
Improved Fuzzy Local Mean Discriminant Analysis via Iterative Optimization for Feature Transformation and Classification

Saeed Maadani⁽¹⁾ Mohammad Hossein Moattar^{(2)*} Yahya Forghani⁽³⁾

(1) Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran

(2) Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran*

(3) Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran

(Date received:1403/08/17

Date accepted:1403/11/16)

Abstract

Fuzzy Local Mean Discriminant Analysis (FLMDA) is a supervised dimensionality reduction method. FLMDA gathers local information for constructing between-class and within-class scatters. However, after feature transformation using FLMDA, the neighboring data points may differ. This fact may degrade the classification performance. In the proposed method, the feature extraction process is repeated based on the list of adjacent data after transformation, and this process continues until convergence. Therefore, it is supposed that the local information is preserved as much as possible and the local discrimination between the instances of different classes is increased. The experiments performed on different University of California, Irvine (UCI) datasets show the superiority of the proposed method compared to similar methods.

Keywords: Fuzzy Local Mean Discriminant Analysis (FLMDA), Linear Discriminant Analysis (LDA), Dimensionality reduction, Feature extraction, Classification

* Corresponding author:

Mohammad Hossein Moattar

Address: Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran

Email: moattar@mshdiau.ac.ir

1- Introduction

In machine learning tasks, dimensionality reduction leads to the discovery of the hidden structure of the data [1]. Dimensionality reduction methods are divided into two groups of methods namely feature selection [2-17] and feature extraction. In features extraction (FE) methods, each extracted feature is a linear or nonlinear combination of initial features. Therefore, FE methods are divided into linear [18, 19] and nonlinear approaches [20-29]. Principle Component Analysis (PCA) [19] and Linear Discriminant Analysis (LDA) [18] are among the most widely used linear feature transformation and extraction methods. Among these approaches, LDA is a supervised feature transformation method. The purpose of LDA is to find features that make the data more separable and performs it by maximizing the between-class scatters and minimizing the within-class scatters.

In LDA, the Euclidean measure or L2 norm is used to calculate the data distribution. Many papers have been presented based on the LDA method such as [30, 31]. Zhong et al. [32] introduced a noise resistant version of LDA called L1-LDA. In their model, instead of L2 norm, L1 norm is used for calculating scatterings. It is claimed that, L1 norm has lower sensitivity to noise. Zhang et al. [33] also claimed that their proposed method unlike L1-norm that is sensitive to outliers, via integrating global and local structure information is robust to outliers. Yang et al. [34] proposed 2D-LDA methodology in 2005 for 2-dimensional data. In their method, there is no need to vectorize image data. It is claimed that the time complexity of the method is lower than the LDA method, and can sometimes lead to the extraction of more discriminative features. Li et al. [35] proposed L1-2DLDA, which combines the advantages of two previously introduced approaches, namely L1-LDA and 2D-LDA.

In the LDA, new features are extracted so that the distance between each data of the same class is minimized and the distance between data from different classes is maximized. While, in particular, when classification methods such as k-nearest neighbor is supposed to be used to classify data, what is needed is to minimize the average distance of each data from the neighbors with the same class labels and maximize the average distance from the neighbors belonging other classes. This idea is developed in [36] and [37] which proposed the local LDA method. In [38] a Fuzzy-Gaussian two-directional inverse FDA is proposed which computes the fuzzy and Gaussian membership values to obtain the class-wise and global means. The Authors of [39] have proposed a new feature extraction method using entropy based fractional fuzzy in which uncertainty is incorporated into 2D-LDA. In this regard, a fuzzy logic-based feature extraction

technique, called fuzzy generalized two-dimensional FLDA is proposed in [42]. Also, a new collaborative representation-based fuzzy discriminant analysis is proposed in [43].

Similarly, some articles such as [40, 41, 44-46] use fuzzy set theory. Jie Xu et al. [36], using the theory of "fuzzy sets", introduced a new feature extraction algorithm in 2016, which they called Fuzzy Local Mean Discriminant Analysis (FLMDA). In this method, the new features are extracted so that the distance of each data point from the weighted average of the adjacent data of the same-class is minimized and the distance between each data point and the weighted average of adjacent other-class data is maximized. In FLMDA, if a data point is on the boundary of the two classes, less weight is given to it due to less certainty for the class label.

After performing FLMDA, the data adjacencies may change, which may in turn causes that the previous adjacencies for which the scatterings are calculated are not valid anymore which may lead to performance degradation. Therefore, in the proposed method, the feature extraction process is iterated based on the FLMDA method and the new list of adjacent data, and this process continues until convergence. Using this approach, we expect that the classification accuracy is enhanced while clearly the complexity of feature mapping may increase. The rest of this paper is organized as follows. In Section 2, the FLMDA approach is described with more detail. The proposed method is explained in Section 3. In Section 4, the results of the experiments are discussed and Section 5 includes the conclusions and direction for future work.

2- Fuzzy Local Mean Discriminant Analysis (FLMDA)

A single label training data belongs to just one class. In other words, the degree of membership of a training data to a class is equal to one and the degree of its membership to other classes is zero. Suppose $u_i(x_j)$ is the predefined degree of membership of the training data x_j to the i^{th} class. Equation (1) shows how to calculate the degree of membership of this training data to the class.

$$\tilde{u}_i(x_j) = \begin{cases} 0.51 + 0.49 \frac{m_{ij}}{k}, & u_i(x_j) = 1; \\ 0.49 \frac{m_{ij}}{k}, & u_i(x_j) = 0; \end{cases} \quad (1)$$

In which $u_i(x_j)$ and $\tilde{u}_i(x_j)$ are the crisp and fuzzy membership degrees, respectively. In Eq. (1), k is a hyper-parameter of the fuzzy k nearest neighbor method, and m_{ij} is the number of nearest neighbors of x_j which have the same label as i . The higher m_{ij} , the higher becomes the membership degree of the x_j to the i^{th} class. In fact, based on this equation, if a training data of a particular class is located at the center of the class, its membership to that class is almost preserved, and if it is located on the boundary of the two classes, its degree of membership to its original class is reduced and the degree of its membership to other classes is increased slightly so that the uncertainty about the label of the data can be exploited at a later stage. According to Eq. (1), the degree of membership of a data point labeled as i to the i^{th} class changes in the range of $[0.51, 1]$, and its membership degree to other classes is at least 0 and at most 0.49. Hence, even after modifying the degrees of membership of the data points in accordance with equation (1), this data has still higher membership of original class.

Suppose $\{x_1, x_2, \dots, x_N\}$ is the training data of c different classes and the set $\{x_{i^s}^1, \dots, x_{i^s}^k\}$ is the k nearest neighbors of x_i which belong to different classes $s \in \{1, 2, \dots, c\}$ and $\Gamma_i^s = \{i^1, \dots, i^k\}$ is the set of indices of these neighbors. The weighted average of these k nearest neighbors of x_i , belonging to class s is defined as:

$$m_i^s = \frac{\sum_{j \in \Gamma_i^s} \tilde{u}_i(x_j) x_j}{\sum_{j \in \Gamma_i^s} \tilde{u}_i(x_j)} \quad (2)$$

In the FLMMDA method, data are transmitted using a linear transformation to a new space with a lower number of dimensions where the ratio of the local between-class scattering and the local within-class scattering is maximized. Local within-class scattering is the sum of distances of each data from the weighted average of its same-class neighboring data, while local between-class scattering is the sum of distances of each data from the weighted average of its other-class neighboring data. More precisely, if the data is M -dimensional, the orthogonal transformation matrix $p = (p_1, p_2, \dots, p_M)$ is determined in such a way that the following target function is minimized:

$$\det\left(\frac{1}{N} \sum_{i=1}^N (p^T x_i - p^T m_i^{l_i}) (p^T x_i - p^T m_i^{l_i})^T\right) = \det\left(p^T \left[\frac{1}{N} \sum_{i=1}^N (x_i - m_i^{l_i}) (x_i - m_i^{l_i})^T\right] p\right) = \det(p^T S_w p) \quad (3)$$

And the following target function is maximized:

$$\det\left(\frac{1}{N(c-1)}\sum_{i=1}^N\sum_{s=1,s\neq l_i}^c(p^T x_i - p^T m_i^s)(p^T x_i - p^T m_i^s)^T\right) = \det\left(p^T\left[\frac{1}{N(c-1)}\sum_{i=1}^N\sum_{s=1,s\neq l_i}^c(x_i - m_i^s)(x_i - m_i^s)^T\right]p\right) = \det(p^T S_b p) \quad (4)$$

Where l_i is the label of x_i and

$$S_w = \frac{1}{N}\sum_{i=1}^N(x_i - m_i^{l_i})(x_i - m_i^{l_i})^T, \quad (5)$$

$$S_b = \frac{1}{N(c-1)}\sum_{i=1}^N\sum_{s=1,s\neq l_i}^c(x_i - m_i^s)(x_i - m_i^s)^T, \quad (6)$$

S_w and S_b are called fuzzy local within-class and between-class scatterings, respectively. To obtain an orthogonal matrix p , it is sufficient to solve the following model:

$$\begin{aligned} \max_p \quad & \frac{\det(p^T S_b p)}{\det(p^T S_w p)} \\ \text{s. t.} \quad & p^T p = I \end{aligned} \quad (7)$$

It can be shown that the optimal values of the orthogonal matrix, p , Eq. (7) are the eigenvector of $S_w^{-1}S_b$ [47].

3- The Proposed Method

In the FLMDA method, data are mapped using a linear transformation to a new space with a lower dimensionality, where the ratio of the local between-class scattering and the local within-class scattering is maximized. However, these scatterings are calculated in original space which means that it is not necessarily the same as the new space. In the other word, after finding new feature space using the FLMDA method, the adjacent points of each data may change. Therefore, in the proposed method, the FLMDA feature extraction process is iterated and, in each iteration, the new list of adjacent data is populated. This process continues until convergence.

The proposed model is similar to the FLMDA model and the objective function is the

same as Eq. 7 and as follows:

$$\begin{aligned} & \max_p \frac{\det(p^T S_b p)}{\det(p^T S_w p)} \quad (8) \\ & \text{s. t.} \quad p^T p = I \end{aligned}$$

in which S_w and S_b are calculated as stated in Equations 5 and 6, and having transformation p , The weighted average of these k nearest neighbors of x_i , belonging to class s is defined as:

$$m_i^s = \frac{\sum_{j \in \Gamma_{i,p}^s} \tilde{u}_i(p^T x_j) p^T x_j}{\sum_{j \in \Gamma_{i,p}^s} \tilde{u}_i(p^T x_j)} \quad (9)$$

In which, in the mapping space, the set $\{p^T x_{i_1}^s, \dots, p^T x_{i_k}^s\}$ is the k nearest neighbors of the data $p^T x_i$ in class s and $\Gamma_{i,p}^s = \{i^1, \dots, i^k\}$ is the set of indices of these k neighbors. $p^T x_i$ denotes the identical x_i training data after a round of mapping.

As mentioned before, unlike the FLMDA model, in the proposed model, the local mean vector of i^{th} training data, namely m_i^s , is variable and dependent on transformation matrix p which should be determined in each iteration of the optimization process. Given that m_i^s is variable, the solution of Eq. (8) is difficult. Therefore, an iterative approximation method is used to solve it. In this iterative method, assuming p to be constant, the values of m_i^s are determined according to Eq. (9), and then, with the assumption of m_i^s being constant, Eq. (8) is solved. Obviously, when m_i^s are assumed to be constant, problem (8) changes to the FLMDA problem. After solving the FLMDA problem and obtaining the optimal value of the matrix p , m_i^s are again determined according to Eq. (9) and the previous process is repeated until convergence. The proposed algorithm is summarized in the following:

The proposed algorithm
Input: The training dataset $\{x_1, x_2, \dots, x_N\}$ and their corresponding labels <p style="text-align: center;">d: final dimension, <i>thr</i>: threshold value</p> Output: The final projection matrix, p
Step 1: initialize $p=I$

Step 2: feature extraction: for each i compute $x_i = p^T x_i$

Step 3: Compute m_i^s using Eq. 9.

Step 4: Fixing m_i^s compute Eq. 8. Set p as the r best Eigen vectors of $S_w^{-1} S_b$

Step 5: If the Frobenius norm of two consecutive values of P the higher than thr go to step 2 else the algorithm is terminated.

4- Experiments

In this section, the proposed method and other methods such as LDA and FLMDA are evaluated using nine real data sets from UCI repository, namely Wine, Glass, Parkinson, Sonar, Yeast, Breast Tissue, Haberman's Survival, Iris and Ecoli. Each time using one of the three mentioned methods, the feature extraction takes place, and then, using the KNN method, the data is classified in the transformation space. 10-fold Cross Validation is used to evaluate the classification accuracy.

Table 1 shows the highest accuracy obtained, the rank of each method among the three methods, the value of optimal parameters and the runtime of each method. The optimal value of the k at the stage of feature extraction and classification is selected from the set of $\{1, 5, 10, 15, 20, 25, 30\}$ and the optimal value for the number of extracted features or the dimension of the data after the feature reduction, namely r , is chosen from the set of $\{1, 2, \dots, M\}$.

We first select the optimal value of k from the set $\{1, 2, 3, \dots, 30\}$ and then repeat the experiments with less values of k , from the set of $\{1, 5, 10, 15, 20, 25, 30\}$ and we compared the results. We saw that the results of the experiments converged for both sets, and both sets of k had almost similar results. It can be inferred that the proposed method has a higher accuracy for the set of different values of k than the other two methods. Therefore, we considered the results for convenience in the set of $\{1, 5, 10, 15, 20, 25, 30\}$.

Table 1: performance comparison of different feature extraction methods

Dataset	Feature extraction method	Accuracy (%)	Training run time (s)	Value of k in classification	Value of k in feature extraction	Final dimension
Wine	LDA	98.8235 (1)	0.1139	5	-	2

	FLMDA	97.0588 (2)	1.2038	1	20	13
	Proposed	98.8235 (1)	2.9011	5	30	13
Iris	LDA	98 (1)	0.2140	10	-	3
	FLMDA	97.3333 (2)	0.8855	10	30	3
	Proposed	98 (1)	2.0997	1	25	3
Glass	LDA	47.1429 (3)	0.2557	5	-	8
	FLMDA	52.3810 (1)	2.1108	1	1	8
	Proposed	49.5238 (2)	5.7013	5	5	6
Sonar	LDA	63 (3)	0.2653	5	-	1
	FLMDA	70.500 (2)	1.2798	25	5	1
	Proposed	72 (1)	5.2143	20	15	1
Parkinson	LDA	82.6316 (3)	0.2117	5	-	1
	FLMDA	87.3684 (2)	1.1802	10	20	1
	Proposed	88.4211 (1)	3.0388	20	25	2
Haberman's Survival	LDA	76.6667 (3)	0.2689	30	-	3
	FLMDA	77.3333 (2)	2.2124	15	5	3
	Proposed	78.3333 (1)	3.8312	15	30	3
Yeast	LDA	56.6216 (3)	0.9221	25	-	8
	FLMDA	57.9054 (2)	63.2307	20	15	2
	Proposed	58.1081 (1)	118.2097	20	15	2
Breast Tissue	LDA	54 (2)	0.2067	1	-	6
	FLMDA	52 (3)	0.8142	1	1	2
	Proposed	58 (1)	3.4271	1	1	2

	LDA	74.8485 (2)	0.2554	5	-	9
Ecoli	FLMDA	74.8485 (2)	4.3023	5	10	7
	Proposed	76.0606 (1)	7.2212	5	25	5

The results of Table 1 show that the proposed method, except in one case (i.e. average accuracy on glass data set), has higher accuracy compared to LDA and FLMDA methods. The lowest rate of accuracy using each of the three LDA, FLMDA and proposed methods was on glass dataset. This may denote that the proposed approach degrades on highly complex datasets and on datasets with medium complexity, the approach outperforms similar approaches. Also as denoted in Table 1, LDA has comparable results as the original FLMDA method and in 3 out of 9 experiments has outperformed FLMDA while for one dataset the results were actually the same. Also except for one dataset (i.e. Wine), the best dimensionality which has resulted the best KNN classification accuracy is almost the same for all three approaches, which is due to the same objective and metric used in these methods.

According to Table 1, the execution time of the LDA algorithm is lower than that of FLMDA. To justify it, it should be said that although in both methods the $S_w^{-1}S_b$ matrix is calculated, but calculating S_b and S_w in the FLMDA method requires more time. Since in each iteration of the proposed algorithm, a problem similar to the FLMDA problem is solved once, the time complexity time of the proposed method is more than the complexity of FLMDA and LDA methods. As seen in Table 1, the implementation time of the proposed method is almost twice as the FLMDA method.

To test the significance of the proposed approach compared with the two other methods, we applied the Wilcoxon ranked sum test with the following hypothesis:

H0: "There is no significant difference between approaches".

H1: "There is significant difference between approaches".

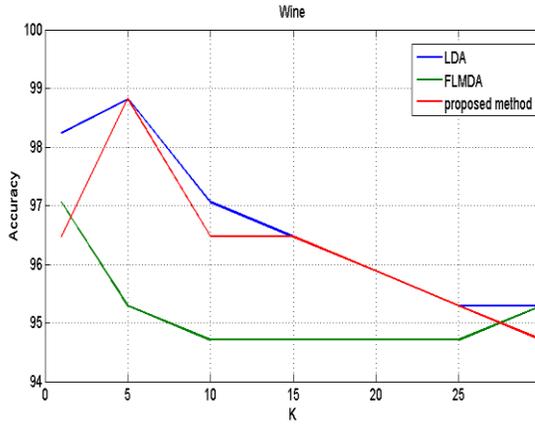
Table 2: Outputs of Wilcoxon ranked sum test for the evaluated approaches.

Algorithm	LDA	FLMDA	Proposed
LDA	-	20	0
FLMDA		-	6
Proposed			-

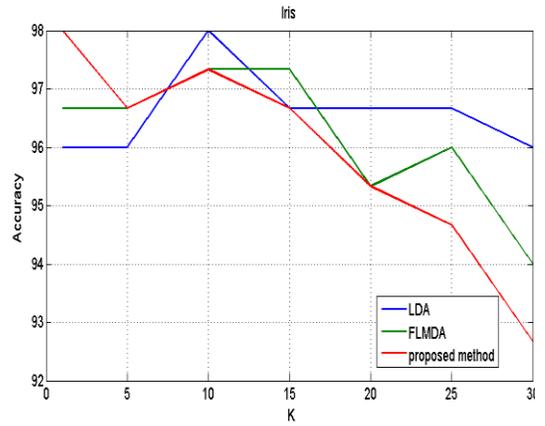
Having 9 experiments for each approach, the Wilcoxon critical value for two-tails test with $\alpha=0.05$ level of significance, is 6. The output of the mentioned test is showed in Table 2. Bold numbers show the test for which the output of the Wilcoxon test is equal to or lower than the critical value and the H_0 hypothesis cannot be rejected which means that there is a significant difference between the approaches.

Figure 1 illustrates the accuracy of three feature extraction methods across varying values of k , paired with the designated classifier, while Figure 2 depicts their accuracy in relation to the final latent space dimension. Across both figures, the overall trends of the three methods remain consistent, displaying similar curve shapes for all datasets. This indicates that the sensitivity of these approaches to k (or the latent space dimension) is comparable, highlighting the importance of selecting an appropriate value for k to optimize performance. The uniformity in curve shapes suggests that all methods share a common dependency on this hyperparameter, making its careful tuning a critical aspect of achieving optimal results.

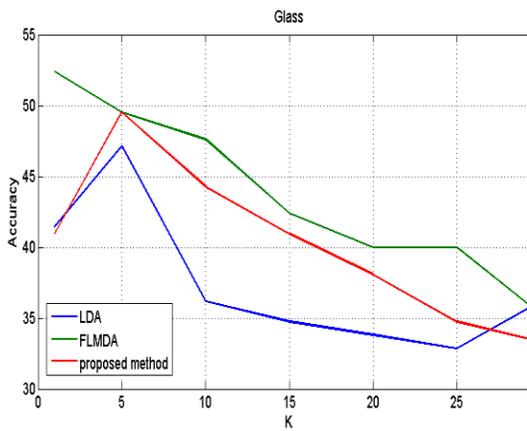
Moreover, despite the similarities in trend, the proposed approach demonstrates a notable advantage in terms of area under the curve (AUC) when compared to the other two methods. This superiority is evident across most datasets, with the Glass dataset being a notable exception. The higher AUC values for the proposed approach suggest its robustness and effectiveness, particularly when the optimal k or latent space dimension is determined. These findings underline the importance of the proposed method as a more reliable option for feature extraction, further emphasizing its potential benefits in applications requiring high sensitivity and accuracy.



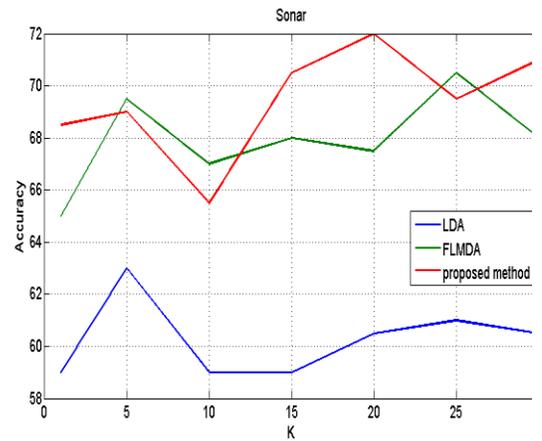
a) Wine dataset



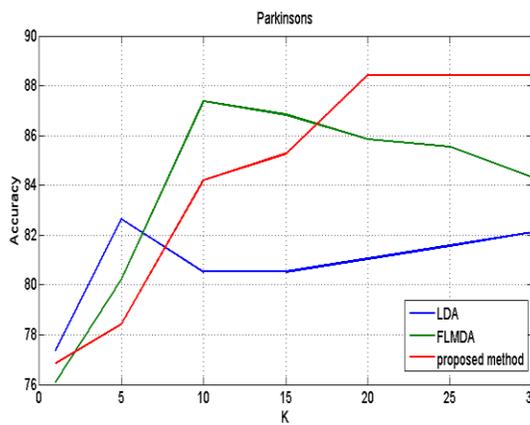
b) Iris dataset



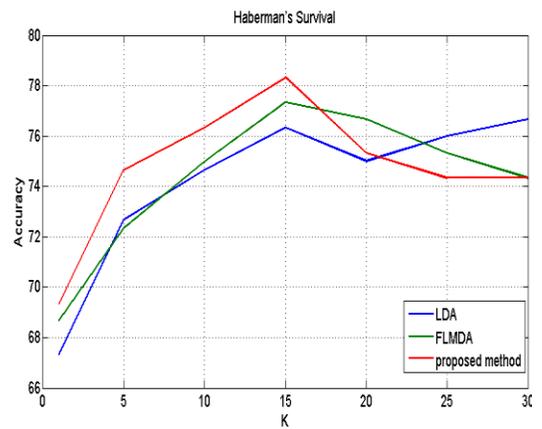
c) Glass dataset



d) Sonar dataset



e) Parkinson dataset



f) Haberman's Survival dataset

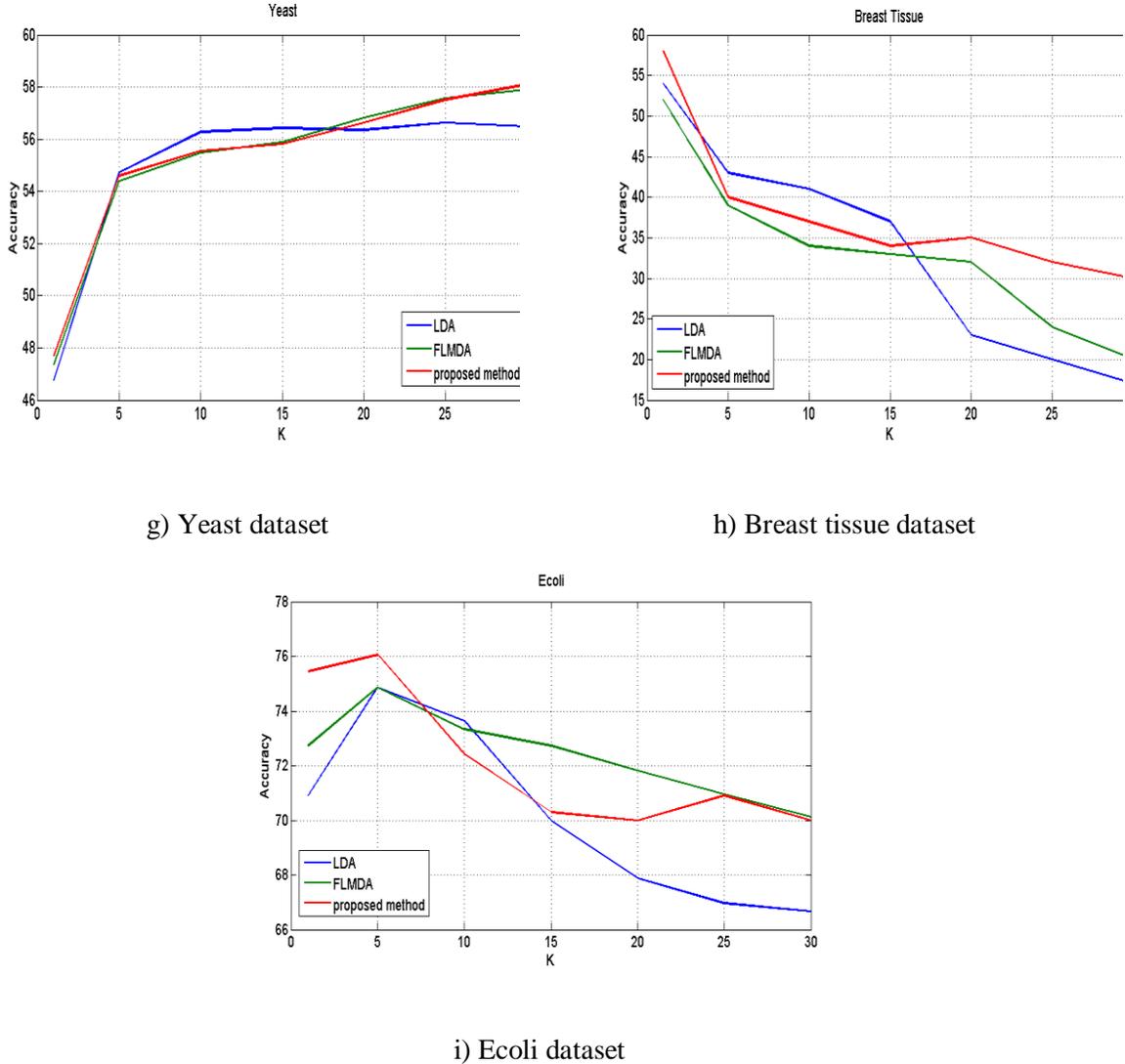
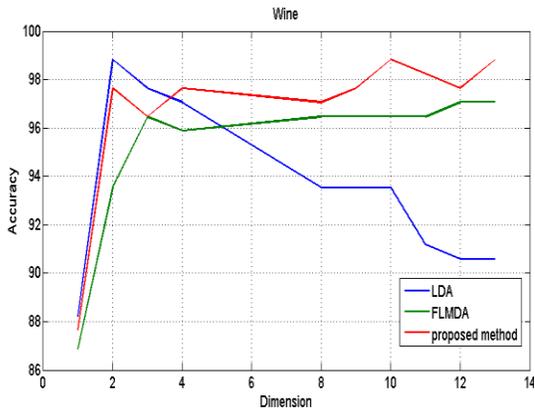


Fig. 1: The effect of the parameter k on the accuracy of the methods for various data sets

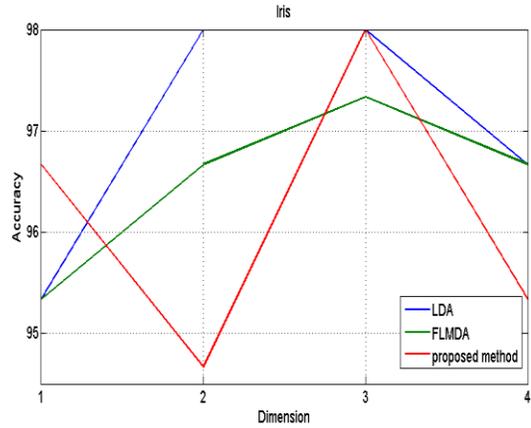
The experiments illustrated in Figure 1 reveal that the value of k significantly influences the performance and accuracy of all three approaches. For datasets such as Wine, Iris, Glass, Breast Tissue, and Ecoli, lower values of k generally yield better results, suggesting that reduced dimensionality enhances performance for these datasets. Conversely, for datasets like Yeast and Parkinson, higher k values improve accuracy, indicating that retaining more dimensions is beneficial in these cases. This variation can be attributed to dataset characteristics, including the size of the dataset and the degree of class imbalance, which affect the suitability of specific k values.

From these observations, it can be concluded that the optimal choice of k is dataset-dependent. When dealing with balanced datasets with a relatively small number of samples, lower k values are advantageous, as they simplify the feature space without

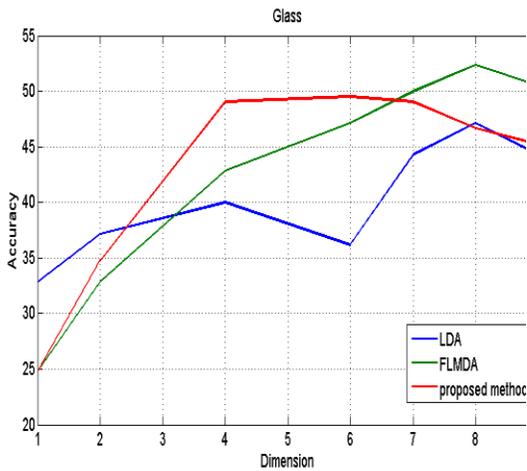
sacrificing critical information. On the other hand, for larger datasets with significant class imbalance, higher k values may preserve more information from the latent space, leading to improved performance. These findings emphasize the importance of tailoring the k parameter to the specific characteristics of the dataset to achieve optimal results.



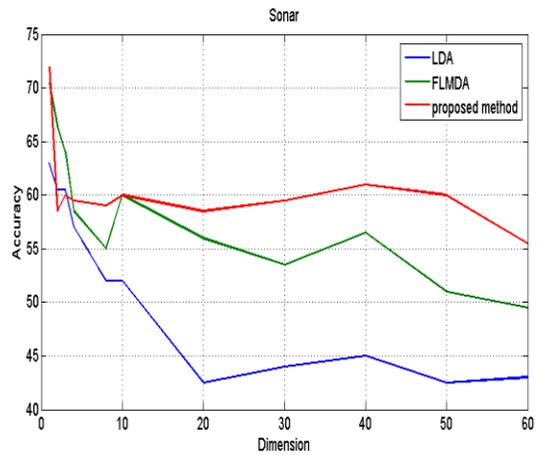
a) Wine dataset



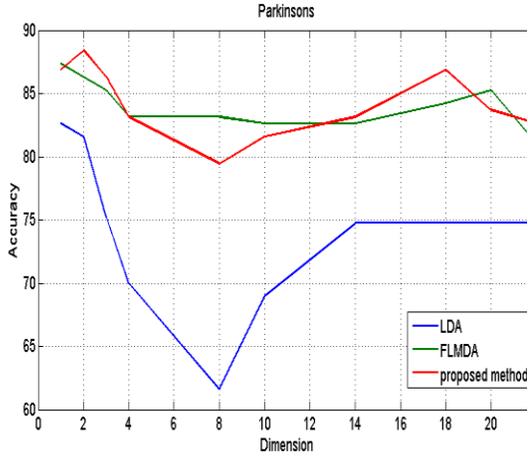
b) Iris dataset



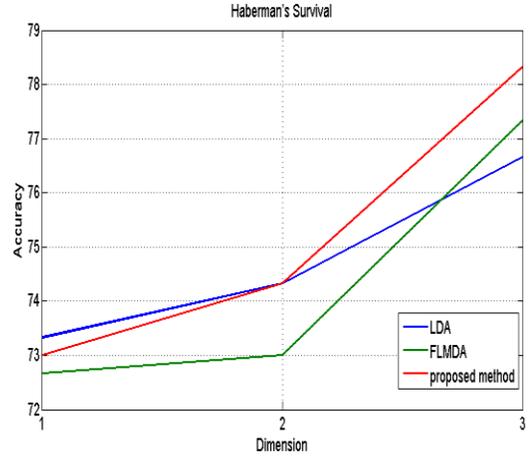
c) Glass dataset



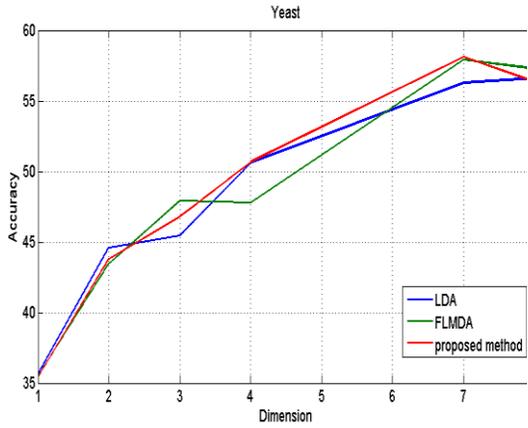
d) Sonar dataset



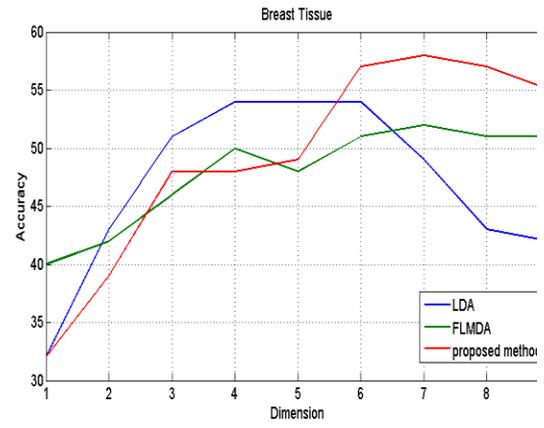
e) Parkinson dataset



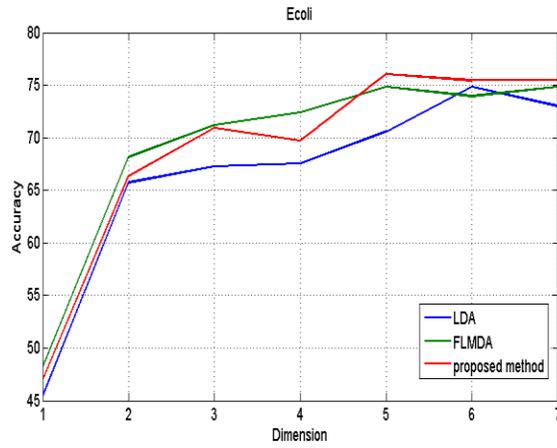
f) Haberman's Survival dataset



g) Yeast dataset



h) Breast tissue dataset



i) Ecoli dataset

Fig. 2: The effect of changing the dimensionality on the accuracy of the methods for different data sets

Figure 2 highlights the impact of dimensionality on the performance of the three feature extraction approaches, with the experimented dimensions varying according to the size of the original feature space. The proposed method demonstrates superior performance compared to the other two approaches in most cases, with the exception of the Glass dataset, previously discussed, and the Iris dataset, where the LDA approach achieves a better area under the curve (AUC). Notably, the performance of the proposed method generally improves as the dimensionality of the mapping space increases, suggesting its robustness and adaptability to higher-dimensional feature transformations.

However, two notable exceptions are observed with the Sonar and Parkinson datasets, where performance degrades as the dimensionality increases. This decline may stem from the smaller latent dimensionality in these datasets, which limits the effectiveness of mapping to higher dimensions. These findings indicate that while the proposed approach excels in most cases, the relationship between dimensionality and performance is not universal and depends on dataset-specific characteristics. It underscores the importance of carefully selecting the dimensionality of the transformation space to optimize results, particularly for datasets with unique feature distributions.

On the other hand, as shown in Figure 2, the LDA approach has outperformed the other two methods in several instances, particularly for datasets like Iris. LDA's superior performance on the Iris dataset can be attributed to its simplicity and effectiveness in handling datasets with well-defined class separations. The linear nature of LDA makes it particularly suitable for such cases, where more complex methods such as FLMDA may not provide significant additional benefits. This suggests that, in scenarios with clear class boundaries and lower-dimensional feature spaces, simpler techniques like LDA can be more effective and computationally efficient than more intricate methods.

Furthermore, success of the proposed approach in many cases highlights its robustness and efficiency in scenarios where the complexity of the data does not require the added sophistication of more advanced techniques. While more complex approaches might offer better results in some cases, the Iris dataset exemplifies situations where LDA's simplicity and low computational cost provide an optimal balance between performance and efficiency. This underscores the importance of choosing the right method based on the complexity of the dataset and the specific characteristics of the problem at hand.

5- Conclusion

In the FLMDA method, the data is transmitted using a linear transformation to a new space with a lower dimensionality, so that the ration between local between-class scatters and the local within-class scatters in that space is maximized. Local within-class scatters are the total distance of each data from the weighted average of the neighboring same-class data, while local between-class scatters is the total distance of each data from the weighted average of the neighboring other-class data. After feature transformation using FLMDA method, the data adjacencies may change. Therefore, in the proposed method, the feature extraction process is iterated based on the FLMDA method and the lists of new adjacent data are generated, and this trend continued until convergence. Experiments are performed on 9 real UCI repository datasets showed that the feature extraction using the proposed method can increase the classification accuracy compared to the LDA and FLMDA. More precisely, the proposed method is more accurate compared to LDA and FLMDA, except in one case, i.e. the glass dataset. The results show that the average accuracy of the proposed method increased by 2.837% compared to LDA and 1.171% compared to FLMDA and by 0.68% compared to the maximum of the two methods. Meanwhile, the implementation time of the proposed method is approximately twice as long as the original FLMDA method .

Even though the trials indicate modest gains in accuracy and computing complexity, it is still worthwhile to talk about how the suggested approach will affect real-world uses. Numerous fields could gain from the enhancements in feature transformation and classification accuracy, including :

- **Medical Diagnosis:** By increasing the accuracy of disease classification in high-dimensional medical datasets such as gene expression profiles, EEG, and ECG, the technique may make it possible to develop more accurate diagnostic instruments for cardiac and neurological conditions.
- **Image Recognition:** Especially for noisy or overlapping datasets, the technique may improve recognition rates in satellite imaging, object detection, and facial recognition .
- **Text and Document Classification:** By enhancing text feature representation, the technique may improve natural language processing applications such as topic modelling, spam detection, and sentiment analysis.
- **Recommendation Systems:** By better classifying user preferences and product categories, the method may improve recommendation systems .

- Financial Data Analysis: By correcting class imbalances and overlapping data, the technique may improve predictive reliability in financial forecasts and fraud detection .

Since a main issue of the proposed approach is the selection of k value, finding a metric for determining the best value of k is crucial as a direction for future works. Grid search can be used to systematically test a range of k values and identify the one with the best performance metric. Alternatively, data-driven methods like the Elbow Method analyze the trade-off between within-class scatter and between-class scatter as k varies. Heuristic approaches based on domain knowledge can also guide the selection of k. Finally, optimization-based methods, such as using genetic algorithms or Bayesian optimization, can automate and refine the k selection process.

Also, due to considerable complexity of the proposed approach, more efficient implementation of the method is necessary for the applicability of the proposed method. Also, we can study on the applicability of the proposed speedup on the recently published approaches such as [38, 39, 42, and 43].

References

- [1] Petscharnig, S., Lux, M. and Chatzichristofis, S. (2017). Dimensionality Reduction for Image Features using Deep Learning and Autoencoders. in Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing. ACM.
- [2] M. Ebrahimi Warkiani, M.H. Moattar, Comprehensive Survey on Recent Feature Selection Methods for Mixed Data: Challenges, Solutions and Future Directions. Neurocomputing, Accepted for publication, 2025.
- [3] Siddiqi, M.A.; Pak, W. Optimizing Filter-Based Feature Selection Method Flow for Intrusion Detection System. Electronics 2020, 9, 2114. <https://doi.org/10.3390/electronics9122114>
- [4] Y. Abroshan, M.H. Moattar. Discriminative Feature Selection Using Signed Laplacian Restricted Boltzmann Machine for Speed and Generalization Improvement of High Dimensional Data Classification. Applied Soft Computing, 153,111274 (2024). <https://doi.org/10.1016/j.asoc.2024.111274>
- [5] Rodrigues, D., et al. (2014). A wrapper approach for feature selection based on bat algorithm and optimum-path forest. Expert Systems with Applications, 41(5): p. 2250-2258.
- [6] M. Fattahi, M.H. Moattar, Y. Forghani. Locally alignment based manifold learning for simultaneous feature selection and extraction in classification problems. Knowledge-Based Systems, 2023. 259,110088.

- [7] Yassi, M. and Moattar, M.H. (2014). Robust and stable feature selection by integrating ranking methods and wrapper technique in genetic data classification. *Biochemical and biophysical research communications*, 446(4): p. 850-856.
- [8] Wang, A., et al. (2017). Wrapper-based gene selection with Markov blanket. *Computers in biology and medicine*, 81: p. 11-23.
- [9] Razavi Ghods, M., Moattar, M. H., and Forghani, Y., “Distance metric learning based on structural neighborhoods for dimensionality reduction and classification performance improvement”, arXiv:1902.03453, 2019. doi:10.48550/arXiv.1902.03453.
- [10] Chen, G. and Chen, J. (2015). A novel wrapper method for feature selection and its applications. *Neurocomputing*, 159: p. 219-226.
- [11] Ma, L., et al. (2017). A novel wrapper approach for feature selection in object-based image classification using polygon-based cross-validation. *IEEE Geoscience and Remote Sensing Letters*, 14(3): p. 409-413.
- [12] Lu, H., et al. (2017). A hybrid feature selection algorithm for gene expression data classification. *Neurocomputing*.
- [13] Apolloni, J., Leguizamón, G. and Alba, E. (2016). Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments. *Applied Soft Computing*, 38: p. 922-932.
- [14] Inbarani, H.H., Bagyamathi, M. and Azar, A.T. (2015). A novel hybrid feature selection method based on rough set and improved harmony search. *Neural Computing and Applications*, 26(8): p. 1859-1880.
- [15] E. Hossein, M. H. Moattar, Evolutionary feature subsets selection based on interaction information for high dimensional imbalanced data classification, *Applied Soft Computing Journal*, Vo. 82. 2019, 105581. DOI: <https://doi.org/10.1016/j.asoc.2019.105581>
- [16] Brahim, A.B. and Limam, M. (2016). A hybrid feature selection method based on instance learning and cooperative subset search. *Pattern Recognition Letters*, 69: p. 28-34.
- [17] Solorio-Fernández, S., Carrasco-Ochoa, J.A. and Martínez-Trinidad, J.F. (2016). A new hybrid filter–wrapper feature selection method for clustering based on ranking. *Neurocomputing*, 214: p. 866-880.
- [18] V. Vahabzadeh, **M.H. Moattar**. Robust microarray data feature selection using a correntropy based distance metric learning approach, *Computers in Biology and Medicine*, 2023. 161, 107056. <https://doi.org/10.1016/j.compbiomed.2023.107056>

- [19] M. Mollae, M. H. Moattar, A novel feature extraction approach based on ensemble feature selection and modified discriminant independent component analysis for microarray data classification, *Biocybernetics and Biomedical Engineering*, Vol. 36, pp. 521-529, 2016. DOI: <https://doi.org/10.1016/j.bbe.2016.05.001>
- [20] Shuichi Shinmura. *New Theory of Discriminant Analysis After R. Fisher*, *Advanced Research by the Feature Selection Method for Microarray Data*, Springer Singapore, 2016. <https://doi.org/10.1007/978-981-10-2164-0>
- [21] Ding, Y., et al. (2016). Image quality assessment method based on nonlinear feature extraction in kernel space. *Frontiers of Information Technology & Electronic Engineering*, 17(10): p. 1008-1017.
- [22] Feng, T.; Shen, Y.; Wang, F. Independent Component Extraction from the Incomplete Coordinate Time Series of Regional GNSS Networks. *Sensors* 2021, 21, 1569. <https://doi.org/10.3390/s21051569>
- [23] Kim, M.-C.; Lee, J.-H.; Wang, D.-H.; Lee, I.-S. Induction Motor Fault Diagnosis Using Support Vector Machine, Neural Networks, and Boosting Methods. *Sensors* 2023, 23, 2585. <https://doi.org/10.3390/s23052585>
- [24] Lin, J. and Chen, Q. (2014). A novel method for feature extraction using crossover characteristics of nonlinear data and its application to fault diagnosis of rotary machinery. *Mechanical Systems and Signal Processing*, 48(1): p. 174-187.
- [25] Leon-Medina, J.X.; Anaya, M.; Tibaduiza, D.A. Locally Linear Embedding as Nonlinear Feature Extraction to Discriminate Liquids with a Cyclic Voltammetric Electronic Tongue. *Chem. Proc.* 2021, 5, 56. <https://doi.org/10.3390/CSAC2021-10426>
- [26] Zhang, Z. and Zha, H. (2004). Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM journal on scientific computing*, 26(1): p. 313-338.
- [27] Kim, K. and Lee, J. (2014). Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. *Pattern Recognition*, 47(2): p. 758-768.
- [28] Orsenigo, C. and Vercellis, C. (2013). A comparative study of nonlinear manifold learning methods for cancer microarray data classification. *Expert systems with Applications*, 40(6): p. 2189-2197.
- [29] A. Jalali Mojahed, M. H. Moattar, H. Ghaffari, Supervised Density-Based Metric Learning Based on Bhattacharya Distance for Imbalanced Data Classification Problems. *Big Data Cogn. Comput.* 2024, 8, 109. <https://doi.org/10.3390/bdcc8090109>
- [30] Pavlinek, M., & Podgorelec, V. (2017). Text classification method based on self-training and LDA topic models. *Expert Systems with Applications*, 80, 83-93.

- [31] Kaznowska, E., Depciuch, J., Łach, K., Kołodziej, M., Koziarowska, A., Vongsvivut, J., ... & Cebulski, J. (2018). The classification of lung cancers and their degree of malignancy by FTIR, PCA-LDA analysis, and a physics-based computational model. *Talanta*, 186, 337-345.
- [32] Zhong, F. and Zhang, J. (2013). Linear discriminant analysis based on L1-norm maximization. *IEEE Transactions on Image Processing*, 22(8): p. 3018-3027.
- [33] Zhang, D., Li, X., He, J., & Du, M. (2018). A new linear discriminant analysis algorithm based on L1-norm maximization and locality preserving projection. *Pattern Analysis and Applications*, 21(3), 685-701.
- [34] Yang, J., et al. (2005). Two-dimensional discriminant transform for face recognition. *Pattern recognition*, 38(7): p. 1125-1129.
- [35] Li, C.N., Shao, Y.H. and Deng, N.Y. (2015). Robust L1-norm two-dimensional linear discriminant analysis. *Neural Networks*, 65: p. 92-104.
- [36] Xu, J., Gu, Z., and Xie, K. (2016). Fuzzy Local Mean Discriminant Analysis for Dimensionality Reduction. *Neural Processing Letters*, 44(3): p. 701-718.
- [37] Qu, L.; Pei, Y. A Comprehensive Review on Discriminant Analysis for Addressing Challenges of Class-Level Limitations, Small Sample Size, and Robustness. *Processes* 2024, 12, 1382. <https://doi.org/10.3390/pr12071382>
- [38] Dey, A., & Ghosh, M. (2019). A novel approach to fuzzy-based facial feature extraction and face recognition. *Informatica*, 43(4).
- [39] Ma, M., Deng, T., Wang, N., & Chen, Y. (2019). Semi-supervised rough fuzzy Laplacian Eigenmaps for dimensionality reduction. *International Journal of Machine Learning and Cybernetics*, 10(2), 397-411.
- [40] Sun, Y., & Lin, C. M. (2021). Design of Multidimensional Classifiers using Fuzzy Brain Emotional Learning Model and Particle Swarm Optimization Algorithm. *Acta Polytechnica Hungarica*, 18(4), 25-45.
- [41] Dey A., Chowdhury S., and Sing J.K. (2022). A new fuzzy and Gaussian distribution induced two-directional inverse FDA for feature extraction and face recognition, *International Journal of Advanced Intelligence Paradigms*, 22(1-2): pp 148-166.
- [42] Ghosh M., and Dey A. (2022) Fractional-weighted entropy-based fuzzy G-2DLDA algorithm: a new facial feature extraction method, *Multimedia Tools and Applications*.
- [43] Chen, C., & Zhou, X. (2022). Collaborative representation-based fuzzy discriminant analysis for Face recognition. *The Visual Computer*, 38(4), 1383-1393.

- [44] Gurubelli, Y., Ramanathan, M., & Ponnusamy, P. (2019). Fractional fuzzy 2DLDA approach for pomegranate fruit grade classification. *Computers and Electronics in Agriculture*, 162, 95-105.
- [45] Zhang, X., Zhu, Y., & Chen, X. (2017). Fuzzy 2d-lda face recognition based on sub-image. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 326-334). Springer, Cham.
- [46] Dey, A., Chowdhury, S., & Sing, J. K. (2018). Feature Extraction Using Fuzzy Generalized Two-Dimensional Inverse LDA with Gaussian Probabilistic Distribution and Face Recognition. In *Advanced Computational and Communication Paradigms* (pp. 553-561). Springer, Singapore.
- [47] Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. 2013: Academic press.