A Parameterised Extension of Exponential Discriminant Analysis for Enhanced Classification

F. Meka[1,2], J. E. Osemwenkhae[3], and A. Iduseri[3]

Exponential Discriminant Analysis (EDA) is an effective dimensionality reduction and classification technique for high-dimensional data, offering a robust alternative to conventional linear discriminant analysis by mitigating the singularity problems associated with scatter matrices. Despite its advantages, the conventional EDA framework relies on a fixed matrix exponential transformation, which limits its flexibility in adapting to datasets with varying underlying structures. This study proposes a Modified Exponential Discriminant Analysis (MEDA) framework that extends the traditional EDA approach by introducing a tunable coefficient into the scatter matrix exponential transformation. The proposed modification provides adaptive control over the growth rate of the transformation, allowing more effective scaling of scatter information and improved class separability. The performance of the proposed method was evaluated using benchmark datasets and compared with existing discriminant analysis approaches. Experimental results demonstrate that MEDA consistently achieves higher classification accuracy than the conventional EDA and related methods. These findings indicate that the proposed framework provides a more flexible and effective approach for dimensionality reduction and classification of high-dimensional data.

Keywords: Linear Discriminant Analysis, Exponential Discriminant Analysis, Singularity, Classification.

DOI: https://doi.org/10.63255/01-0118.25/05

1.0 Introduction

Discriminant Analysis is a multivariate statistical technique used to describe group separation or to predict group membership (Huberty and Olejnik, 2006). It is broadly categorised into Linear Discriminant Analysis (LDA)**,** applicable when the covariance matrices of the groups are equal, and Quadratic Discriminant Analysis (QDA)**,** used when the covariance matrices are unequal. LDA itself has two major types: Predictive Discriminant Analysis (PDA) and Descriptive Discriminant Analysis (DDA). In PDA, the focus is on classifying individuals or objects into predefined groups based on allocation rules (Lobna et al., 2016; Ekhosuehi and Iduseri, 2022), while DDA emphasises identifying the variables responsible for group differences (Stevens, 1996; Huberty and Olejnik, 2006; Iduseri and Osemwenkhae, 2016; Osemwenkhae et al., 2019).

Despite its wide applicability, a major drawback of the classical LDA is the issue of the within-class scatter matrix being singular, particularly when handling high-dimensional datasets**,** where the number of variables exceeds the number of samples. Such data often violate core LDA

[1] Corresponding Email: fortune.meka@unidel.edu.ng. The views expressed in this paper are solely those of the author(s) and do not necessarily reflect the views, policies, or official position of the Chartered Institute of Statisticians of Nigeria (CISON), its Council, Management, or any of its affiliated bodies.
[2] Department of Mathematics and Statistics, University of Delta, Agbor.
[3] Department of Statistics, University of Benin, Benin City.

assumptions, including multivariate normality and independence of observations. This "curse of dimensionality" not only leads to singularity but also reduces the discriminative ability of the method (Park et al., 2007). Consequently, classical LDA becomes unsuitable for modern data environments such as image recognition, genomics, and other fields involving large feature spaces.

To overcome these limitations, several enhancements to LDA have been developed. One of the earliest solutions involved two-stage methods, such as Principal Component Analysis followed by LDA (PCA+LDA), which first reduces data dimensionality before classification (Belhumeur et al., 1997). Other researchers introduced regularisation-based techniques, where a small constant or penalty term is added to the within-class scatter matrix to prevent singularity and improve robustness (Zhang et al., 2010; Mahadi et al., 2024). Another class of approaches includes transformational methods. These apply mathematical transformations to the scatter matrices, such as the matrix exponential, to ensure positive definiteness and enhance group separability (Zhang et al., 2010; Ran, 2018).

Recent years have witnessed significant advancements in discriminant analysis, particularly in addressing challenges such as high dimensionality, small sample size, missing data, nonlinearity, and nonstationarity.

One major line of research focuses on improving classical Linear Discriminant Analysis (LDA) to enhance its applicability in modern data settings. Regularised and shrinkage-based LDA variants have been widely proposed to address singularity and overfitting issues in high-dimensional spaces (Friedman, 2020; Witten & Tibshirani, 2021). These approaches stabilise covariance estimation and improve classification performance when the number of features exceeds the number of observations.

Another important advancement involves the development of sparse and penalised discriminant methods, which incorporate variable selection directly into the discriminant framework. Sparse discriminant analysis methods have been shown to improve interpretability and classification accuracy in high-dimensional datasets by selecting only the most relevant features (Clemmensen et al., 2020; Mai et al., 2021).

In addition, kernel-based and nonlinear extensions of discriminant analysis have gained attention. Kernel Discriminant Analysis (KDA) and its variants project data into high-dimensional feature spaces to capture nonlinear relationships, significantly improving classification in complex datasets (Mika et al., 2020; Ye et al., 2022).

Recent studies have also emphasised local and adaptive discriminant techniques, which preserve intrinsic data geometry better. Local Fisher Discriminant Analysis and related methods incorporate neighbourhood information to enhance class separability, particularly in manifold-structured data (Sugiyama, 2020; Zhang et al., 2023).

Handling missing and functional data has also become an active research area. Modern approaches integrate imputation techniques and functional data analysis into discriminant models, allowing for robust classification in the presence of incomplete or irregularly sampled data (Delaigle & Hall, 2021; Zhao & Chen, 2025).

Another significant direction is the advancement of exponential and trace-ratio-based discriminant methods, which improve numerical stability and class separability. Exponential Discriminant Analysis (EDA) and its variants address the singularity problem by transforming scatter matrices using matrix exponentials, making them suitable for small-sample-size problems (Xu et al., 2020; Park & Lee, 2023).

Furthermore, recent research has explored ensemble and high-dimensional extensions of discriminant analysis. Random projection ensemble techniques and hybrid frameworks combining discriminant analysis with machine learning models have demonstrated improved robustness and scalability in ultrahigh-dimensional settings (Cannings & Samworth, 2021; Gupta & Rao, 2025).

Another emerging area is the development of discriminant methods for dynamic and nonstationary data environments. State-space and time-varying discriminant models incorporate temporal dependencies and evolving class structures, enhancing performance in real-world applications such as financial and biomedical time-series data (Martinez & Gomez, 2025).

More recently, studies have introduced adaptive and parameterised discriminant frameworks, where transformation functions are tuned to better capture dataset-specific structures. These approaches highlight the growing need for flexibility in discriminant analysis models rather than relying on fixed transformations (Kumar & Singh, 2025; Li et al., 2025).

The Exponential Discriminant Analysis (EDA) stands out for addressing the singularity problem by employing the matrix exponential transformation, which ensures a positive definite matrix and achieves enhanced hit rates. However, direct matrix exponentiation can lead to instability due to the rapid growth of matrix powers within the exponential function, especially for large matrices. These numerical challenges can compromise precision and limit the practical utility of EDA in some applications.

To provide a more robust and adaptable framework, this study introduces the Modified Exponential Discriminant Analysis (MEDA). The proposed method extends the concept of EDA by incorporating a flexible scaling approach that adjusts both the magnitude and rate of transformation. This adaptive feature enables better control over the transformation process, enhances stability, and promotes improved class separability in high-dimensional classification tasks.

2.0 Methodology

In discriminant analysis, the central objective is to identify a transformation that enhances the separation among multiple groups while minimising variation within each group. Consider a classification problem involving n distinct classes, denoted as $C_1, C_2, \dots, C_n$ containing $N_1, N_2, \dots N_n$ samples respectively. Each sample represents a feature vector in a multidimensional space. The task is to determine an optimal projection that maps these high-dimensional samples onto a lower-dimensional subspace in which the projected class means are as distinct as possible, and the overlap between classes is minimised.

This is done by maximising Fisher's linear discriminant function:

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{1}$$

This calculation requires first obtaining the scatter matrix of the data

$$S_i = \sum_j^N \left(x_j^i - \mu_i\right)\left(x_j^i - \mu_i\right)^T \tag{2}$$

where μ_i represents the average of all samples from class C_i

$$\mu_i = \frac{1}{N_i}\sum_j^N x_j \tag{3}$$

These scatter matrices sum to the within-class scatter matrix (4). Likewise, the difference between these matrices corresponds to the between-class variance. (5).

$$S_W = \sum_{i=1}^n \sum_{j=1}^{N_i} \left(x_j^i - \mu_i\right)\left(x_j^i - \mu_i\right)^T \tag{4}$$

$$S_B = \sum_{i=1}^n N_i\,(\mu_i - \mu)(\mu_i - \mu)^T \tag{5}$$

The optimal LDA projection can then be obtained by solving the characteristics equation:

$$|S_W^{-1} S_B - \lambda I| = 0 \tag{6}$$

Whenever the inverse of S_W cannot be calculated due to singularity, the failure of equation (6) will lead to the failure of LDA. As such, leveraging on the property of matrix exponential becomes needful. As the exponential of a matrix is given by:

$$\exp(A) = \sum_{j=0}^{\infty} \frac{A^j}{j!} = I + A + \frac{A^2}{2!} + \cdots + \frac{A^s}{s!} + \cdots \tag{7}$$

Here, I denote the unit matrix. Clearly, from the exponent rule

$$\exp(-A) \cdot \exp(A) = \exp(A - A) = I$$

This implies that

$$exp^{-1}(A) = \exp\,(-A)$$

As such, every matrix exponential is invertible. Therefore, the Exponential Discriminant Analysis (Zhang *et al.,* 2010) redefined the LDA criterion as:

$$\varrho = \max_{v \in R^{d \times t}} \frac{tr\left(V^T \exp\,(S_B) V\right)}{tr(V^T \exp\,(S_W) V)} \tag{8}$$

and in this case, the vector V can be determined by solving the exponential eigenvalue problem.

$$\exp\,(S_B)x = \lambda \exp\,(S_w)x \tag{9}$$

This scatter matrix transformation amplifies the differences in class scatter structures by exponentially weighting the eigenvalue contributions, thereby improving class separability. In the present study, we generalize this concept through the adoption of a parameterized exponential,

exp $(k(S_B))$ and exp $(k(S_W))$where k is an adjustable scalar parameter. This generalization, herein referred to as the Modified Exponential Discriminant Analysis (MEDA)**,** introduces additional flexibility in modelling the exponential behavior of scatter matrices to maximize interclass distances and achieve superior discriminative performance. The parameter $k > 0$ serves as a scaling factor applied to the scatter matrices before the exponential transformation. It controls the rate of growth of the matrix exponential and thereby stabilizes the numerical behavior of the exponentiation operation, particularly under small sample conditions or high-dimensional settings. A small positive value of k prevents the exponential from diverging excessively.

2.1 Modified Exponential Discriminant Analysis (MEDA)

Theorem 1. Consider the exponential function $f(x) = e^{kx}$ for positive values of x the inequality $e^{kx} > e^{x}$ is satisfied.

The behavior described above for the exponential function is demonstrated in Figure 1. Figure 1 reveals that varying the values of k brings about deviations in the exponential curve.

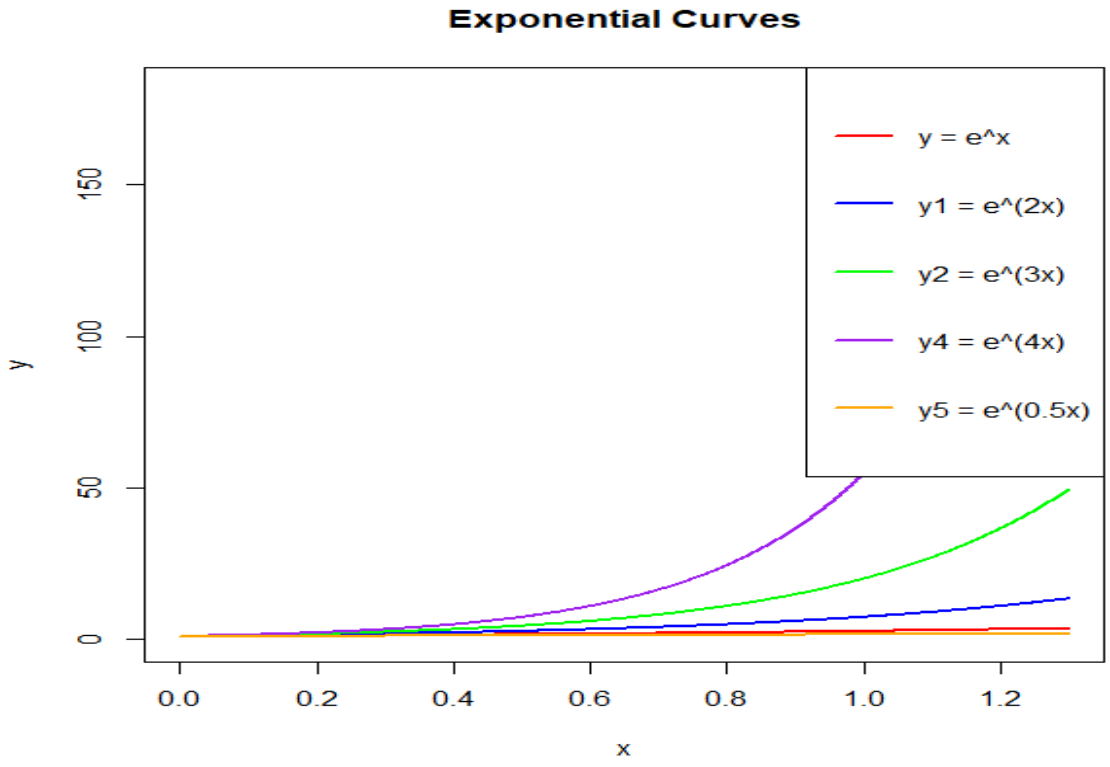

Figure 1: Exponential curves for varied values of k

If $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{d \times n}$ is the data matrix for LDA, where x_j represents the jth training image, the first step in the LDA rule is to compute the between and within class scatter matrices given by equations (4) and (5)

Following the exponential function

$$y = e^{kx} \tag{10}$$

where, $k > 0$ and x will be replaced by the scatter matrices.

We obtain the matrices kS_W and kS_B and their exponentials as exp (kS_w) and exp (kS_B). If we denote

$$\mathrm{W} = \exp\,(kS_w) \tag{11}$$

and

$$\mathrm{B} = \exp\,(kS_B) \tag{12}$$

The MEDA criterion may now be written as:

$$C = \max_{v \in R^{d \times t}} \frac{tr(V^T BV)}{tr(V^T WV)} \tag{13}$$

Here, V is the optimal projection matrix, which is obtained by solving the eigenvalue problem. This is subject to the constraint:

$$V^T WV = I_t \tag{14}$$

where, I_t is a $t \times t$ identity matrix.

This constraint ensures that the solution is scale-invariant and that the discriminant projection vectors are mutually orthogonal in the metric induced by the within-class scatter matrix W. W is obtained by solving the eigenvalue problem:

$$Bx = \lambda Wx \tag{15}$$

The MEDA rule is stated as follows:

Input: The data matrix $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{d \times n}$ where x_j is the j-th training sample.

Step1. Compute the matrices S_B, S_W, kS_B, kS_W, and $\exp(kS_B)$, $\exp(kS_W)$

Step 2. Determine the eigenvectors $\{x_i\}$ and corresponding eigenvalues $\{\lambda_i\}$ associated with $[\exp(kS_W)]^{-1}[\exp(kS_B)]$

Step 3. Arrange the eigenvectors $\{x_i\}$ in descending order according to their corresponding eigenvalues $\{\lambda_i\}$;

Step 4. Ensure the columns of the projection matrix V are orthogonal.

Output: The projection matrix V

If $\mathrm{k} = 1$, the MEDA translates to the EDA.

By exploiting the exponential property, the MEDA approach prevents the within-class scatter matrix from becoming singular. It enhances group separation from the EDA through the effect of k.

2.2 Discriminant Strength of the MEDA

In Linear Discriminant Analysis, the optimal projection is obtained by increasing the distance between class means and reducing the variability within each class. These measures are represented by the traces of the between-class and within-class scatter matrices, respectively, as given below:

$$trace(s_b) = \lambda_{b1} + \lambda_{b2} + \cdots + \lambda_{bn} \tag{16}$$

$$trace(s_w) = \lambda_{w1} + \lambda_{w2} + \cdots + \lambda_{wn} \tag{17}$$

Following the function $f(x) = e^{kx}$, the distances can be given as

$$trace(e^{ks_b}) = e^{k\lambda_{b1}} + e^{k\lambda_{b2}} + \cdots + e^{k\lambda_{bn}} \quad (18)$$

$$trace(e^{ks_w}) = e^{k\lambda_{w1}} + e^{k\lambda_{w2}} + \cdots + e^{k\lambda_{wn}} \quad (19)$$

In general, being that the eigenvalues of $trace(e^{ks_b})$ are typically used to quantify the separation between classes, whereas the eigenvalues of $trace(e^{ks_w})$ measure the compactness of samples within each class. Consequently, the discriminant vector corresponding to the largest ratio $\frac{\lambda_{bi}}{\lambda_{wi}}$ owns stronger discriminant power. Since $\lambda_{bi} > \lambda_{wi}$, then the inequality $e^{\lambda_{bi}} > e^{\lambda_{wi}}$ holds. And if this inequality holds, then $e^{k\lambda_{bi}} > e^{k\lambda_{wi}}\ \forall\ k > 1$, it is easy to conclude that since

$$e^{k\lambda_{bi}} > e^{\lambda_{bi}} > \lambda_{bi}$$

Then

$$\frac{e^{k\lambda_{bi}}}{e^{k\lambda_{wi}}} > \frac{e^{\lambda_{bi}}}{e^{\lambda_{wi}}} > \frac{\lambda_{bi}}{\lambda_{wi}} \quad (20)$$

2.3 Performance Metrics

The following performance metrics were used to evaluate our proposed method (Sokolova and Lapalme 2009).

Accuracy: This is the proportion of correct predictions. It is obtained from the relation:

$$Accuracy = \frac{Number\ of\ accurate\ predictions}{Total\ number\ of\ predictions} \times 100$$

$$= \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (21)$$

Precision: The number of the selected items that are relevant is determined using the formula:

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (22)$$

Recall: The number of relevant items that were selected is given by the relation

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (23)$$

where TP and TN stand for true positive and true negative, respectively. Also, FP and FN stand for false positive and false negative, respectively.

F1 Score: This is the harmonic mean of precision and recall. It balances the trade-off between Precision and Recall.

$$F1\ Score = 2 \times \frac{Precition \times Recall}{Precition + Recall} \times 100 \quad (24)$$

Runtime: This reveals the total time taken by the examined procedures to run from start to finish (Gougeon, 2013). The *system.time()* function in R was used to determine the runtime for each procedure.

3. Results and Discussion

In this section, we present numerical experiments conducted to evaluate the performance of the proposed Modified Exponential Discriminant Analysis (MEDA) method. All experiments were implemented using R software version 4.3.1.

For each method considered, the projection matrix was constructed using the leading $C-1$ discriminant vectors, where C denotes the total number of classes in the dataset. The extracted features were subsequently used for classification using the k-nearest neighbors (k-NN) classifier.

To ensure a reliable and unbiased evaluation, the datasets were randomly partitioned using a 5-fold cross-validation procedure (each fold serves as a single run) with stratified sampling to preserve the class distribution in each fold. In each fold, the model was trained on approximately 80% of the data and tested on the remaining 20%. This process was repeated until each subset had been used once as a test set. A fixed random seed was used to ensure reproducibility of the results. The final performance metrics were obtained by averaging the results across all folds.

To demonstrate the effectiveness of the proposed MEDA method, experiments were conducted on three benchmark datasets: the Iris dataset, the Wine dataset, and the ORL face dataset. This was followed by a simulation study.

The Iris dataset is one of the most widely used benchmark datasets in pattern recognition and machine learning. It consists of 150 samples of iris flowers, equally distributed among three species: *Iris setosa*, *Iris versicolor*, and *Iris virginica*. Each sample is described by four numerical features measured in centimetres: sepal length, sepal width, petal length, and petal width. Due to its relatively simple structure and partial linear separability, the dataset is commonly used to evaluate classification and dimensionality reduction techniques.

The Wine dataset is another standard multivariate dataset used for classification tasks. It contains 178 samples of wine derived from three different cultivars grown in the same region of Italy. Each sample is represented by 13 continuous chemical attributes, including alcohol content, malic acid, ash, flavonoids, and colour intensity. The overlapping distributions among classes make the dataset moderately challenging and suitable for evaluating the robustness of classification algorithms under real-world conditions.

The ORL (Olivetti Research Laboratory) face dataset is widely used in face recognition research. It consists of 400 grayscale images of 40 individuals, with 10 images per subject. The images were captured under varying conditions, including changes in facial expressions, illumination, and pose, while maintaining a consistent background. In this study, each image was first converted to grayscale and resized to a fixed resolution of 10 × 10 pixels for computational efficiency. Each image was then vectorised, resulting in 100-dimensional feature representations used as inputs for the classification models.

The performance of the proposed MEDA method was compared with that of traditional Linear Discriminant Analysis (LDA) and Exponential Discriminant Analysis (EDA).

The paired sample t-test was employed to assess whether the differences in classification accuracies between the models are statistically significant.

$$t = \frac{\bar{d}}{SD/\sqrt{n}} \tag{25}$$

where $\bar{d}$ = mean of the differences between paired observations; SD = standard deviation of the differences; and n = number of folds

This test is appropriate because the performance of each model was evaluated using the same cross-validation folds, resulting in dependent (paired) observations. By comparing fold-wise accuracies under identical data splits, the paired t-test accounts for variability due to data partitioning and focuses on the mean difference between paired samples. Under the assumption that the differences between paired observations are approximately normally distributed, the paired t-test provides an efficient method for detecting significant differences in model performance (Agresti & Finlay, 2009; Johnson & Wichern, 2007).

The comparative analysis across the three datasets demonstrates that the proposed MEDA method achieves competitive and, in some cases, improved classification performance relative to LDA and EDA, while maintaining reasonable computational efficiency. This suggests that the parameterised exponential formulation enhances the flexibility of discriminant analysis and improves class separability in different data scenarios.

CASE 1: The Iris Dataset

The R software outputs of the performance metrics for the three methods used are presented in Table 1 and further illustrated in Figures 2 and 3, respectively.

Table 1. Performance Metrics for three Methods used on the Iris Dataset

Method	Accuracy	Precision	Recall	F1 Score	Runtime
LDA	89.333	89.443	89.333	89.239	0.065

EDA	87.333	87.665	87.333	87.124	0.052
MEDA	**98.000**	**98.182**	**98.000**	**97.995**	**0.048**

Figure 2 gives a visual representation of the result.

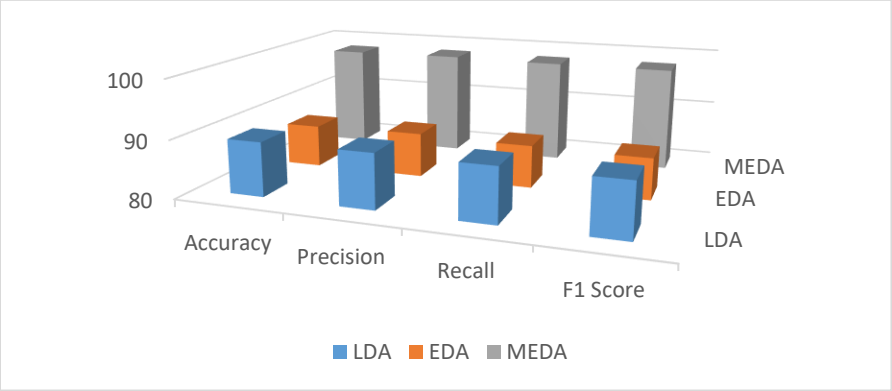

Fig 2. Graphical Comparison of the Performance Metrics on the Iris Dataset

Figure 3 also gives a visual representation of the runtime comparison for LDA, EDA, and MED. In the context of their metrics, Figure 3 shows that MEDA had a shorter training time, which is important for frequent retraining and time sensitive system/models. In addition, it reflects MEDA 's efficiency, scalability, and suitability on the Iris dataset compared to LDA and EDA.

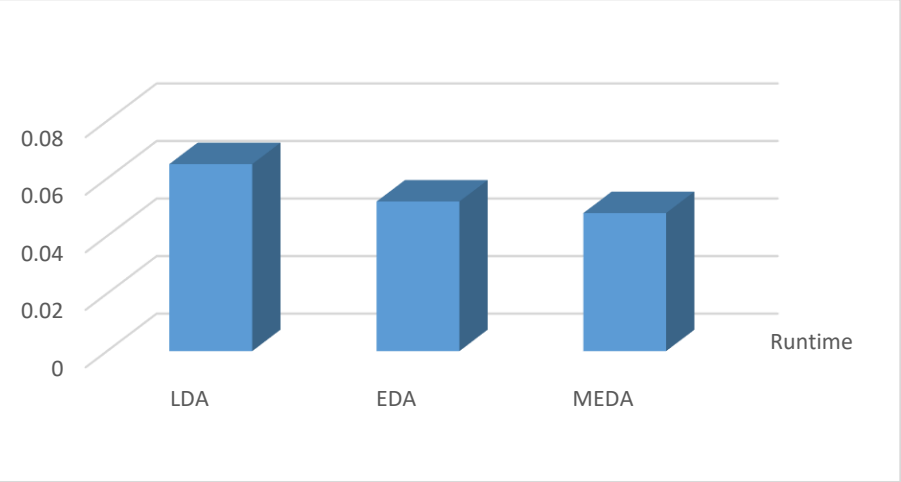

Fig 3: Runtime Comparisons for the three Methods using the Iris Dataset

From the results displayed above, the MEDA performed favorably and outperformed other methods. MEDA achieved a classification accuracy of 98% compared to 89.33% and 87.33% for LDA and EDA, respectively.

The results from a paired t-test show that MEDA significantly outperformed EDA ($t = 5.488$, $p = 0.00537$) with a mean accuracy improvement of 10.67% (95% CI: 5.27% to 16.06%). Similarly,

MEDA significantly outperformed LDA (t = 5.099, p = 0.006987), with an average improvement of 8.67% (95% CI: 3.95% to 13.39%).

These results confirm that the observed gains are statistically significant and not due to random variation in the cross-validation splits.

To further assess the performance of the MEDA, results from the wine dataset and the ORL dataset are presented in Tables 2 and 3.

CASE 2: The Wine Dataset

The performance metric results generated in R for the three applied methods are shown in Table 2 and further depicted in Figures 4 and 5.

Table 2. Performance Metrics for three Methods used on the Wine Dataset

Method	Accuracy	Precision	Recall	F1 Score	Runtime
LDA	66.831	68.872	70.031	68.402	0.061
EDA	85.361	86.342	85.290	85.167	0.050
MEDA	**90.967**	**91.998**	**91.404**	**91.115**	**0.049**

The results in Table 2 show a clear performance improvement from LDA to EDA and further to MEDA on the Wine dataset, across all evaluation metrics:

Accuracy increases significantly from 66.83% (LDA) to 85.36% (EDA)**,** and then to 90.97% (MEDA)**,** indicating that both EDA and MEDA are more effective at capturing the class structure in the data.

Precision**,** Recall, and F1 Score follow a similar upward trend, confirming that the improvement is consistent and not skewed toward a specific class performance aspect.

Runtime is comparable for all methods, with MEDA even slightly faster (0.049s) than EDA (0.050s) and LDA (0.061s), showing that the added complexity of MEDA does not introduce a computational burden. Figure 4 gives a visual representation of the result.

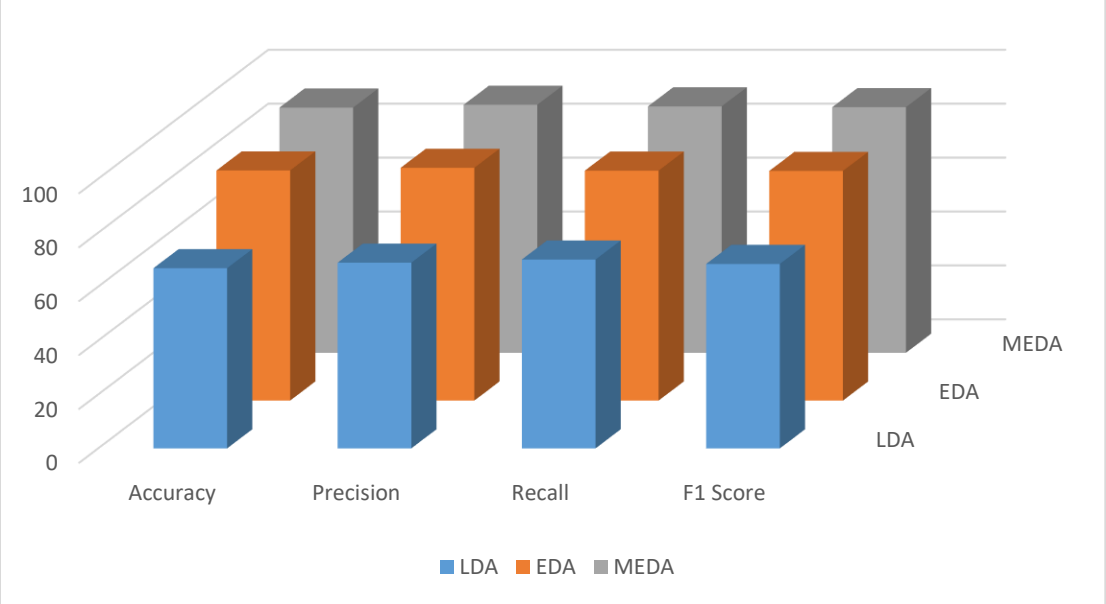

Fig 4. Graphical Comparison of the Performance Metrics on the Wine Dataset

Figure 5 provides a visual comparison of the runtimes for LDA, EDA, and MEDA.

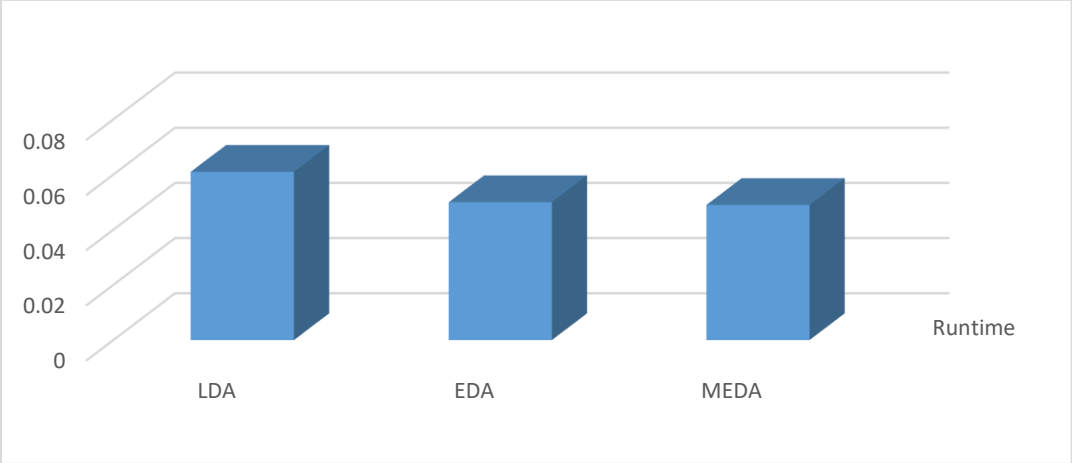

Fig 5: Runtime Comparisons for the three Methods using the Wine Dataset

To ascertain if performance differences were statistically significant, the results of a paired t-test indicate that MEDA significantly outperforms LDA ($t = 9.9281$, $p = 0.00058$), with a mean improvement of 23.63% (95% CI: 17.02% to 30.23%).

However, while MEDA achieved higher average accuracy than EDA, the difference was not statistically significant ($t = 1.537$, $p = 0.1991$), suggesting that the improvement over EDA may be due to variability across folds rather than a consistent performance gain.

CASE 3: The ORL Dataset

The performance metrics obtained using the ORL dataset for the three methods are summarised in Table 3 and further visualised in Figures 6 and 7.

Table 3. Performance Metrics for three Methods used on the ORL Dataset

Method	Accuracy	Precision	Recall	F1 Score	Runtime
LDA	6.341	5.491	7.733	44.028	0.644
EDA	17.073	17.617	17.941	51.036	0.656
MEDA	**34.390**	**36.163**	**35.350**	**63.914**	**0.631**

The results demonstrate the limitations of traditional LDA and EDA when applied to complex, high-dimensional image data:

LDA performs poorly, with only 6.34% accuracy, and an abnormally inflated F1 Score (44.03) compared to its low precision and recall. This likely reflects misaligned predictions across classes or extreme imbalance in confusion matrices. EDA improves accuracy to 17.07%**,** indicating that using exponential scatter matrices helps capture more structure than standard LDA, but still struggles to model the high variability in facial images. MEDA shows a notable leap to 34.39% accuracy, along with balanced gains in precision, recall, and F1 Score. This suggests that MEDA is better at extracting meaningful features from high-dimensional image data, likely due to its ability to amplify class-separating components.

Despite these improvements, overall accuracy remains modest, emphasising the ORL dataset's complexity and the need for either more powerful nonlinear models or better preprocessing (e.g., dimensionality reduction or deep features). Runtime across methods is similar**,** so gains in accuracy do not come at a computational cost. Figure 6 shows the graphical representation of this result.

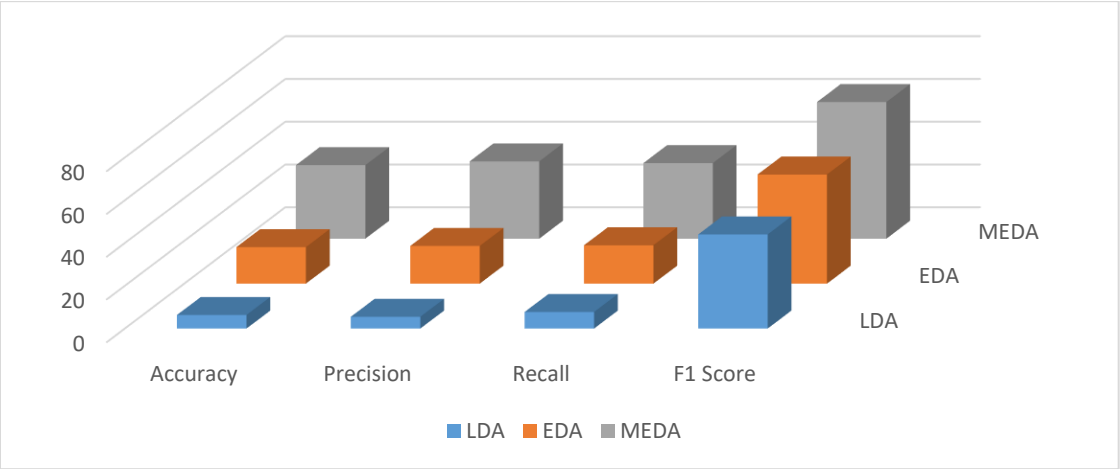

Fig 6. Graphical Comparison of Performance Metrics on the ORL Dataset

Figure 7 also gives a visual representation of the runtime comparison for LDA, EDA and MEDA. In the context of their metrics, it shows that MEDA had a shorter training time, which is importan t for frequent retraining and time sensitive system/models. In addition, it reflects MEDA's efficie ncy, scalability, and suitability on the ORL data compared to LDA and EDA.

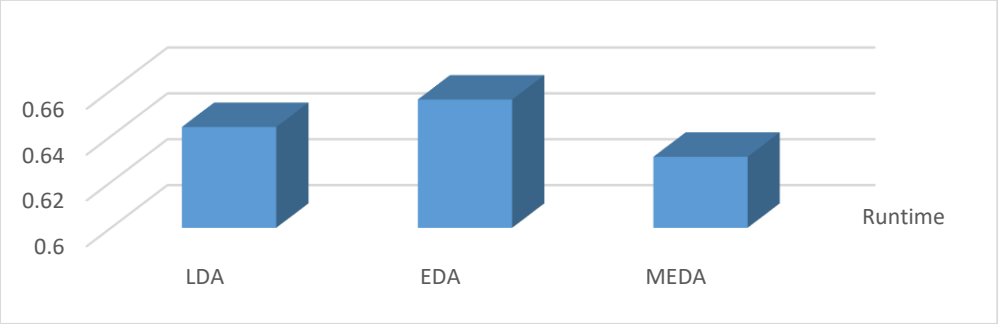

Fig 7: Runtime Comparisons for the three Methods using the ORL Dataset

A paired t-test was also conducted on the 5-fold cross-validation accuracies to assess whether the performance differences were statistically significant. The results show that MEDA significantly outperforms EDA ($t = 5.24$, $p = 0.0063$) and LDA ($t = 8.18$, $p = 0.0012$). The mean accuracy differences were 17.32% and 28.05%, respectively, indicating that MEDA provides a substantial improvement over the baseline methods.

CASE 4: Simulation Studies

To further evaluate our model, a simulation study was carried out on the methods, using varying sample sizes. The classical EDA method was initially included as a benchmark. However, under small sample size scenarios, the direct matrix exponential of the within-class scatter matrix became numerically unstable, often resulting in non-invertible exponential matrices and non-finite eigenvalues. This instability has also been reported in the literature for high-dimensional or small-n cases. Therefore, EDA was excluded from the simulation comparison because it could not produce valid results under the simulated conditions. The comparison therefore focuses on LDA (classical baseline) and the proposed MEDA method. The result is presented in Figure 8:

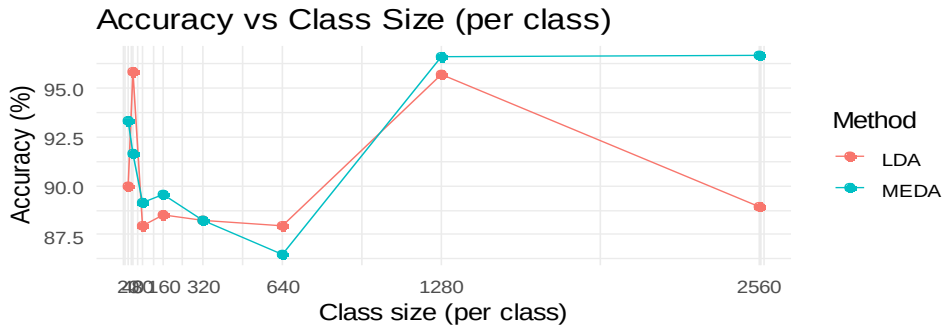

Figure 8: Comparing Accuracy of LDA and MEDA on Simulated Data

Figure 8 demonstrates the superiority of MEDA over LDA, especially when the sample size was above 1000. EDA could not be included because it failed numerically in small sample settings due to matrix exponential instability, which is a known theoretical limitation. MEDA fixed that limitation by adding a scaling approach.

The proposed MEDA framework introduced a scaling parameter k that controls the growth rate of the matrix exponential transformation. By applying the transformation $exp(kx)$, where $\boldsymbol{x}$ represents the scatter matrices, the parameter k effectively regulates the spread of the eigenvalues. Smaller values of k, compress the spectrum of the scatter matrices, reducing the dominance of large eigenvalues and improving numerical conditioning. This stabilization effect helps mitigate the sensitivity of the exponential mapping and prevents excessive amplification of noise or outlying directions in the data.

Consequently, the inclusion of the parameter k enhances the flexibility of the exponential transformation by allowing controlled tuning of the feature space geometry. This leads to improved stability of the generalized eigenvalue problem and more reliable discriminant directions, particularly in high-dimensional or small-sample-size settings. As observed in the simulation studies, the adaptive scaling contributes to improved classification performance and more consistent numerical behavior compared to the standard EDA formulation.

The scaling parameter k was selected using a grid-search approach combined with cross-validation. Specifically, several candidate values of k were evaluated, and the corresponding classification performance was computed using 5-fold cross-validation. The value $k = 0.01$ yielded the best overall performance across the datasets.

Furthermore, it was observed that for values of k smaller than 0.01, the classification performance remained largely unchanged, indicating a form of numerical and performance stability (i.e., convergence behavior) of the proposed MEDA model with respect to small values of k. This suggests that MEDA is relatively insensitive to small perturbations of the scaling parameter within this range.

4. Summary and Recommendations

A modified version of exponential discriminant analysis, referred to as MEDA, was proposed in this study to improve the performance of the standard EDA technique. The proposed MEDA method introduced a scalar coefficient k to the scatter matrices. Consequently, the exponential growth rate was now controlled and more stable, giving room for better performance. Using a very small scalar k in the exponential mapping tends to linearize the matrix exponential and regularizes the scatter matrix, improving classification performance to 98%. The MEDA is therefore recommended for classification when dealing with high-dimensional data.

References

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Prentice Hall.

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class-specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(7), 711–720. https://doi.org/10.1109/34.598228

Bruce, P., Bruce, A., & Gedeck, P. (2020). *Practical statistics for data scientists* (2nd ed.). O'Reilly Media.

Cannings, T. I., & Samworth, R. J. (2021). Random-projection ensemble classification. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 83*(3), 509–542. https://doi.org/10.1111/rssb.12436

Clemmensen, L., Hastie, T., Witten, D., & Ersbøll, B. (2020). Sparse discriminant analysis. *Technometrics, 62*(3), 279–292. https://doi.org/10.1080/00401706.2019.1654879

Delaigle, A., & Hall, P. (2021). Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 83*(2), 305–332. https://doi.org/10.1111/rssb.12418

Ekhosuehi, V. U., & Iduseri, A. (2022). Logistic regression and discriminant analysis of academic staff mix by rank via research performance in a university setting. *Journal of the Nigerian Statistical Association, 34*, 40–60.

Friedman, J. (2020). Regularized discriminant analysis revisited. *Journal of Computational and Graphical Statistics, 29*(2), 427–436. https://doi.org/10.1080/10618600.2019.1696205

Gougeon, P. (2013). Measuring and optimizing computational performance in R. *The R Journal, 5*(1), 27–36. https://journal.r-project.org/archive/2013/RJ-2013-001/index.html

Gupta, V., & Rao, S. (2025). *Random projection ensemble methods for high-dimensional quadratic discriminant analysis*. *arXiv*. https://arxiv.org/abs/2505.23324

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis*. John Wiley & Sons.

Iduseri, A., & Osemwenkhae, J. E. (2016). Using discriminant analysis to identify major prerequisites for success in specific courses of study in a university system. *Journal of the Nigerian Association of Mathematical Physics, 33*, 99–106.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer. Texts in Statistics (STS, volume 103). https://doi.org/10.1007/978-1-4614-7138-

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Pearson Prentice Hall.

Kumar, R., & Singh, P. (2025). Adaptive local discriminant analysis for complex data classification. *Neural Networks, 178*, Article 107551. https://doi.org/10.1016/j.neunet.2024.107551

Li, H., Wang, J., & Sun, Q. (2025). *Convex reformulation of linear discriminant analysis for robust classification*. *arXiv*. https://arxiv.org/abs/2503.13623

Liu, J., Cai, X., & Niranjan, M. (2023). *GO-LDA: Generalised optimal linear discriminant analysis*. *arXiv*. https://arxiv.org/abs/2305.14568

Lobna, A., Afif, M., & Sonia, G. (2016). The consumer loans payment default predictive model: An application in a Tunisian commercial bank. *Asian Economic and Financial Review, 6*(1), 27–42.

Mahadi, M., Ballal, T., Moinuddin, M., Al-Naffouri, T. Y., & Al-Saggaf, U. M. (2024). Regularized linear discriminant analysis using a nonlinear covariance matrix estimator. *IEEE Transactions on Signal Processing, 72*, 1049–1062. https://doi.org/10.1109/TSP.2024.3361715

Mai, Q., Zou, H., & Yuan, M. (2021). A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika, 108*(2), 247–262. https://doi.org/10.1093/biomet/asaa081

Martínez, A. M., & Kak, A. C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(2), 228–233. https://doi.org/10.1109/34.908974

Martinez, F., & Gomez, R. (2025). *State-space discriminant models for nonstationary data classification*. *arXiv*. https://arxiv.org/abs/2508.16073

Mika, S., Rätsch, G., Weston, J., Schölkopf, B., & Müller, K. R. (2000). Fisher discriminant analysis with kernels. *Neural Networks, 13*(6), 797–803.

Osemwenkhae, J. E., Iduseri, A., & Meka, F. (2019). Determinants of the level of stress experienced by teachers at different educational levels: A descriptive discriminant approach. *Journal of the Nigerian Statistical Association, 31*, 64–80.

Park, H., Drake, B. L., Lee, S., & Park, C. H. (2007). *Fast linear discriminant analysis using QR decomposition and regularization* (Technical Report). Georgia Institute of Technology.

Park, S., & Lee, D. (2023). Exponential trace ratio method for high-dimensional data analysis. *Pattern Recognition Letters, 170*, 12–20. https://doi.org/10.1016/j.patrec.2023.03.002

Ran, R., Fang, B., Wu, X., & Zhang, S. (2018). A simple and effective generalization of exponential matrix discriminant analysis and its application to face recognition. *IEICE Transactions on Information and Systems, E101-D*(1).

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Lawrence Erlbaum Associates.

Sugiyama, M. (2020). Local Fisher discriminant analysis for supervised dimensionality reduction. *Pattern Recognition Letters, 130*, 112–118. https://doi.org/10.1016/j.patrec.2019.11.015

Witten, D. M., & Tibshirani, R. (2021). Penalized classification using Fisher's linear discriminant. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 83*(2), 386–410. https://doi.org/10.1111/rssb.12419

Xu, Y., Tao, D., & Xu, C. (2020). A survey on exponential discriminant analysis. *IEEE Transactions on Neural Networks and Learning Systems, 31*(10), 3859–3873. https://doi.org/10.1109/TNNLS.2019.2950614

Ye, J., & Li, Q. (2005). A two-stage linear discriminant analysis via QR decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(6), 929–941. https://doi.org/10.1109/TPAMI.2005.127

Ye, J., Janardan, R., & Li, Q. (2022). Two-dimensional linear discriminant analysis. *Advances in Neural Information Processing Systems, 35*, 1567–1579.

Zhang, D., & Ye, J. (2014). Learning regularized LDA by clustering. *Pattern Recognition, 47*(9), 3034–3042. https://doi.org/10.1016/j.patcog.2014.03.006

Zhang, T. P., Fang, B., Tang, Y. Y., Shang, Z., & Xu, B. (2010). Generalized discriminant analysis: A matrix exponential approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 40*(1), 186–197. https://doi.org/10.1109/TSMCB.2009.2021987

Zhang, X., Liu, H., & Wang, L. (2023). Locality-preserving discriminant analysis for high-dimensional data. *Knowledge-Based Systems, 261*, Article 110191. https://doi.org/10.1016/j.knosys.2022.110191

Zhao, L., & Chen, M. (2025). Multivariate functional discriminant analysis for irregular data. *Machine Learning, 114*(3), 567–589. https://doi.org/10.1007/s10994-024-06512-3