

Hybrid Speech Feature Learning Ensembled with Dual Stage Hierarchical Boosting for Classification of Parkinson's Disease

Zeeshan Hameed¹, Usman Rafiq², Waheed Ur Rehman³, Mahmood Ashraf⁴, Tahir Bashir⁵

^{1,2} Free University of Bozen-Bolzano, Bolzano, 39100, Italy

³Department of Mechatronics Engineering, University of Chakwal, Chakwal, Punjab, 48800, Pakistan.

⁴Department of Communication and Cybersecurity, Bahauddin Zakariya University, Multan, Punjab, 48800, Pakistan.

⁵College of Integrated Circuit Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China.

zeeshan.Hameed@unibz.it, usman.rafiq@unibz.it, waheed.urrehman@uoc.edu.pk,

mahmoodkhn24@gmail.com, tahir.bashir@njupt.edu.cn

Abstract

As a long-term and progressive neurodegenerative disease, Parkinson's disease (PD) is challenging to diagnose at premature stages. Building machine learning models based on speech data for classifying PD objects has proven to be an effective approach. However, PD speech data has highly correlated features, including noise and redundancy, which affect the performance of machine learning models. Although dimensionality reduction algorithms (feature selection and feature extraction) can solve these issues to some extent, they have their limitations. Feature selection algorithms select a subset of the original feature space, thereby losing substantial information by eliminating the features. Feature extraction methods transform the features in a new feature space without reducing the number of features; However, the algorithms suffer from the problem of high aliasing and instability while transforming the high-dimensional data. To address these issues, we propose the HBD-SFDSB Model that leverages hybrid speech feature learning and dual-stage hierarchical Boosting approaches for Accurate PD diagnosis. The main goals of HBD-SFDSB are as follows: (1) constructing a deep chain-based sample space (DCS) by an iterative extraction method. (2) developing a dual-stage feature learning by introducing a hybrid feature learning technique, and (3) performing the dual-stage feature learning recurrently employing the AdaBoost technique to acquire high-quality speech features that significantly enhance the classification accuracy of the machine learning model. The performance of the HBD-SFDSB was evaluated using a self-collected and two publicly available speech datasets. The results show that HBD-SFDSB improves classification accuracy up to 38.71% compared to available feature reduction and deep learning algorithms. Consequently, HBD-SFDSB may provide valuable insights for future research in this field.

Index Terms— Parkinson's disease; dimensionality reduction; ensemble learning; hybrid feature learning.

I. INTRODUCTION

THE Parkinson's disease (PD) affects the central nervous system, leading to a variety of symptoms, particularly motor impairments such as movement slowness, stiffness, and tremors. Additionally, non-motor symptoms, including depression, sleep disturbances, constipation, and dysarthria, also significantly affect patients' quality of life as the disease progresses [1]. Detecting PD in its early stages through speech data processing has become a key area of research, as it provides an effective tool for diagnosis and monitoring [2]. However, the PD speech datasets generally contain aliasing and noise, which complicates the analysis. A central challenge in PD research is efficiently extracting meaningful features from speech data.

Dimensionality reduction techniques aim to decrease the number of features in high-dimensional datasets, creating a new, lower-dimensional representation that retains the most critical information. These techniques are generally divided into two categories: 1) feature selection (FS) and 2) feature transformation (FT) [3]. The FS is the process of selecting a subset of relevant features from the original high-dimensional dataset that best represents the data while reducing its complexity. There are three primary approaches for FS: embedded, wrapper, and filter, which are commonly applied in data

preprocessing [4]. For instance, Rovini et al. [5] employed the P-value approach for feature selection, whereas Sakar C and Kursun O. [6] built a hybrid approach by integrating FS with the SVM classifier, achieving 92% classification accuracy. Peker et al. [7] utilized the Minimum Redundancy Maximum Relevance (mRMR) method along with Complex-Valued Artificial Neural Networks (CVANN) for selected attributes. Benba et al. [8] employed MDVP (Multi-Dimensional Voice Program) for feature selection, followed by k-nearest neighbors (KNN) and SVM for classification. Additionally, Shirvan et al. [9] investigated the combined impact of genetic algorithm and KNN for FS.

While FS methods have advantage of retaining only the most significant features for classification, they do not generate new high-quality features and can result in the loss of potentially valuable information due to feature elimination [10]. Feature transformation is a technique that reduces the dimensionality of high-dimensional data while retaining the maximum possible information from the samples [11]. A method combining fuzzy K-Nearest Neighbors (KNN) with Principal Component Analysis (PCA) for Parkinson's disease (PD) classification was proposed by Chen H et al. [12], which achieved 96% accuracy. Hariharan et al. [13] introduced an approach using PCA and LDA to recognize PD patients. While PCA and LDA

perform well, they may not be suitable for real-world scenarios where data is typically non-linear. PCA and LDA are linear approaches that presume the data lie within a linear subspace, making them inadequate for non-linear datasets [14]. To address this, non-linear variants of these techniques, such as Kernel Principal Component Analysis (KPCA) and Kernel Linear Discriminant Analysis (KLDA), have been developed [15, 16]. However, choosing the appropriate kernel function for data mapping remains a challenging task. An alternative to kernel-based methods is manifold learning, which does not require the selection of a kernel function. LPP is another form of manifold learning that preserves the neighborhood structure of data while minimizing distances between points with similar neighborhood relationships [17, 18]. However, LPP's performance is heavily reliant on the number of neighborhood data samples and faces issues with small sample sizes. Various modifications to LPP have been proposed to address these problems, although challenges remain, such as instability during high-dimensional data mapping and the focus on variance between classes rather than within classes. Additionally, partitioning small datasets can introduce significant randomness due to the differences between training and test data distributions [15, 19-21].

Recent studies have focused on hybrid systems that combine FS and FE techniques to enhance automatic PD detection [22]. Uzer M et al. [23] built a hybrid system by integrating Sequential Forward Selection (SFS) and PCA with a neural network classifier. Amin Ul Haq et al. [24] used an ant-colony optimization algorithm combined with the Relief filtering method to build a hybrid system, utilizing SVM for classification and K-fold cross-validation. These approaches highlight three primary categories of PD speech feature learning: FS, FE, and a combination of both (FSFE or FEFS). Feature selection centers on the correlation between the features, while feature extraction emphasizes retaining critical information in data. The combined FEFS approach often yields superior classification results compared to using FS or FE alone.

Furthermore, enhancing the remote diagnosis of PD necessitates both the extraction of high-quality features and the implementation of a robust classification approach. Ensemble learning, a popular strategy for enhancing classification performance, has been applied in various bioinformatics domains, including PD diagnosis. Kadam V et al. [25] employed a genetic algorithm for attribute selection and then combined a bagging approach with SVM classifiers. Abuhasel et al. [26] proposed a hybrid approach that combines the NEWFM classifier with an adaptive boosting technique to enhance the accuracy of Parkinson's disease diagnosis. Additionally, Yongming et al. [27] introduced a decision tree-based FS method that combines ensemble learning, RF, and SVM for classification.

Despite the insights from the analysis above, an approach often overlooked in research involves combining hybrid feature ensemble learning. Feature extraction techniques are effective in constructing high-quality features, but they often fail to prioritize these features based on their relevance. On the other hand, traditional feature selection methods can prioritize features based on their relevance but do not focus on generating high-quality features. These two approaches to feature reduction (FR) may appear distinct, but could complement each other. Therefore, it is essential to develop an FR

method that not only constructs high-quality features but also ensures feature prioritization based on usefulness. While some studies have made progress in this direction [28, 29], several significant issues remain unresolved. First, merely integrating feature selection and feature extraction methods does not guarantee the generation of high-quality features. This is because, without the use of Deep Chian-based Sample-space (DCS) and ensemble learning, feature extraction and selection are performed only once. Second, current approaches focus solely on the characteristics of samples in the original space and fail to consider the structural information of deeper instances. To address these challenges, combining hybrid ensemble learning with hybrid feature learning on top of DCS could significantly enhance detection performance. The primary contributions of this HBD-SFDSB model are summarized below:

- Development of a deep chain-based hierarchical sample space (DCS) to enhance the model's generalization capabilities.
- Creation of a hybrid feature reduction approach by combining manifold feature extraction techniques with nearest neighbor feature selection methods.
- Formulation of a hybrid ensemble learning model by integrating weighted-boosting with Bagging methods for effective model training.
- An iterative dual-hybrid feature reduction process that extracts high-quality features, leading to significant improvements in classification accuracy.

The structure of this paper is as follows: Section II gives a comprehensive explanation of the principles underlying the HBD-SFDSB model; Section III details the methodology, experimental setup, and results analysis; Section IV highlights the contributions and constraints of this work.

II. MATERIALS AND METHODS

A. Datasets

This study utilized the following three distinct and representative datasets to validate the excellence of the HBD-SFDSB method: the PSDMSTR [30], a self-collected dataset SelfData, and PARKINSON [31]. The PSDMSTR dataset consists of 20 normal participants and 20 Parkinson's disease (PD) patients. The PARKINSON dataset includes 8 normal individuals and 23 patients. The SelfData dataset was recorded by the Army Medical University, Chongqing, China. It includes 21 healthy participants and 10 PD patients. A total of 13 samples were collected for each participant, and each sample contained 26 features, leading to a data matrix of size 1170×26.

The PSDMSTR and PARKINSON datasets are publicly available and accessible through the University of California, Irvine (UCI) repository <https://archive.ics.uci.edu/ml/index.php>. A summary of these datasets is provided in Table I.

TABLE I
Characteristics of the datasets

Dataset	Patients	Normal People	Samples	Features	Classes	Reference
PARKINSON	23	8	195	23	2	[29]
PSDMSTR	20	20	1040	26	2	[28]
Self Data	36	54	1170	26	2	-

B. Proposed Method

1. Symbol Description

In this paper, the data matrix is defined as:

$$X = [x_1, x_2, x_3, \dots, x_N]^T \quad \dots (1)$$

$$= [X^1, X^2, X^3, \dots, X^C]^T \in R^{N \times F}$$

Where N represents the total number of data samples, F represents the number of features, and C is the total number of classes in the dataset X.

The labels of the samples are defined as:

$$y = [y_1, y_2, y_3, \dots, y_N]^T \in R^N \quad \dots (2)$$

The projection matrix is represented as:

$$M = (m_1, m_2, m_3, \dots, m_f) \in R^{F \times f} \quad \dots (3)$$

After the projection, the data space $R^{N \times f}$ is transformed into $R^{N \times f}$, where $F > f$, which is a lower-dimensional subspace.

2) Construction of Deep Chain-Based Sample Space

One of the key innovations in our research is the development of a hierarchical sample space. In this framework, data samples within each sample space are processed independently, allowing for the analysis of relationships across these spaces, which enhances the generalization capability of the model.

In this mechanism, p_j represents the set of data indices assigned to the j -th cluster. The dataset is partitioned into k clusters, denoted as $[\rho_j]_{j=1}^k$. To address the challenges posed by nonlinearly separable data, we use a radial basis kernel function ϕ .

The objective function can be represented as follows:

$$E \{ [\rho_j]_{j=1}^k \} = \sum_{j=1}^k \sum_{a \in \rho_j} \| \phi(x_i) - m_j \|^2 w(a) \quad \dots (4)$$

where: a is the data sample that belongs to the cluster p_j , x_i represents the i -th data sample in the training dataset X_{train} , $w(a)$ is the weighting factor for the sample x_i , and m_j is the center of the j -th cluster, given by:

$$m_j = \frac{\sum_{b_i \in \rho_c} w(b) \phi(x_i)}{\sum_{b_i \in \rho_c} w(b) |\rho_j|} \quad \dots (5)$$

Where b is a data (a point) that belongs to a neighbor cluster p_c . The Euclidean distance between the cluster center m_j and the data sample $\phi(x_i)$ can be expanded as follows:

$$\begin{aligned} & \left\| \phi(x_i) - \frac{\sum_{a_i \in \rho_c} \phi(x_i)}{|\rho_j|} \right\|^2 \quad \dots (6) \\ &= \frac{\phi(x_i) \cdot \phi(x_i)}{|\rho_j|} \\ &+ \frac{\sum_{x_j, x_l \in \rho_c} \phi(x_j) \phi(x_l)}{|\rho_c|} \\ &- \frac{2 \sum_{x_i \in \rho_c} \phi(x_j) \phi(x_l)}{|\rho_c|} \end{aligned}$$

The flowchart illustrating the Deep Chain-Based Sample Space (DCS) mechanism is shown in Fig. 1. This mechanism combines a mean-based clustering algorithm with a radial basis kernel function. The steps are as follows:

- The original dataset is treated as the first sample space.
- The DCS algorithm is applied to create the second sample space.
- This procedure is recursively repeated to generate the third, fourth, and subsequent hierarchical sample spaces, continuing until the n -th sample space is formed.

Each new sample space contain P% of the samples from the previous hierarchical (upper-level) sample space.

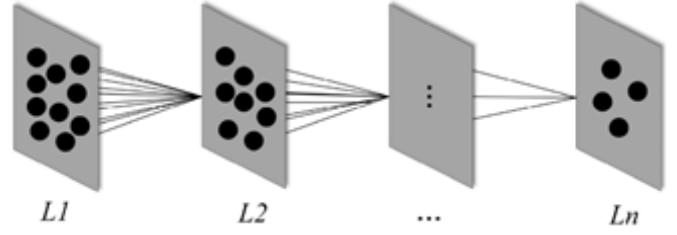


Fig. 1. Flowchart of the Deep Chain-Based Sample Space (DCS) mechanism.

3) Dual-hybrid Feature Reduction Module

The primary goal of this module is to minimize the distance between data points that belong to the same class while maximize the distance that belong to distinct classes, all while maintaining the locality of the data samples. Additionally, the module utilizes the weighted-boosting technique to highlight misclassified samples, and ensemble learning is subsequently applied to generate the final output using weight coefficients.

For parameter tuning and model validation, hold-out cross-validation is employed.

This algorithm is designed to minimize within-class variance, maximize between-class variance, and preserve neighborhood relationships during feature transformation. The methodology integrates feature reduction with a weighted hybrid-AdaBoost classification approach, utilizing decision stumps and SVM as weak and strong classifiers, respectively. The process is validated through dataset segmentation, with detailed mathematical formulations and pseudo-code provided.

4) Feature Extraction Process

A) Within-Class Variance Minimization

To make samples of the same class closest after mapping, the algorithm minimizes the within-class variance matrix. The mathematical formulation is expressed as follows:

$$\min_M \sum_{c=1}^C \left[\| M^T x'_c - M^T \bar{x}'_{w,c} \|^2 \right]_{x'_c \in X_{wc}} \quad \dots (7)$$

which can be rewritten as:

$$\min_M \sum_{c=1}^C \left[M^T (x'_c - \bar{x}'_{w,c}) (x'_c - \bar{x}'_{w,c})^T M \right]_{x'_c \in X_{wc}} \quad \dots (8)$$

or equivalently:

$$\min_M M^T S_W M \quad \dots (9)$$

Where $S_W = \sum_{c=1}^C (x'_c - \bar{x}'_{w,c}) (x'_c - \bar{x}'_{w,c})^T$ represents the within-class variance matrix, $\bar{x}'_{w,c} = \frac{1}{k_{mc}} \sum_{i=1}^{k_{mc}} x'_{i,c}$ describes the center of class c and $x'_{i,c}$ represents the i -th instance of class.

B) Between-Class Variance Maximization

To make samples of the distinct classes far apart after mapping, the algorithm increases the variance of the between-class scatter matrix. The mathematical formulation is expressed as follows:

$$\max_M \sum_{c=1}^C [\|M^T \bar{x}'_{a,c} - M^T \bar{x}_a\|^2]_{x'_{a,c}, x_a \in X} \quad \dots (10)$$

$$\max_P \sum_{c=1}^C [P^T (\bar{x}'_{w,c} - \bar{x}_a) (\bar{x}'_{w,c} - \bar{x}_a)^T P]_{x'_{a,c}, x_a \in X} \quad \dots (11)$$

or equivalently:

$$\max_P P^T S_B P \quad \dots (12)$$

Where $S_B = \sum_{c=1}^C (\bar{x}'_{w,c} - \bar{x}_a) (\bar{x}'_{w,c} - \bar{x}_a)^T$ is the inter-class variance matrix.

C) Neighborhood Preservation

The neighborhood relationships among the data samples are preserved in the process of mapping, which is expressed mathematically as:

$$\sum_{c=1}^C \sum_{j=1}^N \sum_{i=1}^{N_c} Q_{ij}^c [\|M^T x'_{i,c} - M^T x'_j\|^2]_{x'_{i,c}, x'_j \in X_{train}} \quad \dots (13)$$

$$\sum_{c=1}^C \left[M^T \left(\sum_{i=1}^{N_c} x'_{i,c} D^c (x'_{i,c})^T - \sum_{j=1}^N x'_{i,c} Q_{ij}^c (x'_j)^T \right) M \right]_{x'_{i,c}, x'_j \in X_{train}} \quad \dots (14)$$

Or:

$$\sum_{c=1}^C [M^T X'_c (D^c_{ij} - Q^c) X'_{train} M]_{x'_c \in X_{train}} \quad \dots (15)$$

Or:

$$\sum_{c=1}^C [M^T X'_c (D - Q) X'_{train} M]_{x'_c \in X_{train}} \quad \dots (16)$$

$$M X_{train} Z X_{train}^T M \quad \dots (17)$$

Where $Z = D - Q$ is a Laplacian matrix, $D^c_{ij} = \sum_j Q^c_{ij}$

shows the diagonal matrix, $D = \begin{pmatrix} D^1 & & \\ & \ddots & \\ & & D^C \end{pmatrix}$ and $Q =$

$\begin{pmatrix} Q^1 & & \\ & \ddots & \\ & & Q^C \end{pmatrix}$ represents the affinity matrix determined as follows:

$$\text{Heatkernel: } Q_{ij}^c = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}}, & \text{if } x_i^c \in N_k(x_j) \parallel x_j \in N_k(x_i^c) \\ 0, & \text{otherwise} \end{cases} \quad \dots (18)$$

$$\text{Simple-minded: } Q_{ij}^c = \begin{cases} 1, & \text{if } x_i^c \in N_k(x_j) \parallel x_j \in N_k(x_i^c) \\ 0, & \text{otherwise} \end{cases} \quad \dots (19)$$

D) Objective Function

The goal is to minimize the local variance matrix of the similar class, maximize the variance matrix of distinct classes, and preserve the neighborhood formation. The objective function can be expressed as follows:

$$\min_M \text{Tr}[M^T S_W M - M^T S_B M + X_{train} Z X_{train}^T M] \quad \dots (20)$$

which can be expressed as a Lagrange multiplier:

$$L(M, \lambda) = M^T (S_W - \lambda [\mu S_B - \gamma (X_{train} Z X_{train}^T)]) M \quad \dots (21)$$

Taking the derivative with respect to M for the optional solution:

$$\begin{aligned} \frac{\partial}{\partial M^n} [L(M, \lambda)] &= 0 \quad \dots (22) \\ \Rightarrow \frac{\partial \{M^T [S_W - \lambda [\mu S_B - \gamma (X_{train} Z X_{train}^T)]] M\}}{\partial M^n} &= 0 \end{aligned}$$

$$2S_W M - 2\lambda [\mu S_B M - \gamma (X_{train} Z X_{train}^T) M] = 0 \quad \dots (23)$$

$$S_W M = \lambda [\mu S_B - \gamma (X_{train} Z X_{train}^T)] M = 0 \quad \dots (24)$$

Or

$$\lambda M = M S_W [\mu S_B - \gamma (X_{train} Z X_{train}^T)]^{-1} \quad (25)$$

where μ and γ are penalty factors tuned via a substitute approach.

E) Feature Selection Process

After mapping the data onto M, the algorithm ranks features according to their ability to separate classes in local neighbors:

$$W(f_i) \quad (26)$$

$$\begin{aligned} &= P(\text{different value of } f_i \mid \text{distinct class}) \\ &- P(\text{different value of } f_i \mid \text{same class}) \end{aligned}$$

which can be expressed as:

$$\begin{aligned} W(f_i) &= \sum_{c \neq \text{class}(x')} \left[\frac{P(x')}{1 - P[\text{class}(x')]} \sum_{j=1}^k \frac{\text{diff}[f_i, x', M_j(x')]}{mk} \right]_{x' \in} \quad (27) \\ &- \sum_{j=1}^k \frac{\text{diff}[f_i, x', H_j(x')]}{mk} \end{aligned}$$

where $M_j(x')$ describes neighborhood data samples of different classes, $H_j(x')$ denotes neighborhood samples of the same class, and $\text{diff}[f_i, M_j(x')]$ is the variance between the M_j and the data sample x' for feature f_i . The weights are initialized as $W = (w_1, w_2, \dots, w_i) = (0, 0, \dots, 0)$. A sample x' is randomly selected from X'_{train} , and k nearest hits and misses are computed based on distance. The weights are updated over m iterations, and the top-rated features $W(f_i)_{f \in F} (i = 1, \dots, F)$ of X'_{train} are used for training.

F) Weighted Hybrid-AdaBoost Classification

A weighted hybrid-AdaBoost classification method uses decision stumps and SVM as weak and strong classifiers, respectively. The objective function is:

$$\text{sign} \left[\sum_{s=1}^S \rho_s h_s(x'_s) \right] \quad (28)$$

The classifier weight for each data sample s is given by $\rho_s = \frac{1}{2} \ln \left(\frac{1-\epsilon_s}{\epsilon_s} \right)$ where ϵ_s represents the error rate of the classifier.

The initial weights are defined as $\omega(x'_s, y_s)_{y_s \in [-1,1]} = \frac{1}{S}$, for $s = 1 \dots, S$, where S is the total number of samples. The algorithm generates a base learner $h_s(x'_s) : x'_s \rightarrow y_s$ using the training dataset, after which $h_s(x'_s)$ is validated using the validation data X_V . The weights are then updated as follows:

$$\omega_s(x'_s, y_s) = \frac{\omega_{s-1}(x'_s, y_s) \cdot e^{-\rho_s y_s h_s(x'_s)}}{J_s} \quad (29)$$

where J_s is used as normalization parameter, ensuring the sum of all weights is 1. The process iterates n time, producing classifiers ($classifier_1, \dots, classifier_n$), with the final classifier:

$$\begin{aligned} Classifier_{Final} \\ &= \alpha_1 classifier_1 \\ &+ \alpha_2 classifier_2 + \dots + \alpha_n classifier_n \end{aligned} \quad (30)$$

Subject to:

$$\sum_{r=1}^n \alpha_r = 1 \quad (31)$$

where α is calculated by grid search.

G. Workflow Description

Considering the previously mentioned details, the overall methodology of the HBD-SFDSB method unfolds as follows: Initially, the dataset (X) is split into the training set X_R , the testing set X_T , and the validation set X_V . Moreover, X_R is then arbitrarily partitioned into n sub-sets, each containing 80% of the total training data. These subsets are denoted as $X_{r1}, X_{r2}, \dots, X_{rn}$, corresponding to the total number of networks n . Each subset X_n is processed independently through the proposed Deep Chain-Based Sample Space (DCS) mechanism and subjected to the feature reduction algorithm described earlier. The validation set X_V is used to evaluate and validate each classifier trained on its respective X_{rn} subset. After all classifiers are trained and validated, their outputs are combined to produce the final result. The integration is performed through a weighted ensemble strategy that considers the performance of each classifier. Figure 2 illustrates the overall flowchart of the proposed method, while Figure 3 demonstrates the weight-boosting mechanism used during classifier training. The corresponding pseudocode is provided in Algorithm 1, which summarizes the full training and evaluation procedure.

To summarize:

- Split dataset X into training X_R , validation X_V , and testing X_T sets.
- Apply the DCS mechanism to create multiple hierarchical sample spaces (X_1, X_2, X_m) from X_R .
- Independently process each X_{rn} through the proposed feature reduction and hybrid classification procedure.
- Validate each resulting classifier using X_V .
- Integrate all classifiers using a weighted ensemble to produce the final prediction.

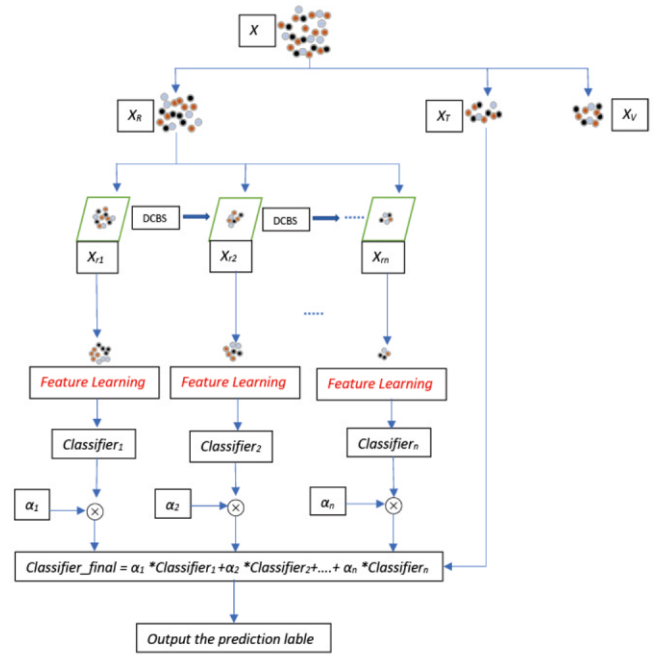


Fig. 2. Flowchart of the HBD-SFDSB model

III. RESULTS

This section provides a detailed evaluation of the HBD-SFDSB method. All experiments were carried out on a Windows 11 (64-bit) system with 16 GB of RAM. The proposed algorithm was implemented and executed using MATLAB R2024a. Moreover, each experiment was repeated 10 times to minimize the impact of randomness and irregularities. Additionally, to get optimal performance, the parameters were adjusted as outlined in Table II.

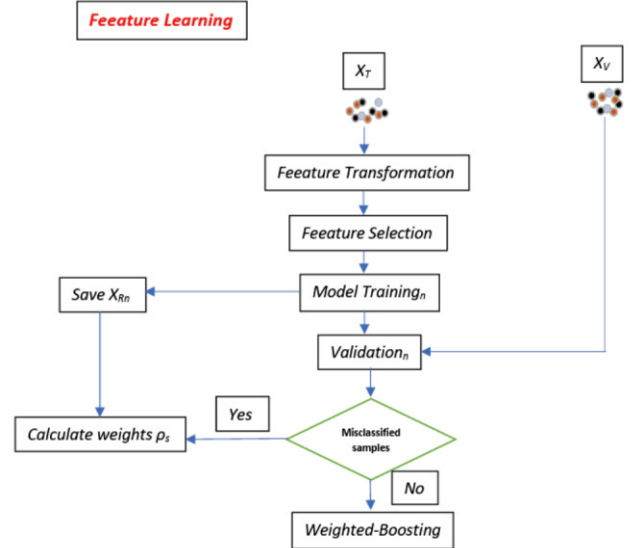


Fig. 3. Cross-validation process of the proposed model

Algorithm 1 Proposed HBD-SFDSB Framework

```

1: function MAIN( $X, Y$ )
2:   Split dataset  $X$  into  $X_R$  (train),  $X_T$  (test), and  $X_V$  (validation)
3:   for  $d = 1$  to  $n$  do ▷  $n$  is the number of sample spaces
4:      $[\alpha, \beta] \leftarrow \text{shape}(X_R)$ 
5:      $k \leftarrow \alpha \times P$  ▷ Number of clusters
6:     Define cluster space  $\mathcal{E}' \in \mathbb{R}^{k \times \mathcal{E}}$ 
7:      $\mathbb{E} \left\{ [\rho_j]_{j=1}^k \right\} = \sum_{j=1}^k \sum_{a \in \rho_j} w(a) \|\phi(x_i) - m_j\|^2$ 
8:      $L^k = \frac{1}{k} \sum_{j=1}^k a_j$  ▷ Cluster labels
9:      $X_R^n = [X_{r1}, X_{r2}, \dots, X_{rn}]$ 
10:   end for
11:   for  $i = 1$  to  $n$  do
12:     while  $\epsilon_s \leq \text{Threshold}$  do
13:        $S_W = \sum_{c=1}^C (x'_c - \bar{x}'_{w,c})^T (x'_c - \bar{x}'_{w,c})$ 
14:        $S_B = \sum_{c=1}^C (\bar{x}'_{w,c} - \bar{x}_a) (\bar{x}'_{w,c} - \bar{x}_a)^T$ 
15:        $Z = D - Q$ 
16:       Solve for projection matrix  $M$ 
17:       Initialize  $W(f_i) = 0$  for all features
18:       Compute intra-class difference  $\text{diff}[f_i, x', H_j(x')]$ 
19:       Compute inter-class difference  $\text{diff}[f_i, x', M_j(x')]$ 
20:       Compute weights  $W(f_i)$ 
21:       Select top-ranked  $f$  features out of  $F$ 
22:       Assign equal weights  $\omega(x'_s, y_s) = \frac{1}{S}$ 
23:       Train model  $\text{Model}_i$ 
24:       Validate model  $\text{Model}_i$  on  $X_V$ 
25:       Identify misclassified samples
26:        $\rho_s = \frac{1}{2} \ln \left( \frac{1 - \epsilon_s}{\epsilon_s} \right)$ 
27:        $\omega_s(x'_s, y_s) = \frac{\omega_{s-1}(x'_s, y_s) \cdot e^{-\rho_s y_s h_s(x'_s)}}{\sum_j \omega_{s-1}(x'_s, y_s) \cdot e^{-\rho_s y_s h_s(x'_s)}}$ 
28:       Add misclassified samples to training set  $X_{ri}$ 
29:     end while
30:     Apply hierarchical boosting:  $\text{sign} \left[ \sum_{s=1}^S \rho_s h_s(x'_s) \right]$ 
31:   end for
32:   Ensemble Mapping:
33:   ClassifierFinal =  $\sum_{i=1}^n \alpha_i \cdot \text{Classifier}_i$ 
34:   Subject to constraint:  $\sum_{i=1}^n \alpha_i = 1$ 
35:   return ClassifierFinal
36: end function

```

TABLE II

Symbols	Description	Values
r_n	Sample ratio	0.8
n	Subspace	10
γ	Penalty factor of $P^{n^T} \left(X_{train}^n Z X_{train}^T \right) P^n$	$10^{-4}, 10^{-3}, \dots, 10^4$
λ	Penalty factor of $P^{n^T} \left(\mu S B - \gamma \left(X_{train}^n Z X_{train}^T \right) P^n \right)$	$10^{-4}, 10^{-3}, \dots, 10^4$
μ	Penalty factor of $P^{n^T} S_\gamma P^n$	$10^{-4}, 10^{-3}, \dots, 10^4$
f	No. of features after FR	5, 10, 15, 20, ...
k	sampling for the affinity Matrix A	5
t	Kernel parameter of A	$10^{-4}, 10^{-3}, \dots, 10^4$
α	weight coefficient	Calculated by grid search

A. Evaluation Criteria

A comprehensive evaluation of the proposed algorithm is essential, and it requires well-established performance metrics. In this study, four evaluation metrics were chosen to assess the performance of the proposed algorithm: prediction accuracy (Acc), prediction correctness (Pre), recall (Rec), and the G-mean value. These metrics were derived from the confusion matrix presented in Table III, which visually represents the performance of the algorithm [18, 32].

TABLE III

Conclusion matrix representation

Real label	Prediction Labels	
	Positive (P)	Negative (N)
Positive (P)	TP	FN
Negative (N)	FP	TN

Accuracy (Acc):

$$Acc = \frac{TN+TP}{FP+FN+TP+TN} \quad (32)$$

Precision (Pre):

$$Pre = \frac{TP}{TP+FP} \quad (33)$$

Recall (Rec):

$$Rec = \frac{TP}{TP+FN} \quad (34)$$

Specificity

$$(Spe): \quad Acc = \frac{TN}{TN+FP} \quad (35)$$

G-mean:

$$G - mean = \sqrt{Rec * Spe} = \sqrt{\frac{TP}{FN+TP} + \frac{TN}{FP+TN}} \quad (36)$$

B. Benchmarking Against Representative Feature Processing Methods

This section compares the performance of our proposed HBD-SFDSB algorithm against various feature processing methods for the classification of Parkinson's disease speech data. The results are evaluated using the following assessment metrics: Accuracy (Acc), Precision (Pre), Recall (Rec), and Geometric Mean (G-mean). We analyzed the performance of the HBD-SFDSB algorithm with a traditional classifier, support vector machine (SVM), feature selection methods including mRMR, Pvalue, SVMRFE, and two classical feature extraction methods, PCA and LDA. Since deep learning acts as a feature extraction method, we also used its two representative methods, the stacked encoder (SE) and deep belief network (DBN).

Three feature selection methods, including mRMR, Pvalue, and SVMRFE, were compared with our proposed HBD-SFDSB algorithm. The results in the Table IV shows that HBD-SFDSB (H), applying the heat kernel approach, substantially outperformed all three feature selection methods for all evaluation metrics. For the Parkinson dataset, the HBD-SFDSB (H) method attained the highest classification accuracy of 96.77%, which is significantly higher compared to the best performance of SVMRFE (linear) with 82.26%. Additionally, the HBD-SFDSB (H) also outperformed in precision, recall, and G-mean, reaching 98% recall and 100% precision, indicating the robustness of the proposed method. Although SVMRFE and mRMR presented good performance in terms of precision and recall, the heat kernel-based HBD-SFDSB (H) constantly offered superior overall classification performance. Moreover, for other datasets, SelfData and PSDMTISR, the simple-minded version of HBD-SFDSB (S) method also outperformed these feature selection methods, showing its adaptability for different types of datasets.

For benchmarking against feature extraction methods, we selected two classical and two deep learning-based methods. The results are in the Table V indicates that the HBD-SFDSB (H) version constantly shows superior performance, particularly for the Parkinson dataset. The HBD-SFDSB (H) achieved an accuracy of 96.77%, outperforming the best-performing feature extraction methods, PCA (64.52%) and LDA (69.35%). Also, HBD-SFDSB (H) offered exceptional results in recall and precision, with 98% recall and 100% precision, indicating a strong capability to classify true positives correctly. For the Self Data and PSDMTISR datasets, though performance dropped marginally, HBD-SFDSB (B) yet outperformed all other feature extraction algorithms in accuracy and G-mean. These results demonstrate the substantial advantage of using

HBD-SFDSB (Fig. 1) over existing feature-learning techniques for complex feature handling and classification tasks.

This study also compares the performance of the HBD-SFDSB method with several state-of-the-art existing methods, including mRMR [3], Relief[33], a deep learning-based technique, LDA-NN-GA [34], Relief-FC-SVM [35], and SFFS-RF [36]. The results, presented in Table 5, demonstrate that HBD-SFDSB consistently outperforms the other algorithms. Notably, for the Parkinson datasets, HBD-SFDSB achieved significant improvements in classification accuracy, with an increase of 38.71%, underscoring its effectiveness. Additionally, for the SelfData and PSDMTSR datasets, HBD-SFDSB generally outperformed most of the other methods, with the largest accuracy improvement of 22.27%.

TABLE IV

Performance comparison with other feature processing techniques

Dataset	Metric	Classifier	Feature Selection			Feature Extraction				HBD-SFDSB	
			mRMR	Pvalue	SVMRFE	PCA	LDA	DBN	SE	(S)	(H)
Parkinson	Acc	SVM (L)	72.58	82.26	80.65	64.52	69.35	64.52	67.74	96.77	95.16
		SVM (RBF)	72.58	79.03	72.58	61.29	75.81	64.52	67.74	93.55	98.39
	Pre	SVM (L)	100.00	100.00	80.65	100.00	96.97	100.00	87.50	100.00	100.00
		SVM (RBF)	100.00	93.02	78.95	76.00	100.00	100.00	87.50	97.92	100.00
	Rec	SVM (L)	74.00	78.32	100.00	56.00	64.00	56.00	70.00	94.00	98.00
		SVM (RBF)	66.00	80.00	90.00	76.00	70.00	56.00	70.00	94.92	98.00
PSDMTSR	Acc	SVM (L)	86.02	88.32	00.00	74.83	76.59	74.83	63.90	97.98	96.95
		SVM (RBF)	81.24	77.46	00.00	00.00	83.67	74.83	63.90	92.83	99.99
	Pre	SVM (L)	48.08	46.47	52.56	57.05	48.40	47.00	60.26	61.22	66.35
		SVM (RBF)	56.41	56.41	55.77	56.73	53.85	47.00	64.26	60.26	61.22
	Rec	SVM (L)	47.86	45.99	56.25	63.10	47.13	46.27	64.29	69.23	72.57
		SVM (RBF)	62.82	59.09	57.38	58.27	54.76	46.27	64.29	61.11	64.00
Self Data	Acc	SVM (L)	42.95	40.38	23.08	33.97	26.28	29.81	64.15	40.38	52.86
		SVM (RBF)	31.41	41.66	44.87	47.44	44.23	29.81	64.15	56.41	51.28
	Pre	SVM (L)	47.80	46.07	43.51	52.18	43.03	44.15	58.28	57.56	64.64
		SVM (RBF)	50.57	54.45	54.69	55.96	52.98	44.15	58.58	60.13	60.45
	Rec	SVM (L)	48.25	44.76	61.14	48.25	45.45	41.26	61.54	64.34	61.54
		SVM (RBF)	47.55	45.45	51.75	45.45	45.45	41.26	61.54	56.64	66.43

IV. CONCLUSION

This study introduces the HBD-SFDSB, a novel feature processing method that combines the strengths of the FS, FX, and ensemble learning techniques. The HBD-SFDSB effectively generates high-quality features, enhancing model learning and enabling accurate diagnosis of PD. This is attained by reducing intra-class variance, increasing inter-class variance, preserving the neighborhood formation, and eliminating correlated features that introduce noise and aliasing during classification. This approach improves both classification accuracy and stability, making it a valuable tool for medical applications. Additionally, HBD-SFDSB is well-suited for small and imbalanced datasets, addressing a significant challenge in this domain. The evaluation of the proposed HBD-SFDSB method demonstrates substantial

TABLE V

Performance Comparison of HBD-SFDSB with other state-of-the-art processing algorithms (%)

Methods \ Datasets	Cassifier	Parkinson	PSDMTSR	SelfData
HBD-SFDSB (S)	SVM (L)	96.77	61.22	64.34
	SVM (RBF)	93.55	60.26	56.64
HBD-SFDSB (H)	SVM (L)	95.16	66.35	61.54
	SVM (RBF)	98.39	61.22	66.43
Relief [31]	SVM (L)	59.59	45.19	47.55
	SVM (RBF)	61.29	47.44	45.45
mRMR [3]	SVM (L)	72.58	48.08	48.25
	SVM (RBF)	95.71	56.41	48.25
LDA-NN-GA [32]	NN	68.39	58.38	63.00
Relief-FC-SVM [33]	SVM (RBF)	80.83	61.38	60.00
SFFS-RF [34]	RF	80.83	60.63	60.00

improvement in classification accuracy, outperforming existing methods across multiple publicly available PD speech datasets, as well as a dataset collected by the Army Medical University, Chongqing, China. Notably, HBD-SFDSB achieves up to 38.71% improvement in classification accuracy over other feature processing algorithms.

For future research, there are several areas for improvement. The integration of deep chain-based sample space and ensemble learning increases the algorithm's processing time. Therefore, future work will focus on optimizing the execution time of the proposed HBD-SFDSB method. Additionally, exploring the performance of the proposed method on larger and more diverse datasets is crucial. Nevertheless, as a novel framework, HBD-SFDSB distinguishes itself from other algorithms, providing valuable insights into PD speech data and setting the stage for future research advancements.

REFERENCES

- [1] M. Yang et al., "Hierarchical Boosting Dual-Stage Feature Reduction Ensemble Model for Parkinsons Disease Speech Data. Diagnostics 2021, 11, 2312," ed.
- [2] A. Tsanas, M. Little, P. McSharry, and L. Ramig, "Accurate telemonitoring of Parkinsons disease progression by non-invasive speech tests," *Nature Precedings*, pp. 11, 2009.
- [3] C. O. Sakar et al., "A comparative analysis of speech signal processing algorithms for Parkinsons disease classification and the use of the tunable Q-factor wavelet transform," *Applied Soft Computing*, vol. 74, pp. 255263, 2019.
- [4] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & electrical engineering*, vol. 40, no. 1, pp. 1628, 2014.
- [5] E. Rovini, C. Maremmani, A. Moschetti, D. Esposito, and F. Cavallo, "Comparative motor pre-clinical assessment in Parkinsons disease using supervised machine learning approaches," *Annals of biomedical engineering*, vol. 46, no. 12, pp. 20572068, 2018.
- [6] C. O. Sakar and O. Kursun, "Telediagnosis of Parkinsons disease using measurements of dysphonia," *Journal of medical systems*, vol. 34, no. 4, pp. 591599, 2010.
- [7] M. Peker, B. Şen, and D. Delen, "Computer-Aided Diagnosis of Parkinsons Disease Using Complex-Valued Neural Networks and mRMR Feature Selection Algorithm," *Journal of healthcare engineering*, vol. 6, no. 3, pp. 281302, 2015.
- [8] A. Benba, A. Jilbab, and A. Hammouch, "Hybridization of best acoustic cues for detecting persons with Parkinson's disease," in 2014 Second World Conference on Complex Systems (WCCS), 2014: IEEE, pp. 622625.
- [9] R. A. Shirvan and E. Tahami, "Voice analysis for detecting Parkinson's disease using genetic algorithm and KNN classification method," in 2011 18th Iranian conference of biomedical engineering (ICBME), 2011: IEEE, pp. 278283.
- [10] R. Babayev, "Improving the Performance of Type-2 Diabetes Prediction Models with Automated Feature-Engineering Methods: A Design Science Research Study," *Colorado Technical University*, 2021.
- [11] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in 2014 science and information conference, 2014: IEEE, pp. 372378.
- [12] H.-L. Chen et al., "An efficient diagnosis system for detection of Parkinsons disease using fuzzy k-nearest neighbor approach," *Expert*

- systems with applications, vol. 40, no. 1, pp. 263271, 2013.
- [13] M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent system for accurate detection of Parkinson's disease," *Computer methods and programs in biomedicine*, vol. 113, no. 3, pp. 904913, 2014.
- [14] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," *Journal of the royal society interface*, vol. 8, no. 59, pp. 842855, 2011.
- [15] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *science*, vol. 290, no. 5500, pp. 23232326, 2000.
- [16] Z. Hameed, H. R. Barzegar, N. El Ioini, and C. Pahl, "Robust-DSN-Performance and Fault Tolerance of a Distributed Storage Network," in *2023 Eighth International Conference on Fog and Mobile Edge Computing (FMEC)*, 2023: IEEE, pp. 115122.
- [17] G. Yu, H. Peng, J. Wei, and Q. Ma, "Enhanced locality preserving projections using robust path based similarity," *Neurocomputing*, vol. 74, no. 4, pp. 598605, 2011.
- [18] Z. Hameed, H. R. Barzegar, N. E. Ioini, and C. Pahl, "Robust-DSN: A Hybrid Distributed Replication and Encoding Network Grouped with a Distributed Swarm Workflow Scheduler," *Electronics*, vol. 13, no. 10, p. 1861, 2024. [Online]. Available:
- [19] L. Zhang, X. Wang, G.-B. Huang, T. Liu, and X. Tan, "Taste recognition in e-tongue using local discriminant preservation projection," *IEEE transactions on Cybernetics*, vol. 49, no. 3, pp. 947960, 2018.
- [20] Y. Liu, X. Tan, Y. Li, and P. Wang, "Weighted local discriminant preservation projection ensemble algorithm with embedded micro-noise," *IEEE Access*, vol. 7, pp. 143814143828, 2019.
- [21] W. Ur Rehman et al., "Motion synchronization for the SHA/EMA hybrid actuation system by using an optimization algorithm," *Automatika*, vol. 62, no. 3-4, pp. 503512, 2021.
- [22] H. Chen, Y. Lin, Y. Li, W. Wang, P. Wang, and Y. Lei, "Hybrid feature embedded sparse stacked autoencoder and manifold dimensionality reduction ensemble for mental health speech recognition," *IEEE Access*, vol. 9, pp. 2872928741, 2021.
- [23] M. S. Uzer, O. Inan, and N. Yılmaz, "A hybrid breast cancer detection system via neural network and feature selection based on SBS, SFS and PCA," *Neural Computing and Applications*, vol. 23, no. 3, pp. 719728, 2013.
- [24] A. Ul Haq et al., "Recognition of the Parkinsons disease using a hybrid feature selection approach," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 1, pp. 13191339, 2020.
- [25] V. J. Kadam, A. A. Kurdukar, and S. M. Jadhav, "An Expert Diagnosis System for Parkinsons Disease Using Bagging-Based Ensemble of Polynomial Kernel SVMs with Improved GA-SVM Features Selection," in *Proceeding of International Conference on Computational Science and Applications: ICCSA 2019, 2020*: Springer, pp. 227234.
- [26] K. A. Abuhasel, A. M. Iliyasu, and C. Faticah, "A combined AdaBoost and NEWFM technique for medical data classification," in *Information science and applications*: Springer, 2015, pp. 801809.
- [27] Y. Li et al., "Classification of Parkinson's disease by decision tree based instance selection and ensemble learning algorithms," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 2, pp. 444452, 2017.
- [28] M. Yang et al., "Hierarchical boosting dual-stage feature reduction ensemble model for Parkinsons disease speech data," *Diagnostics*, vol. 11, no. 12, p. 2312, 2021.
- [29] Z. Hameed, W. U. Rehman, W. Khan, N. Ullah, and F. R. Albogamy, "Weighted hybrid feature reduction embedded with ensemble learning for speech data of parkinsons disease," *Mathematics*, vol. 9, no. 24, p. 3172, 2021.
- [30] B. E. Sakar et al., "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE journal of biomedical and health informatics*, vol. 17, no. 4, pp. 828834, 2013.
- [31] M. Little, P. McSharry, E. Hunter, J. Spielman, and L. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinsons disease," *Nature Precedings*, pp. 11, 2008.
- [32] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Information sciences*, vol. 340, pp. 250261, 2016.
- [33] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE transactions on biomedical engineering*, vol. 59, no. 5, pp. 12641271, 2012.
- [34] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated detection of Parkinsons disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network," *IEEE journal of translational engineering in health and medicine*, vol. 7, pp. 110, 2019.
- [35] O. Cigdem and H. Demirel, "Performance analysis of different classification algorithms using different feature selection methods on Parkinson's disease detection," *Journal of neuroscience methods*, vol. 309, pp. 8190, 2018.
- [36] Z. Galaz et al., "Prosodic analysis of neutral, stress-modified and rhymed speech in patients with Parkinson's disease," *Computer methods and programs in biomedicine*, vol. 127, pp. 301317, 2016.

