Received: March 10, 2022. Revised: April 28, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

A Fast Harmonic Mean Linear Discriminant Analysis for Dimensionality Reduction

Sritha Sreedharan[1]* **Ranjith Nadarajan[1]**

[1]*Department of Computer Applications, KSG College of Arts and Science, Coimbatore, India*
*Corresponding author's Email: srithaphd.nandha@gmail.com

Abstract: Dimensionality reduction is the most prominent process in artificial intelligence and data science because of using a massive amount of high-dimensional information. In recent days, many dimensionality reduction algorithms focused on the harmonic mean-based linear discriminant analysis (HLDA) which was the enhanced version of classical LDA. In particular, two different variants such as HLDA and HLDA pairwise (HLDAp) have been applied to reduce the high-dimensional data by using the harmonic mean between-class distance. However, its computation time complexity was high during the initialization phase since it comprises the matrix Eigen decomposition/inverse. Hence this article proposes the Fast HLDA (FHLDA) and FHLDA pairwise (FHLDAp) algorithms for reducing the high-dimensional data during classification. In this algorithm, a joint diagonalization scheme is introduced instead of Eigen decomposition depending on Taylor expansion for lessening the number of iterations in the initialization step to produce the discriminant. It does not a choice to a sweeping task so that every element of the Eigenvector matrix at every iteration is calculated to minimize the computation time burden. As well, a first-order approximation of the inverse Eigenvector matrix and the complete matrix of Eigenvectors are updated at every iteration. On the contrary, the overlap among the samples of dissimilar classes tends to miscategorization. So, the optimal discriminant vector is discovered to solve this issue by extending the between-class scatter matrices. Finally, the experimental outcomes show that the FHLDA and FHLDAp algorithms achieve 10 % higher average accuracy than LDA, ALDA, WLDA, G2DLDA, HLDA, and HLDAp algorithms.

Keywords: Dimensionality reduction, High-dimensional data classification, LDA, HLDA, Eigen decomposition, Joint diagonalization, Taylor expansion

1. Introduction

As technology has been advanced, massive quantities of high-dimensional information will become more widely available in the area of artificial intelligence. This information involves high-resolution photos, data files, genomic information, and so on. Nonetheless, high-dimensional information provides challenges to the classification algorithm regarding reliability and efficacy. The curse of dimensionality is a phenomenon that happens while constructing the classifier is affected by a massive quantity of features to be decided [1]. Few learning algorithms are less prone to this issue because they never focus on selecting the feature subsets; however, rather utilize the classifier on each accessible characteristic. Even if greater the quantity of dimensions has no influence on the prediction efficacy, the computational cost may be high, so avoiding the model from being employed with the high-dimensional information for several purposes. For this perspective, dimensionality reduction algorithms have been developed for extracting relevant data and attributes which support the process of high-dimensional information.

Typically, dimensionality reduction is essential for different purposes such as machine learning, object detection, and so on [2]. Many learning algorithms have been available to perform high-dimensional information categorization. Specifically, these algorithms were utilized for minimizing the size of high-dimensional information by attribute

choice, and projection [3].

Also, such learning was used for ensuring the diversity of using classifications, random or fixed sample choice schemes like bagging, and AdaBoost. But, classical learning algorithms have a few challenges: (i) many learning algorithms enhance the diversity of classifiers in either the sample space or attribute space, and the conversion from sample to feature was not reported to enhance the diversity of classifications, (ii) many learning algorithms for high-dimensional information were achieved by instant dimensionality reduction instead of expanding the number of attributes for enhancing the diversity of classifications before reducing the dimensionality, (iii) various learning algorithms have been optimized by designing a specific objective function rather than using other unsupervised learning algorithms like density-based clustering [4].

The most well-known dimensionality reduction algorithms are the principal component analysis (PCA), LDA, and independent component analysis (ICA) [5, 6]. But, these linear algorithms cannot analyze the complicated nonlinear high-dimensional information. Standard LDA [7] has a few disadvantages: 1) they apply arithmetic mean of between-class distances and 2) they do not consider a pairwise between-class distance, so few classes may overlap with each other in subspace. To combat these problems, two different formulations of harmonic mean-based LDA: HLDA and HLDAp were developed [8] for achieving dimensionality reduction by considering the harmonic mean between-class distance. But, it has a high computation time complexity during the initialization phase which involves the matrix Eigen decomposition/inverse.

Therefore in this paper, the FHLDA and FHLDAp algorithms are proposed by using the joint diagonalization scheme depending on Taylor expansion to minimize the iterations during the initialization phase for creating the discriminant. This algorithm does not an alternative to the sweeping process so that each element of the matrix of Eigenvectors at every iteration are directly computed for reducing the computation time complexity. Also, a first-order approximation of the inverse matrix of Eigenvectors and the entire matrix of Eigenvectors is updated at every iteration. But, the overlap among the samples of different classes may tend to miscategorization. This problem is resolved by finding the optimal discriminant vector is obtained by maximizing the between-class scatter matrices.

As a result, this algorithm can minimize the computational time complexity of dimensionality reduction for high-dimensional information categorization.

The remaining sections of this paper are prepared as follows: Section II discusses the related works on dimensionality reduction in different uses. Section III explains the FHLDA and FHLDAp algorithms and section IV exhibits their efficacy. Section V concludes this work and gives the future scope.

2. Literature survey

The analysis of dimension reduction using the deep convolutional neural network (DCNN) and improved PCA [9] was proposed. Still, it needs to automatically optimize the number of variables to improve efficiency. Also, it does not operate appropriately and the correlation between the variables was poor.

A new theoretical model of group-oriented optimized multiple kernel learning was developed [10] to reduce the dimensionality depending on the cooperative representation categorization. But, it has a high computational difficulty and less robustness while the training image distribution was varied from the testing images.

A wide hierarchical sub-network-based neural network (Wi-HSNN) with entrance SNN (En-SNN) and exit SNN (Ex-SNN) was presented [11] for reducing the dimensionality of high-level features and categorizing the food patterns. Still, the accuracy was not effective. As well, it was not able to encode the raw features and create the latent space in an unsupervised manner.

An auto-encoder-based dimensionality reduction and categorization were designed [12] based on CNN for hyperspectral images. But, it degrades the efficiency since it creates blur images while increasing the number of images. An enhanced hybrid-graph discriminant learning (EHGDL) was designed [13] for reducing the dimensionality and categorization of hyperspectral images. But, it has a high computational difficulty and running time.

A new hybrid dimensionality reduction forest (HDRF) was [14] designed based on ensemble pruning for increasing the diversity of a combined model from feature and sample space. But, it needs to develop an adaptive scheme for optimizing the efficiency and pruning rate of the ensemble forest algorithm.

An adapted LDA (ALDA) [15] was developed with parameter selection to categorize high-dimensional medicinal data. But, it needs to enhance the efficiency by using various covariance matrices

for the K groups.

An enhanced joint dimension reduction and dictionary learning model was developed [16] using an auto-encoder to categorize the high-dimensional information. But, it was sensitive to the outliers because -normalization was less robust to outliers. The robustness was also less while considering the class imbalanced data distribution.

A novel dimensionality reduction technique called auto-weighted LDA (WLDA) [17], which learns the similarity matrix and updates in the subspace concurrently so that the neighborhoods were analyzed in the optimal subspaces rather than the actual space. In addition, an iterative re-weighted optimization strategy was used to assign a small weight to the pairwise points with large distances and vice versa. But, it didn't deal with large-scale unlabeled datasets.

A generalized 2D LDA (G2DLDA) was presented [18], which uses a random L-norm to estimate the between-class and within-class scatter. Also, an effective learning method was developed to solve a series of convex problems with closed-form solutions. However, it does not consider nonlinear scenario.

3. Scientific contribution of the research

The above-studied LDA variants have a few challenges: i) they consider arithmetic mean of between-class distances, and ii) they do not consider the pairwise between-class distance and so few labels may overlap with every other in subspace. Partitioning every label from the overall average does not ensure each pairwise class is partitioned. Utilizing the arithmetic mean of pairwise between-class distances is similar to using distances between every class mean and overall mean. Also, the challenges of arithmetic mean-based between-class distance occurred in several other formulations of LDA, namely, HLDA [8] and HLDAp [8].

These challenges are solved by proposing the FHLDA algorithm which considers the harmonic mean-based pairwise between-class distance while reducing the dimensionality of instances. It can increase the classification accuracy while increasing the subspace dimension due to the consideration of pairwise between-class distance and the utilization of harmonic (arithmetic) mean of between-class distances. If the subspace dimension reaches few values, then the regular increase of subspace dimension for the class overlap issue is solved by the Taylor expansion-based Eigen decomposition and therefore the classification accuracy is increased. The details of the FHLDA and FHLDA-pairwise (FHLDAp) are presented below section.

4. Proposed methodology

In this section, the FHLDA and FHLDAp algorithms are explained briefly. The schematic overview of these algorithms for dimensionality reduction is illustrated in Fig. 1.

4.1 Fast harmonic linear discriminant analysis (FHLDA)

Consider $\mathcal{X} \in \mathbb{R}^{p \times n}$ is the data matrix and $\mathcal{X} = (x_1, \ldots, x_n)$ where p refers to the data dimension and n refers to the number of data samples. Also, k is the class index, c is the required subspace dimension, and K is the overall amount of classes. Table 1 presents the notations used in this study.

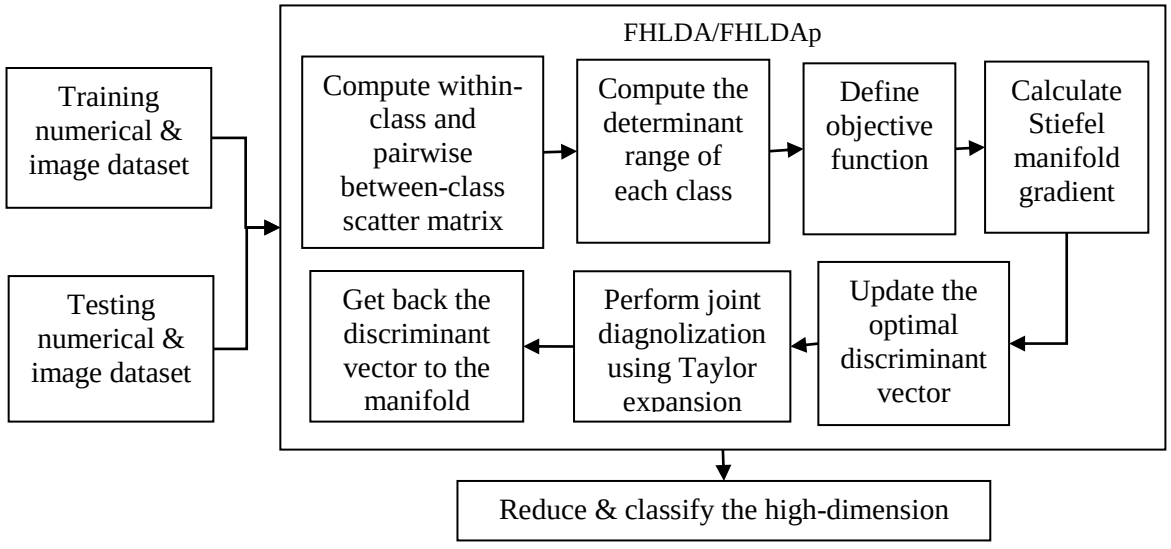

Figure. 1 Schematic overview of dimensionality reduction using proposed algorithm

Table 1. Notations used in this study

$\mathcal{X}$	Data matrix
$\mathcal{p}$	Data dimension
n	Number of data samples
$\mathcal{k}$	Class index
c	Required subspace dimension
K	Overall amount of classes
$\mathcal{G}$	Conversion matrix to a c-dimensional subspace
$\mathcal{S}_{\mathcal{b}}$	Between-class scatter matrix
$\mathcal{S}_t$	Overall scatter matrix
$n_{\mathcal{k}}$	Amount of samples in $\mathcal{k}$
$x_i \in \mathcal{k}$	Data x_i belongs to $\mathcal{k}$
$m_{\mathcal{k}}$	Average of $\mathcal{k}$
m	Average of the whole dataset
$\mathcal{B}_{\mathcal{k}\ell}$	Between-class scatter matrix for $\mathcal{k}$ and ℓ
$(\alpha'_{\mathcal{k}}, \beta'_{\mathcal{k}})$	New discriminant range of $\mathcal{k}$
η	Step size
$\mathcal{J}_1$	Objective function
$\mathcal{A}, \mathcal{U}$	Eigenvectors and diagonal matrices
$\hat{\mathcal{A}}$	Predicted matrix
t	Iteration
$\mathcal{Q}_t$	Matrix
δ	Number of iterations to reach the convergence rapidly
$\mathcal{Z}$	Updating matrix
$\mathcal{S}_w$	Global mean of within-class

Assume $\mathcal{G} \in \mathbb{R}^{\mathcal{p}\times c}$ is the conversion matrix to a c-dimensional subspace. The between-class scatter matrix ($\mathcal{S}_{\mathcal{b}}$), within-class scatter matrix ($\mathcal{S}_w$), and the overall scatter matrix ($\mathcal{S}_t$) are described as follows:

$$\mathcal{S}_t = \mathcal{S}_{\mathcal{b}} + \mathcal{S}_w \tag{1}$$

$$\mathcal{S}_{\mathcal{b}} = \sum_{\mathcal{k}=1}^{K} n_{\mathcal{k}}(m_{\mathcal{k}} - m)(m_{\mathcal{k}} - m)^T \tag{2}$$

$$\mathcal{S}_w = \sum_{\mathcal{k}=1}^{K} n_{\mathcal{k}} \mathcal{W}_{\mathcal{k}} \tag{3}$$

$$\mathcal{W}_{\mathcal{k}} = \frac{1}{n_{\mathcal{k}}} \sum_{x_i \in \mathcal{k}} (x_i - m_{\mathcal{k}})(x_i - m_{\mathcal{k}})^T \tag{4}$$

In Eqs. (2)-(4), $n_{\mathcal{k}}$ refers to the number of samples in $\mathcal{k}$, $x_i \in \mathcal{k}$ refers to the data x_i belongs to $\mathcal{k}$, $m_{\mathcal{k}}$ refers to the average of $\mathcal{k}$, and m refers to the average of the whole dataset.

$$m_{\mathcal{k}} = \frac{\sum_{x_i \in \mathcal{k}} x_i}{n_{\mathcal{k}}}, m = \frac{\sum_{i=1}^{n} x_i}{n} \tag{5}$$

A pairwise between-class scatter matrix $\mathcal{B}_{\mathcal{k}\ell}$ for $\mathcal{k}$ and ℓ as follows:

$$\mathcal{B}_{\mathcal{k}\ell} = (m_{\mathcal{k}} - m_{\ell})(m_{\mathcal{k}} - m_{\ell})^T \tag{6}$$

In this FHLDA, the discriminant range of every class is determined to obtain the optimal $\mathcal{G}$. If the new discriminant range of $\mathcal{k}$ is represented by $(\alpha'_{\mathcal{k}}, \beta'_{\mathcal{k}})$, then for given 2 classes $\mathcal{k}$ and ℓ, their discriminant ranges can satisfy the below criteria:

$$(\alpha'_{\mathcal{k}}, \beta'_{\mathcal{k}}) \cap (\alpha'_{\ell}, \beta'_{\ell}) = \phi, 1 \le \mathcal{k} < l \le K \tag{7}$$

So, the new linear discriminant criterion is defined. For any given data x, if $z = \mathcal{G}^T x$, then,

$$x \in \begin{cases} \mathcal{k}, & z \in (\alpha'_{\mathcal{k}}, \beta'_{\mathcal{k}}) \\ Subsample, & Or\ else \end{cases} \tag{8}$$

Then, the objective factor of HLDA is defined as:

$$\min_{\mathcal{G}} \mathcal{J}_1(\mathcal{G}) = \sum_{\mathcal{k}<l} n_{\mathcal{k}} n_{\ell} \frac{Tr(\mathcal{G}^T \mathcal{S}_w \mathcal{G})}{Tr(\mathcal{G}^T \mathcal{B}_{\mathcal{k}\ell} \mathcal{G})}, \quad \text{s.t. } \mathcal{G}^T \mathcal{G} = I \tag{9}$$

The gradient of Eq. (9) is provided as:

$$\nabla \mathcal{J}_1 \triangleq \frac{\partial \mathcal{J}_1}{\partial \mathcal{G}} = 2 \sum_{\mathcal{k}<l} n_{\mathcal{k}} n_{\ell} \frac{\mathcal{S}_w \mathcal{G}}{Tr(\mathcal{G}^T \mathcal{B}_{\mathcal{k}\ell} \mathcal{G})} - 2 \sum_{\mathcal{k}<l} n_{\mathcal{k}} n_{\ell} \mathcal{B}_{\mathcal{k}\ell} \mathcal{G} \frac{Tr(\mathcal{G}^T \mathcal{S}_w \mathcal{G})}{\left(Tr(\mathcal{G}^T \mathcal{B}_{\mathcal{k}\ell} \mathcal{G})\right)^2} \tag{10}$$

Restraint $\mathcal{G}^T \mathcal{G} = I$ enforces $\mathcal{G}$ on the Stiefel manifold. The differences of $\mathcal{G}$ on this manifold is parallel converse which provides few limits to the gradient. The gradient which reserves the manifold pattern is,

$$\nabla \mathcal{J}_1 - \mathcal{G}[\nabla \mathcal{J}_1]^{\mathrm{T}} \mathcal{G} \tag{11}$$

So, the new $\mathcal{G}$ is determined as:

$$\mathcal{G} \leftarrow \mathcal{G} - \eta\left(\nabla \mathcal{J}_1 - \mathcal{G}[\nabla \mathcal{J}_1]^{\mathrm{T}} \mathcal{G}\right) \tag{12}$$

In Eq. (12), η refers to the step size. Because of the truth that the manifold preserving gradient of Eq. (11) only enforces the restraint $\mathcal{G}^T \mathcal{G} = I$ to first-order at each iteration, $\mathcal{G}$ is get back to the manifold using the joint diagonalization scheme. The principle of this scheme is to diagonalize a group of $\mathcal{U}$ non-defective matrices as:

$$\mathcal{G}^{(u)} = \mathcal{A} \mathcal{D}^{(u)} \mathcal{A}^{-1}, \ \forall u = 1, \ldots, \mathcal{U} \tag{13}$$

In Eq. (13), the matrix of Eigenvectors $\mathcal{A}$ and $\mathcal{U}$ diagonal matrices $\mathcal{D}^{(u)}$ are unknown. Hence, assuming that all matrices ($\mathcal{G}$) of size $N \times N$. It is a necessity to discover a predicted matrix $\hat{\mathcal{A}}$ such that, for each matrix $\mathcal{G}^{(u)}$, $\hat{\mathcal{A}}^{-1} \mathcal{G}^{(u)} \hat{\mathcal{A}}$ are as diagonal as

promising. The matrix $\hat{\mathcal{A}}$ is predicted iteratively by successive updates. At every iteration t, a matrix $\mathcal{Q}_t$ is determined for reducing the diagonalization criterion and the matrix set is updated as:

$$\mathcal{T}_{t+1}^{(u)} = \mathcal{Q}_t^{-1}\mathcal{T}_t^{(u)}\mathcal{Q}_t,\ \forall t = 1, \dots, \delta, \forall u = 1, \dots, \mathcal{U} \quad (14)$$

In Eq. (14), $\mathcal{T}_1^{(u)} = \mathcal{G}^{(u)}$, and δ is the number of iterations to reach the convergence rapidly. So, if δ iterations enable to perform the joint diagonalization, then $\hat{\mathcal{A}}$ can be equivalent to $\prod_{t=1}^{\delta}\mathcal{Q}_t$, and each $\mathcal{T}_\delta^{(u)}$ matrices are diagonal. At every iteration, the condition depending on a quadratic measure of diagonality is given as:

$$C(\mathcal{Q}_t) = \sum_{u=1}^{\mathcal{U}} \left\| Zdiag\left\{\mathcal{Q}_t^{-1}\mathcal{T}_t^{(u)}\mathcal{Q}_t\right\}\right\|^2 \quad (15)$$

It is vital to compute N^2 elements of the matrix $\mathcal{Q}_t$ at every iteration. In this FHLDA algorithm, $\mathcal{Q}_t$ is decomposed as:

$$\mathcal{Q}_t = (I + \mathcal{Z}_t) \quad (16)$$

In Eq. (16), $\mathcal{Z}_t$ is equivalent to $Zdiag\{\mathcal{Q}_t\}$. Therefore, the updated data matrices at all iterations become:

$$\mathcal{T}_{t+1}^{(u)} = (I + \mathcal{Z}_t)\mathcal{T}_t^{(u)}(I + \mathcal{Z}_t),\ \forall u = 1, \dots, \mathcal{U} \quad (17)$$

So that the condition (15) focuses only on $\mathcal{Z}_t$, and is written as:

$$C(\mathcal{Q}_t) = \tilde{C}(\mathcal{Z}_t) = \sum_{u=1}^{\mathcal{U}} \left\| Zdiag\left\{(I + \mathcal{Z}_t)\mathcal{T}_t^{(u)}(I + \mathcal{Z}_t)\right\}\right\|^2 \quad (18)$$

For simplifying the notations, the index t is neglected by reducing the past condition. It is needed to estimate the condition while performing on an analytic form of $\mathcal{Z}$. For this reason, consider that it is nearer to the solution and so that $\|\mathcal{Z}\| \ll 1$, then its first-order Taylor expansion is calculated which gives:

$$\begin{aligned}(I + \mathcal{Z})^{-1}\mathcal{T}^{(u)}(I + \mathcal{Z}) &\approx (I - \mathcal{Z})\mathcal{T}^{(u)}(I + \mathcal{Z}) \\ &\approx \mathcal{T}^{(u)} - \mathcal{Z}\mathcal{T}^{(u)} + \mathcal{T}^{(u)}\mathcal{Z} - \mathcal{Z}\mathcal{T}^{(u)}\mathcal{Z} \\ &\approx \mathcal{T}^{(u)} - \mathcal{Z}\mathcal{T}^{(u)} + \mathcal{T}^{(u)}\mathcal{Z}\end{aligned} \quad (19)$$

Further, each matrix $\mathcal{T}^{(u)}$ is decomposed as:

$$\mathcal{T}^{(u)} = \Lambda^{(u)} + \mathcal{O}^{(u)} \quad (20)$$

In Eq. (20), $\Lambda^{(u)} = diag\{\mathcal{T}^{(u)}\}$, and $\mathcal{O}^{(u)} = \mathcal{Z}diag\{\mathcal{T}^{(u)}\}$. Also, assume that it is nearer to the solution, each $\mathcal{T}^{(u)}$ matrix is almost diagonal and $\|\mathcal{O}^{(u)}\| \ll 1$. This second estimation yields:

$$\begin{aligned}\mathcal{T}^{(u)} - \mathcal{Z}\mathcal{T}^{(u)} + \mathcal{T}^{(u)}\mathcal{Z} \approx \\ \Lambda^{(u)} + \mathcal{O}^{(u)} - \mathcal{Z}\Lambda^{(u)} + \Lambda^{(u)}\mathcal{Z}\end{aligned} \quad (21)$$

Therefore, an alternative to the below estimated joint diagonalization condition is defined as:

$$C_a(\mathcal{Z}) = \sum_{u=1}^{\mathcal{U}} \left\| \mathcal{Z}diag\{\mathcal{O}^{(u)} - \mathcal{Z}\Lambda^{(u)} + \Lambda^{(u)}\mathcal{Z}\}\right\|^2 \approx \tilde{C}(\mathcal{Z}) \quad (22)$$

Eq. (22) can be rewritten as:

$$\begin{aligned}C_a(\mathcal{Z}) = \sum_{u=1}^{\mathcal{U}}\sum_{\substack{m,n=1 \\ m\neq n}}^{N}\left(\mathcal{O}_{mn}^{(u)} + \mathcal{Z}_{mn}\Lambda_{mm}^{(u)} - \right. \\ \left.\mathcal{Z}_{mn}\Lambda_{nn}^{(u)}\right)^2 = \sum_{\substack{m,n=1 \\ m\neq n}}^{N} f(\mathcal{Z}_{mn})\end{aligned} \quad (23)$$

Where

$$f(\mathcal{Z}_{mn}) = \sum_{u=1}^{\mathcal{U}}\left(\mathcal{O}_{mn}^{(u)} + \left(\Lambda_{mm}^{(u)} - \Lambda_{nn}^{(u)}\right)\mathcal{Z}_{mn}\right)^2 \quad (24)$$

The minimization of $C_a(\mathcal{Z})$ is similar to the minimization of each $f(\mathcal{Z}_{mn})$ separately. To obtain the elements of the updating matrix $\mathcal{Z}$,

$$\begin{aligned}\frac{\partial f(\mathcal{Z}_{mn})}{\partial \mathcal{Z}_{mn}} = 2\sum_{u=1}^{\mathcal{U}}\left(\mathcal{O}_{mn}^{(u)} + \left(\Lambda_{mm}^{(u)} - \right.\right. \\ \left.\left.\Lambda_{nn}^{(u)}\right)\mathcal{Z}_{mn}\right)\left(\Lambda_{mm}^{(u)} - \Lambda_{nn}^{(u)}\right), \forall m, n; m \neq n\end{aligned} \quad (25)$$

At last, $\mathcal{Z}_{mn}$ is deduced as:

$$\mathcal{Z}_{mn} = -\frac{\sum_{u=1}^{\mathcal{U}}\mathcal{O}_{mn}^{(u)}\left(\Lambda_{mm}^{(u)} - \Lambda_{nn}^{(u)}\right)}{\sum_{u=1}^{\mathcal{U}}\left(\Lambda_{mm}^{(u)} - \Lambda_{nn}^{(u)}\right)^2}, \forall m, n; m \neq n \quad (26)$$

Observe that each iteration comprises only one update of matrices $\hat{\mathcal{A}}$ and $\mathcal{T}^{(u)}$. The termination criterion is given as:

$$\frac{|C(\mathcal{Q}_t) - C(\mathcal{Q}_{t-1})|}{C(\mathcal{Q}_{t-1})} \leq 10^{-6} \quad (27)$$

Thus, this scheme gets back $\mathcal{G}$ to the manifold with the reduced computational time complexity. Algorithm 1 describes the steps to solve Eq. (9). The objective is optimized in an iterative manner.

Besides, by extending this FHLDA to multi-class categorization issues, it is guaranteed that the efficiency of classification is improved by reducing the high-dimensional data.

4.2 Fast harmonic linear discriminant analysis pairwise (FHLDAp)

In several datasets, different classes have different within-class covariance and so the global mean of within-class $\mathcal{S}_w$ utilized in Eq. (9) would vary considerably from every class. But, the mean of 2 classes is possible to be nearer to every 2 classes. So, the utilization of the global mean of within-class distances (variances) of each class is a weak representation than the utilization of the mean of 2 class covariances. This pairwise mean is combined with the FHLDA of Eq. (9). To do this, the pairwise within-class covariance (scatter matrix) of $\hbar$ and ℓ is defined as:

$$\mathcal{W}_{\hbar l} = \frac{1}{n_\hbar + n_\ell}(n_\hbar \mathcal{W}_\hbar + n_\ell \mathcal{W}_l) \tag{28}$$

In Eq. (28), $\mathcal{W}_\hbar$ and $\mathcal{W}_l$ are given in Eq. (4). Then, the objective factor of HLDA is modified to:

$$\min_{\mathcal{G}} \mathcal{J}_2(\mathcal{G}) = \sum_{\hbar<l} n_\hbar n_\ell \frac{Tr(\mathcal{G}^T \mathcal{W}_{\hbar l}\mathcal{G})}{Tr(\mathcal{G}^T \mathcal{B}_{\hbar\ell}\mathcal{G})}, \text{ s.t. } \mathcal{G}^T\mathcal{G} = I \tag{29}$$

Here, the restraint $\mathcal{G}^T\mathcal{G} = I$ guarantees the columns of solution $\mathcal{G}$ are linearly independent. This Eq. (29) is called FHLDAp. Also, the Stiefel gradient descent scheme is used for solving the minimization dilemma. The gradient of Eq. (29) is:

Algorithm 1: FHLDA

Input: Data matrix $\mathcal{X} \in \mathbb{R}^{\mathcal{p}\times n}$ having n data samples in $\mathcal{p}$-dimensional space; class indicator matrix $\mathcal{Y} \in \mathbb{R}^{n\times K}$, K denotes the number of classes; subspace dimension c

Output: Projection matrix $\mathcal{G} \in \mathbb{R}^{\mathcal{p}\times c}$

Initialize $\mathcal{G}$;

Determine $\mathcal{S}_w$ and $\mathcal{B}_{\hbar\ell}$ using Eq. (3) & (6);

Determine the discriminant range of every class represented by $(\alpha'_\hbar, \beta'_\hbar), i = 1, \dots, K$;

while$(objective\ function\ Eq.(9)\ not\ converge)$

Determine the Stiefel manifold gradient using Eq. (11);

Update $\mathcal{G}$ using Eq. (12);

Get back $\mathcal{G}$ to the manifold using the joint diagonalization;

Execute ***Algorithm 2***;

end while

Algorithm 2: Joint diagonalization using Taylor expansion

Begin

Get all matrices $(\mathcal{G})$ of size $N \times N$;

Define the termination criterion (ϵ) and the maximum number of iterations δ;

Initialize $\hat{\mathcal{A}}$;

$t = 1$;

while$(\epsilon\ is\ not\ satisfied\ \&\ t \le \delta)$

for$(m = 1{:}N)$

for$(n = 1{:}N)$

if$(m \ne n)$

Calculate $\mathcal{Z}_{mn}$ using Eq. (26);

end if

end for

end for

Determine $\mathcal{Q} = I + \mathcal{Z}$;

Determine $\mathcal{Q}^{-1}$;

for$(u = 1{:}\mathcal{U})$

$\mathcal{T}^{(u)} = \mathcal{Q}^{-1}\mathcal{T}^{(u)}\mathcal{Q}$;

end for

$\hat{\mathcal{A}} \leftarrow \hat{\mathcal{A}}\mathcal{Q}$;

$t = t + 1$;

Update ϵ;

end while

Algorithm 3: FHLDAp

Input: Data matrix $\mathcal{X} \in \mathbb{R}^{\mathcal{p}\times n}$ having n data samples in $\mathcal{p}$-dimensional space; class indicator matrix $\mathcal{Y} \in \mathbb{R}^{n\times K}$, K denotes the number of classes; subspace dimension c

Output: Projection matrix $\mathcal{G} \in \mathbb{R}^{\mathcal{p}\times c}$

Initialize $\mathcal{G}$;

Determine $\mathcal{B}_{\hbar\ell}$ and $\mathcal{W}_{\hbar l}$ using Eq. (6) & (28);

Determine the discriminant range of every class represented by $(\alpha'_\hbar, \beta'_\hbar), i = 1, \dots, K$;

while$(objective\ function\ Eq.(29)\ not\ converge)$

Determine the Stiefel manifold gradient using Eq. (11);

Update $\mathcal{G}$ using Eq. (12);

Get back $\mathcal{G}$ to the manifold using the joint diagonalization;

Execute ***Algorithm 2***;

end while

$$\nabla\mathcal{J}_2 \triangleq \frac{\partial \mathcal{J}_2}{\partial \mathcal{G}} =$$

$$\sum_{\hbar<l} 2n_\hbar n_\ell \left[\frac{\mathcal{W}_{\hbar l}\mathcal{G}}{Tr(\mathcal{G}^T\mathcal{B}_{\hbar\ell}\mathcal{G})} - \mathcal{B}_{\hbar\ell}\mathcal{G}\frac{Tr(\mathcal{G}^T\mathcal{W}_{\hbar l}\mathcal{G})}{\left(Tr(\mathcal{G}^T\mathcal{B}_{\hbar\ell}\mathcal{G})\right)^2}\right] \tag{30}$$

After, the natural gradient of Eqs. (11) & (12) are used to enforce $\mathcal{G}$ on the stiefel manifold.

Table 2. Data attributes in different datasets

Data	Dimension p	Sample Number n	Class Number K
PIE	1024	680	68
UMIST	644	360	20
MediaMill	120	6601	74
Barcelona	48	139	4

5. Experimental results

In this section, the efficiency of FHLDA and FHLDAp algorithms for dimensionality reduction is analyzed by executing them in MATLAB 2017b and compared with the classical algorithms: standard LDA, ALDA [15], WLDA [17], G2DLDA [18], HLDA [8], and HLDAp [8]. The analysis is performed on both single-label and multi-label classification challenges. In this experiment, 2 single-label and 2 multi-label datasets are used. Table 2 summarizes the data attributes. The 2 single-label datasets are PIE and UMIST whereas the 2 multi-label datasets are Barcelona and Mediamill.

The PIE [19] is a face dataset from carnegie mellon robotics institute. Totally, it involves 68 various people and 10 images for every people with various poses, various illumination criteria, and various expressions. Images were resized to 32×32 (1024 pixels). UMIST [20] is a dataset of 360 face images (sheffield face dataset) taken from 20 people of diverse race, gender, and look. Every people contain 18 images which are resized to 28×23 (644 pixels or dimensions). MediaMill dataset [21] comprises multi-label data from video object recognition issues. It contains 74 classes and 6601 samples. Barcelona dataset [22] has image moments of 139 images with 4 classes: buildings, flora, humans, and sky. Every image has a minimum of two classes.

The comparison analysis for both single-label and multi-label classification is conducted in terms of precision, recall, f-measure and accuracy.

- Precision is the percentage of exactly classified classes at true positive (TP) and false positive (FP) rates.

$$Precision = \frac{TP}{TP+FP} = \frac{No.of\ exactly\ classified\ classes}{No.of\ exactly\ classified\ classes+ No.of\ inexactly\ classified\ classes}$$

- Recall is the percentage of exactly classified classes at TP and false negative (FN) rates.

$$Recall = \frac{TP}{TP+FN} = \frac{No.of\ exactly\ classified\ classes}{No.of\ exactly\ classified\ classes+ No.of\ inexactly\ classified\ classes}$$

- $Accuracy$ is the percentage between an exact classification of classes and the overall amount of tests performed.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

TP is a solution where the classifier classifies the classes as themselves e.g., buildings are classified as buildings. True negative (TN) is a solution where the classifier classifies the buildings as the sky or any other classes. FP is a solution where the classifier inexactly classifies the flora as buildings or any other classes. FN is a solution where the classifier inexactly classifies the sky as any other type of class.

5.1 Single-label classification analysis

Fig. 2 depicts the precision results for single-label classification using different algorithms executed on PIE and UMIST datasets. It indicates that the FHLDA and FHLDAp algorithms attain a higher precision than all the other algorithms. For the PIE dataset, the precision of FHLDA and FHLDAp is 86.1 % and 86.72 %, respectively, which are greater than the precision ranges of all other algorithms for single-label classification. Similarly, for the UMIST dataset, the precision of FHLDA and FHLDAp is 85.54 % and 86.1 %, accordingly, which are higher than all other algorithms.

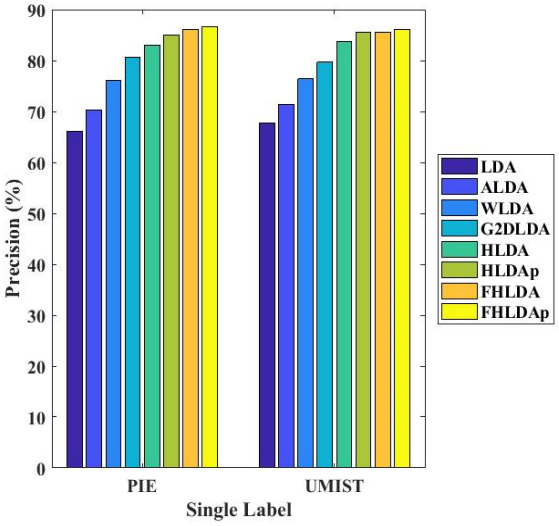

Figure. 2 Comparison of precision for single-label classification

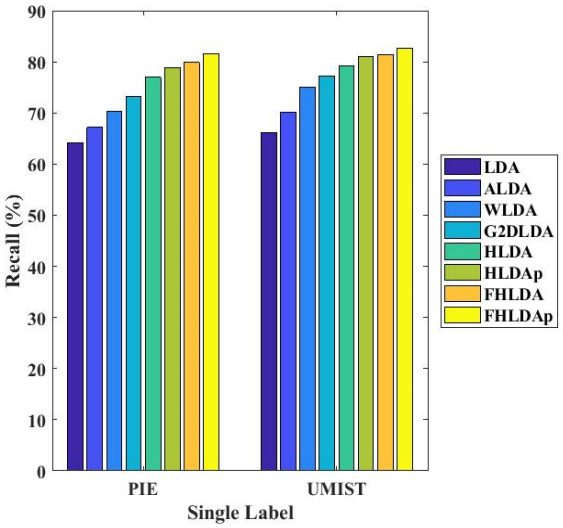

Figure. 3 Comparison of recall for single-label classification

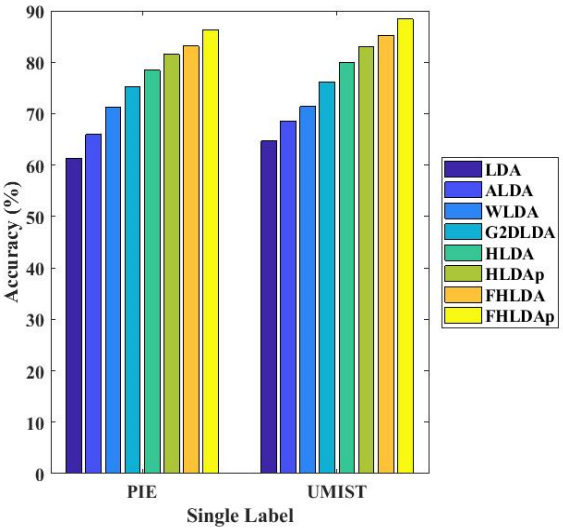

Figure. 4 Comparison of accuracy for single-label classification

Fig. 3 shows the recall outcomes for single-label classification using various algorithms tested on PIE and UMIST datasets. It notices that the FHLDA and FHLDAp algorithms achieve a better recall compared to all other algorithms. For the PIE dataset, the recall of FHLDA and FHLDAp is 79.85 % and 81.52 %, respectively, which are higher than the recall of all other algorithms for single-label classification. Also, for the UMIST dataset, the recall of FHLDA and FHLDAp is 81.45 % and 82.66 %, accordingly, which are greater than the other algorithms such as LDA, ALDA, WLDA, G2DLDA, HLDA, and HLDAp.

Fig. 4 displays the accuracy of different algorithms for single-label classification executed on PIE and UMIST datasets. It indicates that the

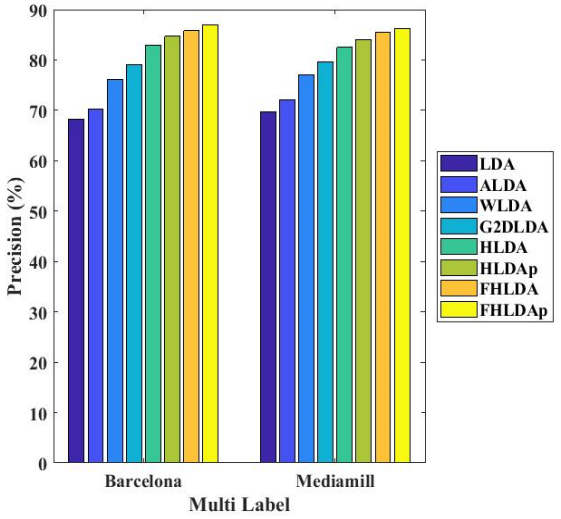

Figure. 5 Comparison of precision for multi-label classification

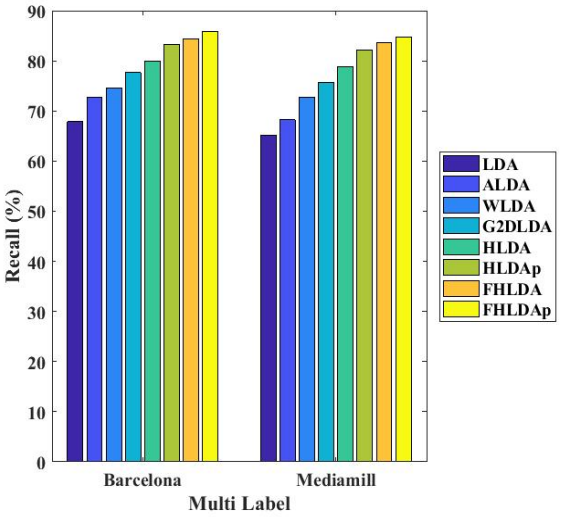

Figure. 6 Comparison of recall for multi-label classification

FHLDA and FHLDAp algorithms accomplish an increased accuracy than all other algorithms. For the PIE dataset, the accuracy of FHLDA and FHLDAp is 83.14 % and 86.72 %, accordingly, which are enhanced than the accuracy of all other algorithms for single-label classification. Likewise, for the UMIST dataset, the accuracy of FHLDA and FHLDAp is 85.23 % and 88.42 %, respectively, which are enhanced compared to the LDA, ALDA, WLDA, G2DLDA, HLDA, and HLDAp algorithms.

5.2 Multi-label classification analysis

Fig. 5 exhibits the precision obtained by different algorithms for multi-label classification tested on Barcelona and Mediamill datasets. It analyzes that the FHLDA and FHLDAp can

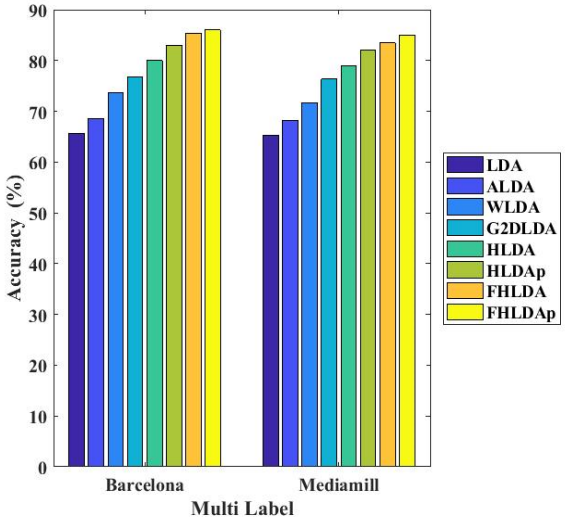

Figure.7 Comparison of accuracy for multi-label classification

increase the precision than all other algorithms. For the Barcelona dataset, the precision of FHLDA and FHLDAp is 85.78 % and 87.01 %, accordingly, which are improved than all the other algorithms for multi-label classification. Also, for the Mediamill dataset, the precision of FHLDA and FHLDAp is 85.45 % and 86.31 %, respectively, which are improved than the LDA, ALDA, WLDA, G2DLDA, HLDA, and HLDAp algorithms.

Fig. 6 shows the recall of different algorithms for multi-label classification executed on Barcelona and Mediamill datasets. It observes that the FHLDA and FHLDAp algorithms have a higher recall than all other algorithms. For the Barcelona dataset, the recall of FHLDA and FHLDAp is 84.36 % and 85.83 %, respectively, which are better than all the other algorithms for multi-label classification. Similarly, for the Mediamill dataset, the recall of FHLDA and FHLDAp is 83.64 % and 84.72 %, accordingly, which are greater compared to the LDA, ALDA, WLDA, G2DLDA, HLDA, and HLDAp algorithms.

Fig. 7 illustrates the accuracy of various algorithms for multi-label classification tested on Barcelona and Mediamill datasets. It addresses that the FHLDA and FHLDAp algorithms can maximize the accuracy compared to all other algorithms. For the Barcelona dataset, the accuracy of FHLDA and FHLDAp is 85.41 % and 86.01 %, respectively, which are larger than all other algorithms for multi-label classification. Additionally, for the Mediamill dataset, the accuracy of FHLDA and FHLDAp is 83.45 % and 85 %, accordingly, which are larger than the LDA, ALDA, WLDA, G2DLDA, HLDA, and HLDAp algorithms.

In addition to these metrics, multi-label classification is analyzed based on other few metrics such as hamming loss (HL), ranking loss (RL), 1-error (1-E), coverage (Covg), and mean precision (MP).

- HL is used to analyze the number of the instance-label set is miscategorized i.e., a label not belonging to the instance is classified or a label belonging to the instance is not classified.
- RL is used to determine the mean ratio of label sets which are reversely sorted for the instance. The efficiency is accurate if RL is 0; the lower the range of RL, the better the efficiency.
- 1-E is used to evaluate how many times the top-sorted label is not in the collection of appropriate labels for the instance. The efficiency is accurate if 1-E is 0; the lower the range of 1-E, the better the efficiency.
- Coverage is used to analyze how far it is essential, on the mean, to go down the list of labels for covering each appropriate label of the instance. The lower the range of coverage, the better the efficiency.
- MP is used to measure the mean ratio of labels sorted greater than the other relevant label. The efficiency is accurate if the MP is 1; the higher the range of MP, the better the efficiency.

These evaluation metrics results for multi-class classification when dimensionality reduction algorithms are applied are reported in Table 3.

From these analyzes, it is noticed that the FHLDA and FHLDAp algorithms realize the enhanced classification efficiency by reducing the high-dimensional data. The HL, RL, 1-E and coverage of FHLDA and FHLDAp algorithms are efficiently decreased while increasing the MP ranges of multi-label classification with the aid of dimensionality reduction using classical algorithms.

6. Conclusion

In this study, the FHLDA and FHLDAp algorithms were presented for dimensionality reduction by applying the joint diagonalization scheme. This scheme was based on the Taylor expansion which decreases the number of iterations in the HLDA algorithm during creating the new discriminants. In these algorithms, all elements of the Eigenvector matrix were determined at all iterations to lessen the computational cost.

Table 3. Multi-label classification analysis

Dataset	Metrics	LDA	ALDA	WLDA	G2DLDA	HLDA	HLDAp	FHLDA	FHLDAp
Barcelona	HL	0.394	0.349	0.341	0.333	0.312	0.316	0.268	0.273
	RL	0.281	0.267	0.260	0.249	0.236	0.235	0.194	0.211
	1-E	0.114	0.105	0.093	0.082	0.071	0.075	0.052	0.046
	Covg	2.221	2.215	2.206	2.194	2.150	2.187	2.019	2.005
	MP	0.779	0.792	0.804	0.811	0.895	0.887	0.916	0.928
Mediamill	HL	0.093	0.085	0.079	0.067	0.056	0.057	0.041	0.036
	RL	0.122	0.114	0.108	0.099	0.086	0.090	0.078	0.069
	1-E	0.155	0.142	0.136	0.130	0.121	0.123	0.114	0.105
	Covg	25.64	25.50	25.41	25.29	25.20	25.00	25.06	24.97
	MP	0.646	0.655	0.669	0.683	0.698	0.694	0.712	0.724

Additionally, the first-order approximation of the inverse Eigenvector matrix and the entire matrix of Eigenvectors was modified at all iterations. Besides, the optimized discriminant vector was obtained for avoiding the overlap among the samples of diverse classes via maximizing the between-class scatter matrices. To conclude, the findings revealed that the FHLDA and FHLDAp algorithms have higher performance on both single-label and multi-label classification using different datasets compared to all other classical dimensionality reduction algorithms.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Ranjith Nadarajan; Methodology, Sritha Sreedharan Marydasan; Software, Simulation, Sritha Sreedharan Marydasan; Writing- Original draft preparation, Sritha Sreedharan Marydasan; Visualization, Investigation, Supervision, Ranjith Nadarajan; Reviewing and Editing, Ranjith Nadarajan.

References

[1] I. Syarif, "Dimensionality reduction algorithms on high dimensional datasets", *EMITTER International Journal of Engineering Technology*, Vol. 2, No. 2, pp. 28-38, 2014.

[2] A. Ullah, U. Qamar, F. H. Khan, and S. Bashir, "Dimensionality reduction approaches and evolving challenges in high dimensional data", In: *Proc. of the 1st International Conf. On Internet of Things and Machine Learning*, pp. 1-8, 2017.

[3] M. T. Islam and L. Xing, "A data-driven dimensionality-reduction algorithm for the exploration of patterns in biomedical data", *Nature Biomedical Engineering*, pp. 1-12.

[4] P. Chen, L. Jiao, F. Liu, Z. Zhao, and J. Zhao, "Adaptive sparse graph learning based dimensionality reduction for classification", *Applied Soft Computing*, Vol. 82, pp. 1-15, 2019.

[5] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, and T. Baker, "Analysis of dimensionality reduction techniques on big data", *IEEE Access*, Vol. 8, pp. 54776-54788, 2020.

[6] F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)", *Computer Science Review*, Vol. 40, p. 100378, 2021.

[7] S. Sadaghiyanfam and M. Kuntalp, "Comparing the performances of PCA (principle component analysis) and LDA (linear discriminant analysis) transformations on PAF (paroxysmal atrial fibrillation) patient detection", In: *Proc. of the 3rd International Conf. on Biomedical Imaging, Signal Processing*, pp. 1-5, 2018.

[8] S. Zheng, C. Ding, F. Nie, and H. Huang, "Harmonic mean linear discriminant analysis", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 31, No. 8, pp. 1520-1531, 2018.

[9] J. Ma and Y. Yuan, "Dimension reduction of image deep feature using PCA", *Journal of Visual Communication and Image Representation*, Vol. 63, pp. 1-8, 2019.

[10] W. Yan, H. Sun, Q. Sun, Z. Zheng, X. Gao, Q. Zhang, and Z. Ren, "Multiple kernel dimensionality reduction based on collaborative representation for set oriented image classification", *Expert Systems with Applications*, Vol. 137, pp. 380-391, 2019.

[11] W. Zhang, J. Wu, and Y. Yang, "Wi-HSNN: a subnetwork-based encoding structure for dimension reduction and food classification via

harnessing multi-CNN model high-level features", *Neurocomputing*, Vol. 414, pp. 57-66, 2020.

[12] M. Ramamurthy, Y. H. Robinson, S. Vimal, and A. Suresh, "Auto encoder based dimensionality reduction and classification using convolutional neural networks for hyperspectral images", *Microprocessors and Microsystems*, Vol. 79, pp. 1-27, 2020.

[13] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 58, No. 8, pp. 5336-5353, 2020.

[14] W. Chen, Y. Xu, Z. Yu, W. Cao, C. P. Chen, and G. Han, "Hybrid dimensionality reduction forest with pruning for high-dimensional data classification", *IEEE Access*, Vol. 8, pp. 40138-40150, 2020.

[15] K. T. Le, C. Chaux, F. J. Richard, and E. Guedj, "An adapted linear discriminant analysis with variable selection for the classification in high-dimension, and an application to medical data", *Computational Statistics & Data Analysis*, Vol. 152, pp. 1-18, 2020.

[16] Y. Li, Y. Chai, H. Zhou, and H. Yin, "A novel dimension reduction and dictionary learning framework for high-dimensional data classification", *Pattern Recognition*, Vol 112, pp. 1-39, 2021.

[17] R. Lu, Y. Cai, J. Zhu, F. Nie, and H. Yang, "Dimension reduction of multimodal data by auto-weighted local discriminant analysis", *Neurocomputing*, Vol. 461, pp. 27-40.

[18] C. N. Li, Y. H. Shao, W. J. Chen, Z. Wang, and N. Y. Deng, "Generalized two-dimensional linear discriminant analysis with regularization", *Neural Networks*, Vol. 142, pp. 73-91, 2021.

[19] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database", In: *Proc. of the Fifth IEEE International Conf. on Automatic Face Gesture Recognition*, pp. 53-58, 2002.

[20] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using LDA-based algorithms", *IEEE Transactions on Neural Networks*, Vol. 14, No. 1, pp. 195-200, 2003.

[21] C. G. Snoek, M. Worring, J. C. V. Gemert, J. M. Geusebroek, and A. W. Smeulders, "The challenge problem for automated detection of 101 semantic concepts in multimedia", In: *Proc. of the 14th ACM International Conf. On Multimedia*, pp. 421-430, 2006.

[22] S. Zheng and C. Ding, "Kernel alignment inspired linear discriminant analysis", In: *Proc. of Joint European Conf. on Machine Learning and Knowledge Discovery in Databases*, Springer, Berlin, Heidelberg, pp. 401-416, 2014.