k nearest misses are found from the sample set of different classes of each x', and then the weight of each feature is updated. Finally, top  $W(f_i)_{f \in F}$  (i = 1, ..., F) features of  $X'_{train}$  with the highest weights will be selected for training the model.

$$W(f_{i}) = W(f_{i}) + \sum_{c \neq class(x')} \left[ \frac{P(x')}{1 - P[class(x')]} \sum_{j=1}^{k} \frac{diff[f_{i}, x', M_{j}(x')]}{mk} \right] - \sum_{j=1}^{k} \frac{diff[f_{i}, x', H_{j}(x')]}{mk}$$
(9)

In order to train the model effectively, the proposed algorithm contains a weightedboosting technique with cross-validation by validation dataset  $X_V$ . This technique gives higher weights to misclassified samples and lower weights to the classified samples and the sum of all weights is 1.

The objective function of AdaBoost is as follows

$$sign\left[\sum_{s=1}^{S} \rho_s h_s(x'_s)\right]$$

where *S* is the number of samples,  $\rho_s = \frac{1}{2}ln\left(\frac{1-\epsilon_s}{\epsilon_s}\right)$  is the weight of the classifier  $h_s(x')$  for sample *s*. Initially, equal weights  $\left[\omega(x'_s, y_s)_{y_s \in [-1,1]} = \frac{1}{5}, s = 1, \dots, S\right]$  is assigned to each of the training example *s*. The learning algorithm uses the training data set and x'

The weights for each example can be updated as follows

$$\omega_{s}(x'_{s}, y_{s}) = \frac{\omega_{s-1}(x'_{s}, y_{s}).e^{-\rho_{s}y_{s}h_{s}(x'_{s})}}{R_{s}}$$
(10)

where  $R_s$  is a normalization factor to ensure that sum of all weights is equal to 1.

Finally, the proposed algorithm is used on *n* different randomly sampled subsets  $X_{train}^n$  of  $X_T$ , so *n* different classifiers (*classifier*<sub>1</sub>,...,*classifier*<sub>n</sub>) will be obtained. Therefore, the final classifier can be obtained as follows:

$$Classifier_{Final} = \beta_1 classifier_1 + \beta_2 classifier_2 +, \dots, +\beta_n classifier_n = \sum_{i=1}^{n} \beta_i classifier_i$$
(11)

where  $\beta$  is a weight coefficient obtained by Bayesian fusion in which sample class is determined based on maximum posterior probability. For instance, n number of classifiers are obtained using n number of training datasets, and n number of predictions results obtained by using validation dataset. The final probability matrix is obtained by joining the prediction of n classifiers. The weight coefficient  $\beta$  can be calculated as

$$\beta_i = \frac{N_i}{N_{valid}}, \ (i = 1, 2, \dots, n)$$
(12)

where  $N_i$  is the number of times *i*th classifier is nominated as prediction result,  $N_{valid}$  is the number of samples in the validation dataset.

Based on the above description, the complete process of the proposed algorithm is as follows. The dataset X is divided into three parts; training  $X_R$ , test  $X_T$ , and validation  $X_V$  datasets. The training data  $X_R$  is further randomly subdivided into n parts (n is the number of networks and in each part, 80% of the whole training data is chosen randomly). Each of the  $X_{tn}$  training datasets are processed separately. After that, the proposed algorithm is used for feature reduction of each  $X_{tn}$  training datasets and the  $X_V$  is used to validate each classifier. Lastly, the final output is obtained by integrating all independent networks. The flow chart and pseudo-code for the proposed method can be seen below.

Figure 1 shows the flowchart of the proposed algorithm and Figure 2 shows the weight-Boosting process. The details of the pseudo-code of the proposed algorithm are given in Algorithm 1 below.

#### 2.3. Classifiers

The experiments were performed using three types of classifiers including RF (random forest), SVM, and ELM (extreme learning machine). Random forest classifier uses multiple trees for training and prediction, has been used for the classification of PD speech data [31,32]. SVM finds the hyperplane in the sample space to maximize the classification interval from different classes by using support vectors and is being used in various fields including PD speech data classification [33,34]. ELM is a comprehensive feed-forward network consists of the single-hidden layer which is used as a classifier in PD speech classification [35,36].

## 2.4. Evaluation Criteria

In this study, five model evaluation criteria are used including precision (Pre), recall (Rec), accuracy (Acc), G-mean, and Area under the ROC (AUC). These criteria are built from a confusion matrix that contains classified and misclassified samples of each class. G-mean is used to evaluate the results of imbalanced data samples, while AUC evaluates the performance of the classifiers. AUC is a numerical value that can be calculated by

the area under the ROC curve (AUC). In this study, the classification of PD is a two-class problem, and the formation of the confusion matrix can be seen in Table 2.

The evaluation criteria can be constructed as follows

$$Accuracy (Acc \%) = \frac{TP + TN}{TP + FP + FN + TN} \times 100$$
(13)

$$Precision (Pre) = \frac{TP}{TP + FP}$$
(14)

$$Recall (Rec) = \frac{TP}{TP + FN}$$
(15)

$$G - mean = \sqrt{Rec * Spe} = \sqrt{\frac{TP}{TP + FN} * \frac{TN}{FP + TN}}$$
(16)

where Spe is specificity.



Figure 1. Flowchart of the proposed algorithm.

## 2.5. Experimental Environment

Windows version 7 with a 64-bit operating system was used for the experiment and the memory size was 8GB. The algorithm was executed in MATLAB version 2018a. Multiple

experiments were carried out to evaluate the performance of the proposed PD diagnosis system. All experiments were performed under the same experimental conditions and each experiment was repeated 10 times and the average was recorded to mitigate the effect of irregularity. To accomplish the optimal performance of the model, the parameters were tuned according to Table 3. Moreover, 300 trees were used for random forest (RF), and 5000 hidden neurons were used for ELM.

Table 2. Confusion matrix for a two-class problem.

T.I.	-1	Pred	iction
Lab	el	Positive	Negative
Paal	positive	True Positive (TP)	False Negative (FN)
Keai –	negative	False Positive (FP)	True Negative (TN)
Algorithm 1. The pseud	lo-code of the propose	ed algorithm	
<b>Input:</b> Data set <i>X</i> , labels <b>Output:</b> Final Prediction	s Y n labels <i>Classifier<sub>Final</sub></i>		
Begin			
1: Randomly divid	le dataset X into train	ing data $X_R$ , test data $X_T$ , a	nd validation data $X_V$ .
2: Randomly selec	t training sets as ( $r_s *$	$(X_T)$ into $X_{t1}, X_{t2}, X_{t3}, \dots$	$X_{tn}$ where <i>n</i> is the
number of subspaces, ar	nd $r_s = 0.8$		
3: <b>for</b> $i = 1$ to $n$ (where $n = 1$ to	here $n$ is the number of	of stacks or networks)	
Function = H	lybrid-Feature-reduc	tion ( $X_{tn}$ )	
Feature	e extraction	·· , 1 · · ·	
Ca	alculate $S_W$ (matrix of	intra-class variance)	
	alculate $S_B$ (matrix of	inter-class variance)	
	alculate $\Delta$ (Laplacian)	inatrix)	
S	actuate A (Annuty in lye for $P^n$ (The project	attin matrix)	
Feature Sele	ction		
Initia	alize weight $W(f_i) = 0$	) for each of the features	
Calc	ulate diff $\int f \cdot x' M \cdot (x)$	(difference of sample $r'$	and neighbor sample of
the same class)		)] (anterence of sample x	und neighbor bumpie o
Cala	alata di GG G ad II (ad		
	unate $uij j [j_i, x, \pi_j(x)]$	) (difference of sample x a	and heighbor sample of
the different classes)	$A_{1} = A_{1}(f)$		
Calci	ulate $W(f_i)$	f < Г	
Selec	ad boosting	J < F	
vvergitt	· 1 · 1.	(1) 1	
A	ssign equal weights	$\omega(x'_s, y_s)_{y_s \in [-1,1]} = \frac{1}{S}, s =$	$1 \dots, 5$ to each of the
training samples	T : 11/3/ 11	\ \	
	Irain model ( <i>Niodel</i> <sub>n</sub> )	.)	
	Validate ( $Nlodel_n$ ) Find Miss classified	ammlaa	
		$(1-\epsilon_{c})$ (1.1.1.)	
	Calculated $\rho_s = \frac{1}{2}ln$	$\left(\frac{\varepsilon_s}{\epsilon_s}\right)$ (Weights)	
	Update weights $\omega_s(z)$	$(s, y_s)$	
	Apply weighted-boo	sting $sign\left[\sum_{s=1}^{5}\rho_{s}h_{s}(x'_{s})\right]$	
end for			
4: Ensemble mapp	ping		
Calculate /	$\beta_i = rac{N_i}{N_{valid}}$ , $(i = 1, 2, \dots$	., <i>n</i> ) (weight coefficient)	
Calculate	$Classifier_{Final} = \sum_{i=1}^{n} \beta$	nclassifier <sub>n</sub> (Predicted outp	out)
End	<i>t</i> =1		



Figure 2. Weighted-boosting process.

Table 3	. The parameters	for the	proposed	algorithm.
---------	------------------	---------	----------	------------

Symbol	Meaning	Parameter Settings
r <sub>s</sub>	Random sampling ratio	0.8
п	Number of subspaces	5
γ	Penalty factor for $P^{nT} \left( X'_{train}^n Z X'_{train}^n \right) P^n$	$10^{-4}, 10^{-3}, \dots, 10^{4}$
λ	Penalty factor for $P^{nT} \left[ \mu S_B - \gamma \left( X'^n_{train} Z X'^{nT}_{train} \right) \right] P^n$	$10^{-4}, 10^{-3}, \dots, 10^{4}$
μ	Penalty factor for $P^{nT}S_BP^n$	$10^{-4}, 10^{-3}, \dots, 10^{4}$
f	Features after feature reduction	5, 10, 15, 20,
k	Nearest neighbor samples in affinity Matrix A	5
t	Kernel parameter of affinity Matrix A	$10^{-4}, 10^{-3}, \dots, 10^{4}$
β	Fusion weight coefficient	Calculated by Bayesian fusion

## 3. Results

In this section, the results of the proposed algorithm are analyzed and compared with feature selection, feature transformation, and some state-of-the-art algorithms which have been extensively used for PD diagnosis. For validation purposes, hold-out-cross-validation is used in which the dataset is randomly and equally divided into training (1/3), validation (1/3) and test (1/3) sets. Since each subject contains multiple samples in the dataset, this validation process can effectively avoid data overlapping.

#### 3.1. Comparison with Feature Extraction Methods

In this section, the results of proposed and various feature extraction methods were compared. These feature extraction methods involved LDA [4,15,37], PCA [31,32,38,39], KPCA [37], LPP [37,40] (simple-minded and heat kernel) LDPP (simple-minded and heat kernel). The classification results can be seen in Table 4. The results in Table 4 shows that the proposed algorithm achieved better performance on multiple evaluation criteria irrespective of datasets and classifiers. It is worth noticing that for PSDMTSR datasets, the proposed method always performed well compared with feature extraction methods. This indicates that the proposed method not only achieves higher classification accuracy but also can deal with imbalanced data samples effectively. Same as PSDMTSR, the proposed algorithm performed better for the PARKINSONS dataset as well on the accuracy, precision, recall, and G-mean which also shows the superiority of the proposed method. For the results using SelfData, although feature extraction methods help to improve the performance as to compare to N\_DR (results without dimension reduction), the improvement of the proposed method is most noticeable. Compared with N\_DR, the classification accuracy was improved for ELM by 18.66% (from 0.4867–0.6733) and for SVM (linear) by 8.67% (from 0.56–0.6467). Moreover, it is worth noticing that affinity matrix with simple-minded offered lower classification accuracy than that of the heat kernel. It is because while making the affinity matrix, the heat kernel provides weights to the neighbor samples according to the distance from the sample under observation (greater weights are given to the closest neighbor samples). However, simple-minded does not provide this functionality.

			Ac	c			1	Pre			R	lec			G-M	ean	
Dataset\Classifier		SVM (linear)	SVM (RBF)	RF	ELM												
	N_DR	0.6833	0.7583	0.7583	0.7417	0.746	0.7463	0.7529	0.7647	0.8125	0.975	0.9625	0.9	0.5876	0.5629	0.5804	0.6185
PARKINSONS	PCA	0.75	0.75	0.75	0.7167	0.7515	0.7297	0.7398	0.7386	0.9375	1	0.975	0.9	0.5929	0.5	0.5408	0.5612
	KPCA	0.7167	0.675	0.7583	0.6917	0.7061	0.6727	0.7453	0.7431	0.9875	1	0.975	0.825	0.4157	0.1581	0.5629	0.5921
	LDA	0.7083	0.7167	0.7417	0.6667	0.748	0.7093	0.7629	0.7061	0.8625	0.9875	0.9125	0.8625	0.5874	0.4157	0.6042	0.487
	LPP(S)	0.7667	0.75	0.725	0.75	0.7591	0.7297	0.7145	0.7954	0.975	1	0.9875	0.8625	0.5842	0.5	0.4444	0.6729
TARRENSONS	LPP(H)	0.7417	0.7833	0.7417	0.7083	0.7533	0.7723	0.7337	0.7639	0.925	0.9625	0.975	0.8125	0.589	0.6396	0.5178	0.6374
	LDPP(S)	0.7583	0.75	0.75	0.6917	0.7686	0.7309	0.7393	0.7371	0.925	1	0.975	0.8375	0.627	0.5	0.5408	0.5788
	LDPP(H)	0.7667	0.7667	0.7917	0.725	0.7657	0.756	0.7715	0.7571	0.95	0.975	0.9875	0.875	0.6164	0.5842	0.6285	0.6098
	Proposed (S)	0.8167	0.7583	0.8167	0.85	0.7899	0.7382	0.7921	0.8396	1	1	0.9875	0.975	0.6708	0.5244	0.6849	0.7649
	Proposed (H)	0.85	0.8083	0.85	0.8917	0.8356	0.781	0.8194	0.9069	0.975	1	1	0.95	0.7649	0.6519	0.7416	0.8581
	N_DR	0.6125	0.575	0.5625	0.6188	0.6152	0.5721	0.5545	0.6458	0.65	0.6375	0.6	0.625	0.6114	0.5716	0.5612	0.6187
	PCA	0.575	0.55	0.5813	0.575	0.5924	0.5561	0.5816	0.5914	0.6625	0.625	0.6625	0.5625	0.5683	0.5449	0.5755	0.5749
	KPCA	0.5188	0.55	0.5688	0.5063	0.6633	0.6883	0.5655	0.496	0.35	0.4625	0.7	0.45	0.4905	0.543	0.5534	0.5031
	LDA	0.6	0.5563	0.5938	0.55	0.6037	0.5515	0.592	0.5565	0.65	0.6125	0.6375	0.6	0.5979	0.5534	0.5921	0.5477
PEDMTER	LPP(S)	0.5625	0.5625	0.5813	0.6	0.5828	0.5602	0.5891	0.6344	0.625	0.65	0.675	0.575	0.559	0.5557	0.5736	0.5995
TODMIOR	LPP(H)	0.5438	0.5688	0.6063	0.5563	0.5486	0.5765	0.6068	0.5839	0.625	0.5875	0.625	0.575	0.5376	0.5684	0.606	0.5559
	LDPP(S)	0.5688	0.5813	0.5	0.6	0.5618	0.5959	0.5082	0.6234	0.7125	0.65	0.5875	0.6125	0.5503	0.5772	0.4923	0.5999
	LDPP(H)	0.5438	0.5	0.5438	0.5688	0.5943	0.5201	0.54	0.5819	0.5625	0.525	0.6	0.5375	0.5434	0.4994	0.5408	0.5679
	Proposed (S)	0.7375	0.6625	0.6875	0.7563	0.7534	0.6628	0.6854	0.764	0.775	0.75	0.7375	0.7875	0.7365	0.6567	0.6857	0.7556
	Proposed (H)	0.7125	0.6625	0.7063	0.7625	0.7219	0.6425	0.6773	0.8171	0.7375	0.7875	0.8375	0.7375	0.7121	0.6506	0.6939	0.7621
	N_DR	0.56	0.5433	0.5367	0.4867	0.3962	0.3305	0.3825	0.3687	0.3583	0.2417	0.2583	0.35	0.4988	0.4242	0.4319	0.4497
	PCA	0.5733	0.5467	0.54	0.5267	0.3628	0.4588	0.3669	0.4339	0.2083	0.2667	0.25	0.4417	0.4125	0.4422	0.4282	0.5076
	KPCA	0.5867	0.5967	0.5367	0.58	0.3917	0.4417	0.3791	0.4619	0.0667	0.075	0.3083	0.1667	0.2494	0.2661	0.4609	0.3776
	LDA	0.5867	0.5667	0.5133	0.49	0.3288	0.1742	0.3455	0.3256	0.1583	0.0833	0.2833	0.3333	0.3716	0.2722	0.4346	0.4451
SelfData	LPP(S)	0.56	0.5067	0.5633	0.5033	0.3468	0.3575	0.4558	0.4105	0.1833	0.225	0.2917	0.5083	0.3856	0.3953	0.466	0.5041
Schibuta	LPP(H)	0.5833	0.59	0.5333	0.4533	0.2171	0.1485	0.3181	0.3044	0.0833	0.075	0.15	0.3083	0.2764	0.2646	0.344	0.4118
	LDPP(S)	0.5967	0.5867	0.55	0.4867	0.3329	0.2432	0.2744	0.3615	0.1917	0.1333	0.2	0.3583	0.4076	0.3443	0.3958	0.4528
-	LDPP(H)	0.5667	0.5967	0.5233	0.4767	0.3078	0.3219	0.3032	0.3564	0.1417	0.2083	0.1833	0.4083	0.347	0.4222	0.3708	0.4618
	Proposed (S)	0.6467	0.6267	0.6733	0.6733	0.5808	0.3405	0.6807	0.6816	0.1833	0.1	0.375	0.525	0.4186	0.3127	0.5719	0.6367
	Proposed (H)	0.6133	0.64	0.6567	0.6733	0.4511	0 5592	0.4595	0 5945	0 1922	0 1667	0 3417	0 5167	0.4062	0 3991	0 5442	0.6339

Note: H indicates heat kernel and S means the simple-minded method used to calculate the affinity matrix (Equation (4) and (5) respectively).

#### 3.2. Comparison with Feature Selection Methods

In this section, the proposed algorithm was compared with feature selection methods included mRMR [9,41–43], *p*-value, ReliefF [41,44,45], SBS [46], SFS [15,37], and SVM\_RFE. The results can be seen in Table 5.

The results in Table 5 illustrate that irrespective of classifier and dataset, the proposed algorithm performed well compared to feature selection algorithms in all evaluation indicators. For PARKINSONS and PSDMTSR datasets, the proposed algorithm achieved better results in terms of accuracy, precision, recall, and G-mean which shows the effectiveness of the proposed algorithm. For the SelfDataset, the advantage of the proposed algorithm is obvious for all classifiers in terms of accuracy and precision. Moreover, for recall and G-mean, the proposed algorithm with RF and ELM classifiers outperformed feature selection methods. The improvement of the proposed algorithm with ELM is most obvious which is 18.66% compared with N-DR and even 15% compared with top feature selection methods (mRMR, ReliefF).

## 3.3. Significance Analysis

For further verification, significance analysis difference between proposed and feature reduction algorithms was performed, and the results were recorded in Table 6. The results in Table 6 shows that no matter which classifier do we use, the proposed algorithm shows significant improvement compared with most of the existing dimensionality reduction methods.

Dataset\Classifier			A	Acc		Pre			Rec				G-Mean				
Dataset	Classifier	SVM (Linear)	SVM (RBF)	RF	ELM												
	N_DR	0.6833	0.7583	0.7583	0.7417	0.746	0.7463	0.7529	0.7647	0.8125	0.975	0.9625	0.9	0.5876	0.5629	0.5804	0.6185
PARKINSONS	mRMR	0.7167	0.6667	0.725	0.7417	0.7227	0.673	0.7198	0.761	0.95	0.975	0.975	0.9125	0.4873	0.2208	0.4684	0.6042
	ReliefF	0.75	0.675	0.725	0.6917	0.7543	0.6758	0.728	0.7146	0.9375	0.9875	0.95	0.9125	0.5929	0.2222	0.5111	0.4776
	Pvalue	0.7667	0.675	0.75	0.7167	0.7691	0.6758	0.7529	0.7505	0.95	0.9875	0.9625	0.8875	0.6164	0.2222	0.5593	0.5769
	SBS	0.7583	0.675	0.725	0.7583	0.7691	0.6727	0.7279	0.7546	0.925	1	0.95	0.9625	0.627	0.1581	0.5111	0.5804
	SFS	0.7917	0.675	0.7333	0.7333	0.7873	0.6727	0.7358	0.7602	0.95	1	0.95	0.9	0.6718	0.1581	0.5339	0.6
	SVM_FRE	0.75	0.675	0.7333	0.6917	0.7396	0.6727	0.7265	0.7291	0.9875	1	0.975	0.85	0.5211	0.1581	0.4937	0.5646
	Proposed (S)	0.8167	0.7583	0.8167	0.85	0.7899	0.7382	0.7921	0.8396	1	1	0.9875	0.975	0.6708	0.5244	0.6849	0.7649
	Proposed (H)	0.85	0.8083	0.85	0.8917	0.8356	0.781	0.8194	0.9069	0.975	1	1	0.95	0.7649	0.6519	0.7416	0.8581
	N_DR	0.6125	0.575	0.5625	0.6188	0.6152	0.5721	0.5545	0.6458	0.65	0.6375	0.6	0.625	0.6114	0.5716	0.5612	0.6187
	mRMR	0.5688	0.5375	0.6188	0.5813	0.597	0.45	0.6032	0.5986	0.4	0.175	0.675	0.5875	0.5431	0.3969	0.6162	0.5812
	ReliefF	0.5938	0.4875	0.5313	0.6188	0.5314	0.2551	0.5305	0.633	0.525	0.3875	0.6125	0.6375	0.5898	0.4771	0.525	0.6185
	Pvalue	0.5688	0.475	0.5625	0.6313	0.6023	0.3851	0.5681	0.6309	0.6125	0.25	0.625	0.6875	0.5671	0.4183	0.559	0.6287
PSDMTSR	SBS	0.5688	0.4813	0.5313	0.6125	0.5828	0.3078	0.5236	0.6198	0.5	0.2875	0.5875	0.625	0.5646	0.4405	0.5283	0.6124
	SFS	0.55	0.4938	0.5563	0.575	0.5754	0.4937	0.5686	0.5737	0.4625	0.2625	0.575	0.625	0.543	0.4362	0.5559	0.5728
	SVM_FRE	0.5375	0.5313	0.5375	0.5313	0.4834	0.5	0.5383	0.5395	0.45	0.4125	0.5875	0.5875	0.5303	0.5178	0.5352	0.5283
	Proposed (S)	0.7375	0.6625	0.6875	0.7563	0.7534	0.6628	0.6854	0.764	0.775	0.75	0.7375	0.7875	0.7365	0.6567	0.6857	0.7556
	Proposed (H)	0.7125	0.6625	0.7063	0.7625	0.7219	0.6425	0.6773	0.8171	0.7375	0.7875	0.8375	0.7375	0.7121	0.6506	0.6939	0.7621
	N_DR	0.56	0.5433	0.5367	0.4867	0.3962	0.3305	0.3825	0.3687	0.3583	0.2417	0.2583	0.35	0.4988	0.4242	0.4319	0.4497
	mRMR	0.5433	0.6	0.5267	0.5233	0.1642	0	0.3175	0.2822	0.1	0	0.1583	0.225	0.2896	0	0.3497	0.4031
	ReliefF	0.5833	0.5933	0.5	0.5233	0.2722	0.025	0.3014	0.3798	0.1917	0.0083	0.2	0.2917	0.4023	0.0905	0.3742	0.4446
	Pvalue	0.58	0.5933	0.5433	0.5	0.2889	0.075	0.3511	0.3865	0.25	0.0167	0.2167	0.35	0.4472	0.1277	0.4061	0.4583
SelfData	SBS	0.5133	0.6	0.5067	0.5	0.2933	0	0.2931	0.3938	0.2917	0	0.25	0.4333	0.4391	0	0.4116	0.4857
	SFS	0.5733	0.5967	0.5267	0.5033	0.2792	0.0333	0.3825	0.3697	0.1333	0.0083	0.2333	0.3417	0.3399	0.0908	0.4105	0.4569
	SVM_FRE	0.5667	0.6	0.5433	0.52	0.1883	0	0.265	0.3789	0.175	0	0.1333	0.35	0.3806	0	0.33	0.4708
	Proposed (S)	0.6467	0.6267	0.6733	0.6733	0.5808	0.3405	0.6807	0.6816	0.1833	0.1	0.375	0.525	0.4186	0.3127	0.5719	0.6367
	Proposed (H)	0.6133	0.64	0.6567	0.6733	0.4511	0.5592	0.6585	0.594	0.1833	0.1667	0.3417	0.5167	0.4062	0.3991	0.5442	0.6339

#### Table 5. Comparison with feature selection methods.

Detecto	Classifiers\Methods								1	Methods						
Datasets	Classi	ners\methous	N_DR	mRMR	ReliefF	p-Value	SBS	SFS	SVM_RFE	PCA	KPCA	LDA	LPP (S)	LPP (H)	LDPP (S)	LDPP (H)
	SVM	Proposed (S)	0.045	0.013	0.07	0.168	0.242	0.52	0.087	0.196	0.009	0.051	0.279	0.041	0.111	0.111
	(linear)	Proposed (H)	0.008	0.002	0.013	0.074	0.012	0.01	0.018	0.003	< 0.001	0.002	0.004	0.009	0.012	0.001
	SVM	Proposed (S)	1	< 0.001	0.023	0.015	0.004	0.004	0.004	0.758	0.004	0.052	0.726	0.394	0.343	0.798
PARKINSONS	(RBF)	Proposed (H)	0.111	< 0.001	0.002	< 0.001	< 0.001	< 0.001	< 0.001	0.025	< 0.001	0.012	0.01	0.193	0.066	0.052
	DE	Proposed (S)	0.045	0.017	0.032	0.053	0.003	0.015	0.008	0.022	0.045	0.041	0.024	0.019	0.011	0.279
	KF	Proposed (H)	0.007	0.003	0.003	0.018	0.003	0.007	< 0.001	0.003	0.003	0.006	0.005	0.002	< 0.001	0.01
	ELM	Proposed (S)	0.004	0.006	< 0.001	0.008	0.032	0.004	< 0.001	< 0.001	0.006	< 0.001	0.03	0.028	0.007	< 0.001
	ELIVI	Proposed (H)	< 0.001	0.004	< 0.001	< 0.001	0.003	0.007	0.003	< 0.001	0.002	0.001	0.004	0.013	0.001	< 0.001
	SVM	Proposed (S)	0.005	0.003	0.014	0.004	0.002	0	0.002	0	0	0.004	0.004	0.002	0	< 0.001
(lii 	(linear)	Proposed (H)	0.008	0.004	0.035	0	0.005	< 0.001	0.01	0.002	0	0.016	0.007	0.001	0.002	< 0.001
	SVM	Proposed (S)	0.007	0.003	< 0.001	< 0.001	< 0.001	< 0.001	0.007	0.005	0.041	< 0.001	0.003	0.015	0.006	0.028
PSDMTSR	(RBF)	Proposed (H)	0.034	0.005	0.003	0.003	< 0.001	< 0.001	0.007	0.019	0.027	0.019	0.013	0.03	0.045	0.032
TODATION	PE	Proposed (S)	0.008	0.017	0.002	< 0.001	< 0.001	< 0.001	< 0.001	0.028	0.002	0.009	0.009	0.022	< 0.001	< 0.001
	Kr	Proposed (H)	0.002	0.013	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.004	< 0.001	0.003	< 0.001	0.022	< 0.001	< 0.001
	ELM	Proposed (S)	0.007	< 0.001	0.007	0.032	0.036	< 0.001	< 0.001	0.006	< 0.001	< 0.001	0.008	< 0.001	< 0.001	0.002
	ELIVI	Proposed (H)	0.019	< 0.001	0.006	0.014	0.007	< 0.001	< 0.001	0.009	< 0.001	< 0.001	0.013	< 0.001	0.006	< 0.001
	SVM	Proposed (S)	0.015	0.006	0.032	0.025	0.004	0.038	0.012	0.007	0.032	0.038	0.011	0.004	0.009	0.009
	(linear)	Proposed (H)	0.065	0.031	0.204	0.195	0.013	0.14	0.072	0.058	0.247	0.21	0.061	0.108	0.138	0.061
	SVM	Proposed (S)	0.016	0.087	0.042	0.063	0.087	0.108	0.087	0.006	0.147	0.027	0.004	0.066	0.044	0.147
SelfData	(RBF)	Proposed (H)	0.009	0.013	0.01	0.016	0.013	0.013	0.013	0.003	0.028	0.004	< 0.001	0.003	0.045	0.057
	DE	Proposed (S)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	Kr	Proposed (H)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.002	< 0.001	< 0.001	<0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
-	ELM	Proposed (S)	<0.001	<0.001	< 0.001	0.003	< 0.001	< 0.001	<0.001	< 0.001	0.002	< 0.001	< 0.001	< 0.001	< 0.001	<0.001
	LLIVI	Proposed (H)	< 0.001	0.002	< 0.001	0.003	< 0.001	< 0.001	< 0.001	< 0.001	0.002	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

**Table 6.** *p*-values of significance analysis between proposed and other feature-reduction methods ( $\alpha = 0.05$ ).

## 3.4. Performance on AUC and Results Visualization

For a more comprehensive performance evaluation of the proposed algorithm, the performance of each algorithm on AUC was recorded by using the PSDMTSR dataset. The results can be seen in Table 7.

Methods	Classifiers	SVM (Linear)	SVM (RBF)	RF	ELM
N	_DR	0.5542	0.6109	0.5026	0.5835
	PCA	0.5207	0.5953	0.5243	0.5543
	КРСА	0.549	0.5969	0.5163	0.4655
Feature	LDA	0.5429	0.6141	0.5408	0.5175
Extraction	LPP(S)	0.5447	0.6281	0.5343	0.5841
	LPP(H)	0.5498	0.5859	0.5131	0.5114
	LDPP(S)	0.5557	0.6234	0.4722	0.55
	LDPP(H)	0.5372	0.5422	0.5259	0.5137
	mRMR	0.5421	0.6234	0.606	0.4998
	ReliefF	0.5879	0.55	0.4775	0.5269
Feature	Pvalue	0.5474	0.5281	0.4719	0.5832
Selection	SBS	0.5534	0.5203	0.4739	0.5502
	SFS	0.5793	0.5094	0.5218	0.5359
	SVM_FRE	0.5422	0.6047	0.4864	0.4662
Dropood	Simple minded (S)	0.6655	0.6609	0.5401	0.6087
rioposed	Heat kernel (H)	0.6082	0.6859	0.5603	0.6455

Table 7. Performance of feature reduction methods on AUC.

14 of 19

It can be seen in Table 7 that the proposed method always has the highest AUC compared with existing dimensionality reduction methods used for speech data of PD, which shows that the proposed method can perform better in the classification of PD objects from speech data. It can be seen that irrespective of classifier and dataset, the proposed algorithm always gives the highest accuracy.

## 3.5. Comparison with State-of-the-Art Algorithms

In this section, the proposed algorithm was compared with some of the state-of-the-art speech feature reduction algorithms. These methods include:

**Relief-RF and Relief-SVM** [41]: In this study, four feature reduction algorithms (LASSO, mRMR, Relief, and LLBFS (local learning-based feature selection) were used. These selected features were mapped to a binary classification response using RF and SVM classifiers and the best performance was achieved on the relief feature selection method with an SVM-linear classifier. The feature subsets were selected by using a cross-validation (CV) approach (only the training set). The CV process was repeated a total of ten times and the features which appeared most frequently were selected.

**mRMR** [43]: The primary objective of this study was to compare the efficiency of feature reduction algorithms. The author used mRMR for feature selection with seven different classifiers (multilayer Perceptron, SVMs with RBF and linear kernels, logistic regression, naïve Bayes, K-nearest neighbor, and RF). Moreover, the results were combined using stacking strategies.

**LDA-NN-GA** [4]: The author divides the dataset into test and training sets using the LOSO technique. Afterward, the LDA feature extraction method was used for feature reduction of the training dataset, and then the training dataset with reduced dimension was fed to the GA-optimized BP neural network to train the model, and finally, evaluate the performance using the test dataset.

**ReliefF-FC-SVM(RBF)** [37]: This technique ranks the features using Fisher criterion (FC) based ReliefF algorithm. After that, top K features were selected for training and testing the model using the SVM-RBF classifier.

**SFFS-RF** [42]: In this study, the author used the sequential floating feature selection (SFFS) method for feature selection and RF as a classifier.

**KPCA-SVM(RBF)** [37]: For this study, a feature extraction method KPCA was used as a feature reduction method and an SVM with RBF kernel was used as a classifier.

It can be seen in Table 8 that irrespective of dataset and classifier, the proposed algorithm performed better compared with state-of-the-art algorithms. For SelfData and PSDMTSR datasets, the proposed algorithm performed better in all cases. For the PARKIN-SONS dataset, the proposed algorithm achieved higher accuracy in most of the cases. It was noticed that the state-of-the-art algorithms did not achieve the same results as described in corresponding papers. One probable reason might be the difference in experimental conditions. For instance, since the method of splitting the dataset into training, test, and validation sets were different, the number of samples used for training the model in this study was significantly lower compared to the corresponding papers. Moreover, Table 9 describes the significant analysis between proposed and state of the art methods.

## 3.6. Influence of Parameter on Accuracy

In this section, the influence of penalty-factor and dimensionality on classification accuracy is analyzed. The PARKINSONS dataset was used for this experiment with the ELM classifier. The result can be seen in Figure 3.

It can be analyzed from Figure 3 that at the beginning, with the increase of dimensionality he classification accuracy increases first and then decreases. On the other hand, at any fixed dimension, the classification accuracy reduces with the increase of penalty-factor  $\gamma$ .

Methods	<b>Datasets</b>	PSDMTSR	PARKINSONS	SelfData
DaliafE	SVM (linear)	0.5938	0.75	0.5833
Kellerr	RF	0.5313	0.725	0.5
	SVM (linear)	0.5688	0.7167	0.5433
mRMR	SVM (RBF)	0.5375	0.6667	0.6
	RF	0.6188	0.725	0.5267
SFF	S-RF	0.6063	0.8083	0.6
ReliefF-FC	-SVM(RBF)	0.6138	0.8167	0.6267
LDA-N	NN-GA	0.6138	0.8083	0.63
KPCA-S	VM(RBF)	0.55	0.675	0.5967
	SVM (linear)	0.7375	0.8167	0.6467
Proposed (S)	SVM (RBF)	0.6625	0.7583	0.6267
Tioposed (b)	RF	0.6875	0.8167	0.6733
	ELM	0.7563	0.85	0.6733
	SVM (linear)	0.7125	0.85	0.6133
Proposed (H)	SVM (RBF)	0.6625	0.8083	0.64
1000000 (11)	RF	0.7063	0.85	0.6567
	ELM	0.7625	0.8917	0.6733

 Table 8. Accuracy comparison with the state-of-the-art feature reduction methods.



**Figure 3.** The Influence of parameter  $\gamma$  and dimension on accuracy.

Dataset		Classifiam Matheda					Methods				
Dataset	Classifiers	s\Methods	ReliefF-SVM (Linear)	ReliefF-SVM (RBF)	mRMR-SVM (Linear)	mRMR-SVM (RBF)	mRMR-RF	LDA-NN-GA	ReliefF-FC-SVM (EBF)	SFFS-RF	KPCA-SVM (RBF)
	SVM	Proposed (S)	0.07	0.057	0.013	< 0.001	0.032	0.84	0.07	0.057	0.013
	(linear)	Proposed (H)	0.013	0.002	0.002	< 0.001	0.005	0.096	0.013	0.002	0.002
	SVM (PRE)	Proposed (S)	0.78	0.309	0.052	< 0.001	0.223	0.111	0.78	0.309	0.052
PARKINSONS	SVIVI (KDF)	Proposed (H)	0.173	0.023	0.007	< 0.001	0.015	1	0.173	0.023	0.007
multitoonto	PE	Proposed (S)	0.07	0.032	0.005	< 0.001	0.017	0.678	0.07	0.032	0.005
	Kr	Proposed (H)	0.013	0.003	<0.001	< 0.001	0.003	0.052	0.013	0.003	<0.001
	FIM	Proposed (S)	0.009	0.005	0.003	< 0.001	0.003	0.138	0.009	0.005	0.003
	ELIVI	Proposed (H)	0.003	0.004	<0.001	< 0.001	0.001	0.023	0.003	0.004	<0.001
_	SVM	Proposed (S)	0.014	< 0.001	0.003	< 0.001	0.004	0.003	0.024	0.027	0.003
	(linear)	Proposed (H)	0.035	< 0.001	0.004	0.002	0.034	0.027	0.052	0.045	0.004
	CVM (DDE)	Proposed (S)	0.137	0.011	0.067	0.003	0.111	0.09	0.309	0.235	0.041
PSIDMTSR	5 ( iii (iibi )	Proposed (H)	0.259	0.027	0.034	0.005	0.226	0.177	0.177	0.31	0.027
15DWI15K	DF.	Proposed (S)	0.022	0.002	0.025	< 0.001	0.017	0.01	0.075	0.045	0.016
	KF	Proposed (H)	0.014	< 0.001	0.006	< 0.001	0.013	0.007	0.044	0.029	0.005
	FIN	Proposed (S)	0.001	< 0.001	< 0.001	< 0.001	0.003	0.001	0.001	0.001	< 0.001
	ELM	Proposed (H)	0.004	< 0.001	<0.001	< 0.001	0.007	0.005	0.005	0.003	< 0.001
	SVM	Proposed (S)	0.032	< 0.001	0.006	0.055	0.003	0.504	0.425	0.396	0.086
	(linear)	Proposed (H)	0.204	< 0.001	0.031	0.509	0.01	0.475	0.565	0.8	0.44
	CVM (DDE)	Proposed (S)	0.083	< 0.001	0.005	0.087	0.011	0.343	1	0.485	0.147
SelfData	SVIVI (KDF)	Proposed (H)	0.035	< 0.001	0.004	0.013	0.002	0.496	0.373	0.358	0.028
SchData	DE	Proposed (S)	0.002	< 0.001	<0.001	< 0.001	< 0.001	0.006	0.003	0.087	0.002
	Kľ	Proposed (H)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.137	0.095	0.23	<0.001
	ELM	Proposed (S)	0.008	< 0.001	0.003	< 0.001	< 0.001	0.083	0.066	0.099	0.002
	ELM	Proposed (H)	0.006	< 0.001	< 0.001	0.003	< 0.001	0.096	0.077	0.173	0.003

**Table 9.** *p*-values of significance analysis between proposed and state-of-the-art feature-reduction methods ( $\alpha = 0.05$ ).

## 4. Discussion

Speech information processing has been widely used for the diagnosis of Parkinson's disease due to the convenience of the collection of speech data and rich information contained in it. The feature reduction methods can help to improve classification accuracy. However, since speech samples are affected by the emotional fluctuations of the speakers, samples often contain high noise and aliasing characteristics which are usually overlooked by the researchers. To alleviate these problems, we proposed a weighted hybrid feature reduction embedded with an ensemble learning technique that fully considers the above-mentioned characteristics.

The experimental results indicate that compared with existing feature reduction algorithms, the proposed method offers the highest accuracy, precision, and recall which shows the effectiveness of the proposed method. Moreover, the proposed algorithm also offers the highest G-mean which shows that it can handle imbalanced data precisely. The experiments on SelfData which evaluate the treatment of PD patients show that the classification results of the proposed algorithm on the accuracy, precision, and recall are better compared with existing feature reduction algorithms. This implies that the proposed algorithm can effectively differentiate between PD patients with or without treatment.

#### 5. Conclusions

In this study, we developed a weighted hybrid feature reduction embedded with ensemble learning algorithm and proved its applicability for the detection of patients suffering from Parkinson's disease. Our findings acknowledge that compared with existing feature reduction methods the proposed algorithm helps to attain the highest accuracy (improved 19% compared with LDPP on SelfData), precision (improved 23% compared with LDPP on SelfData using ELM), recall (improved 11% compared with LDPP on SelfData using ELM), and G-mean (improved 17% compared with LDPP on SelfData using ELM). Moreover, we believe that the proposed weighted hybrid feature reduction embedded with ensemble learning technique is truly convenient for both the patients and health organizations, whereas the clinical experts and their dedicated time are required for medical

assessment. We consider these developments as a key technical milestone that would be a step towards the computer-supported way out which enhance the convenience for voice medical treatment of a larger PD population.

Overall, this study still has substantial space for future work. The introduction of ensemble learning is a source of higher time consumption. For that reason, for the future studies, we will work to improve the processing time of the proposed method. Moreover, the comparatively limited data used in this research advises caution to generalize the existing outcomes that require further verification using new samples before this system would be used in clinical exercise.

**Author Contributions:** Methodology, W.U.R.; software, W.K.; investigation, Z.H.; project administration, N.U.; funding acquisition, F.R.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** The APC is paid by Taif University Re-searchers Supporting Project Number (TURSP-2020/331), Taif University, Taif, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors wish to acknowledge the support from Taif University Researchers Supporting Project Number (TURSP-2020/331), Taif University, Taif, Saudi Arabia.

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- 1. Babayev, R. Improving the Performance of Type-2 Diabetes Prediction Models with Automated Feature-Engineering Methods: A Design Science Research Study; Colorado Technical University: Colorado Springs, CO, USA, 2021.
- De la Fuente-Mella, H.; Rubilar, R.; Chahuán-Jiménez, K.; Leiva, V. Modeling COVID-19 cases statistically and evaluating their effect on the economy of countries. *Mathematics* 2021, 9, 1558. [CrossRef]
- 3. Velasco, H.; Laniado, H.; Toro, M.; Catano-López, A.; Leiva, V.; Lio, Y. Modeling the Risk of Infectious Diseases Transmitted by Aedes aegypti Using Survival and Aging Statistical Analysis with a Case Study in Colombia. *Mathematics* **2021**, *9*, 1488. [CrossRef]
- Ali, L.; Zhu, C.; Zhang, Z.; Liu, Y. Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network. *IEEE J. Transl. Eng. Health Med.* 2019, 7, 1–10. [CrossRef]
- Trier, Ø.D.; Jain, A.K.; Taxt, T. Feature extraction methods for character recognition-a survey. *Pattern Recognit.* 1996, 29, 641–662. [CrossRef]
- 6. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. Comput. Electr. Eng. 2014, 40, 16–28. [CrossRef]
- Rovini, E.; Maremmani, C.; Moschetti, A.; Esposito, D.; Cavallo, F. Comparative motor pre-clinical assessment in Parkinson's disease using supervised machine learning approaches. *Ann. Biomed. Eng.* 2018, 46, 2057–2068. [CrossRef]
- Sakar, C.O.; Kursun, O. Telediagnosis of Parkinson's disease using measurements of dysphonia. J. Med. Syst. 2010, 34, 591–599. [CrossRef]
- 9. Peker, M.; Sen, B.; Delen, D. Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm. *J. Healthc. Eng.* **2015**, *6*, 281–302. [CrossRef]
- 10. Benba, A.; Jilbab, A.; Hammouch, A. Hybridization of best acoustic cues for detecting persons with Parkinson's disease. In Proceedings of the 2014 Second World Conference on Complex Systems (WCCS), Agadir, Morocco, 10–12 November; pp. 622–625.
- Shirvan, R.A.; Tahami, E. Voice analysis for detecting Parkinson's disease using genetic algorithm and KNN classification method. In Proceedings of the 2011 18th Iranian Conference of Biomedical Engineering (ICBME), Tehran, Iran, 14–16 December 2011; pp. 278–283.
- Khalid, S.; Khalil, T.; Nasreen, S. A survey of feature selection and feature extraction techniques in machine learning. In Proceedings of the 2014 Science and Information Conference, London, UK, 27–29 August 2014; pp. 372–378.
- 13. Wang, X.; Paliwal, K.K. Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognit.* **2003**, *36*, 2429–2439. [CrossRef]
- 14. Chen, H.-L.; Huang, C.-C.; Yu, X.-G.; Xu, X.; Sun, X.; Wang, G.; Wang, S.-J. An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Syst. Appl.* **2013**, *40*, 263–271. [CrossRef]
- 15. Hariharan, M.; Polat, K.; Sindhu, R. A new hybrid intelligent system for accurate detection of Parkinson's disease. *Comput. Methods Programs Biomed.* **2014**, *113*, 904–913. [CrossRef] [PubMed]
- Tsanas, A.; Little, M.A.; McSharry, P.E.; Ramig, L.O. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. J. R. Soc. Interface 2011, 8, 842–855. [CrossRef] [PubMed]

- 17. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, 290, 2323–2326. [CrossRef] [PubMed]
- 18. Zhang, L.; Wang, X.; Huang, G.-B.; Liu, T.; Tan, X. Taste recognition in E-tongue using local discriminant preservation projection. *IEEE Trans. Cybern.* **2018**, *49*, 947–960. [CrossRef] [PubMed]
- 19. Yu, G.; Peng, H.; Wei, J.; Ma, Q. Enhanced locality preserving projections using robust path based similarity. *Neurocomputing* **2011**, 74, 598–605. [CrossRef]
- 20. Uzer, M.S.; Inan, O.; Yılmaz, N. A hybrid breast cancer detection system via neural network and feature selection based on SBS, SFS and PCA. *Neural Comput. Appl.* **2013**, *23*, 719–728. [CrossRef]
- 21. Ul Haq, A.; Li, J.; Memon, M.H.; Ali, Z.; Abbas, S.Z.; Nazir, S. Recognition of the Parkinson's disease using a hybrid feature selection approach. J. Intell. Fuzzy Syst. 2020, 39, 1319–1339. [CrossRef]
- Kadam, V.J.; Kurdukar, A.A.; Jadhav, S.M. An Expert Diagnosis System for Parkinson's Disease Using Bagging-Based Ensemble of Polynomial Kernel SVMs with Improved GA-SVM Features Selection. In Proceedings of the International Conference on Computational Science and Applications, Cagliari, Italy, 1–4 July 2020; pp. 227–234.
- 23. Abuhasel, K.A.; Iliyasu, A.M.; Fatichah, C. A combined AdaBoost and NEWFM technique for medical data classification. In *Information Science and Applications;* Springer: Berlin/Heidelberg, Germany, 2015; pp. 801–809.
- 24. Li, Y.; Yang, L.; Wang, P.; Zhang, C.; Xiao, J.; Zhang, Y.; Qiu, M. Classification of Parkinson's disease by decision tree based instance selection and ensemble learning algorithms. *J. Med. Imaging Health Inform.* **2017**, *7*, 444–452. [CrossRef]
- 25. Lauraitis, A.; Maskeliūnas, R.; Damaševičius, R.; Krilavičius, T. Detection of speech impairments using cepstrum, auditory spectrogram and wavelet time scattering domain features. *IEEE Access* **2020**, *8*, 96162–96172. [CrossRef]
- Guimarães, M.T.; Medeiros, A.G.; Almeida, J.S.; y Martin, M.F.; Damaševičius, R.; Maskeliūnas, R.; Mattos, C.L.C.; Rebouças Filho, P.P. An Optimized Approach to Huntington's Disease Detecting via Audio Signals Processing with Dimensionality Reduction. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.
- 27. Zhang, H.-H.; Yang, L.; Liu, Y.; Wang, P.; Yin, J.; Li, Y.; Qiu, M.; Zhu, X.; Yan, F. Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples. *Biomed. Eng. Online* **2016**, *15*, 1–22. [CrossRef]
- 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 J. Biomed. Health Inform.* 2013, 17, 828–834. [CrossRef]
- 29. Little, M.; McSharry, P.; Hunter, E.; Spielman, J.; Ramig, L. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Nat. Preced.* 2008. [CrossRef]
- 30. Boersma, P.; Van Heuven, V. Speak and unSpeak with PRAAT. Glot Int. 2001, 5, 341–347.
- 31. Rusz, J.; Tykalová, T.; Krupička, R.; Zárubová, K.; Novotný, M.; Jech, R.; Szabó, Z.; Růžička, E. Comparative analysis of speech impairment and upper limb motor dysfunction in Parkinson's disease. *J. Neural Transm.* **2017**, 124. [CrossRef]
- 32. Zhan, A.; Little, M.A.; Harris, D.A.; Abiola, S.O.; Dorsey, E.; Saria, S.; Terzis, A. High frequency remote monitoring of Parkinson's disease via smartphone: Platform overview and medication response detection. *arXiv* **2016**, arXiv:1601.00960.
- Khan, T.; Westin, J.; Dougherty, M. Classification of speech intelligibility in Parkinson's disease. *Biocybern. Biomed. Eng.* 2014, 34, 35–45. [CrossRef]
- Benba, A.; Jilbab, A.; Hammouch, A. 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. 2016, 19, 449–456. [CrossRef]
- Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: A new learning scheme of feedforward neural networks. In Proceedings of the 2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541), Budapest, Hungary, 25–29 July 2004; pp. 985–990.
- 36. Liu, Y.; Tan, X.; Li, Y.; Wang, P. Weighted Local Discriminant Preservation Projection Ensemble Algorithm With Embedded Micro-Noise. *IEEE Access* 2019, *7*, 143814–143828. [CrossRef]
- Yang, S.; Zheng, F.; Luo, X.; Cai, S.; Wu, Y.; Liu, K.; Wu, M.; Chen, J.; Krishnan, S. Effective dysphonia detection using feature dimension reduction and kernel density estimation for patients with Parkinson's disease. *PLoS ONE* 2014, 9, e88825. [CrossRef]
- El Moudden, I.; Ouzir, M.; ElBernoussi, S. Automatic speech analysis in patients with parkinson's disease using feature dimension reduction. In Proceedings of the 3rd International Conference on Mechatronics and Robotics Engineering, Paris, France, 8–12 February 2017; pp. 167–171.
- 39. El Moudden, I.; Ouzir, M.; ElBernoussi, S. Feature selection and extraction for class prediction in dysphonia measures analysis: A case study on Parkinson's disease speech rehabilitation. *Technol. Health Care* **2017**, *25*, 693–708. [CrossRef]
- 40. Lei, H.; Zhao, Y.; Wen, Y.; Luo, Q.; Cai, Y.; Liu, G.; Lei, B. Sparse feature learning for multi-class Parkinson's disease classification. *Technol. Health Care* **2018**, *26*, 193–203. [CrossRef]
- 41. Tsanas, A.; Little, M.A.; McSharry, P.E.; Spielman, J.; Ramig, L.O. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 1264–1271. [CrossRef] [PubMed]
- Galaz, Z.; Mekyska, J.; Mzourek, Z.; Smekal, Z.; Rektorova, I.; Eliasova, I.; Kostalova, M.; Mrackova, M.; Berankova, D. Prosodic analysis of neutral, stress-modified and rhymed speech in patients with Parkinson's disease. *Comput. Methods Programs Biomed.* 2016, 127, 301–317. [CrossRef]

- 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. *Appl. Soft Comput.* 2019, 74, 255–263. [CrossRef]
- 44. Cigdem, O.; Demirel, H. Performance analysis of different classification algorithms using different feature selection methods on Parkinson's disease detection. *J. Neurosci. Methods* **2018**, *309*, 81–90. [CrossRef]
- 45. Tuncer, T.; Dogan, S.; Acharya, U.R. Automated detection of Parkinson's disease using minimum average maximum tree and singular value decomposition method with vowels. *Biocybern. Biomed. Eng.* **2020**, *40*, 211–220. [CrossRef]
- 46. Kursun, O.; Gumus, E.; Sertbas, A.; Favorov, O.V. Selection of vocal features for Parkinson's Disease diagnosis. *Int. J. Data Min. Bioinform.* **2012**, *6*, 144–161. [CrossRef] [PubMed]