A DUAL EMPIRICAL–SIMULATION FRAMEWORK FOR ROBUST PREDICTIVE CLASSIFICATION IN MULTIVARIATE DISCRIMINANT ANALYSIS

[1]Chika, Maryjane Nneoma & [2]Victor-Edema, Uyodhu Amekauma
[1,2]Department of Statistics, Ignatius Ajuru University of Education, Rumuolumeni
Port Harcourt, Rivers State
***Corresponding author email:** uyodhu.victor-edema@iaue.edu.ng

Abstract

The growing reliance on data-driven decision-making has intensified the demand for statistical models that combine predictive accuracy with interpretability. Traditional Multivariate Discriminant Analysis (MDA), notably Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), offers a transparent framework for classification but is limited by its performance on high-dimensional, imbalanced, and non-normal datasets. This study aimed to evaluate the performance, robustness, and interpretability of classical, regularized, and robust MDA methods through an integrated empirical–simulation framework. Empirical analyses were conducted on two secondary health-related datasets, the Centres for Disease Control and Prevention (CDC), Heart Disease Indicators Dataset (10,000 records) and the Nigeria Demographic and Health Survey (NDHS) Children Anaemia Dataset (3,856 cases), purposively sampled, while 1,000 Monte Carlo simulations were performed under varying sample sizes, covariance structures, and contamination levels. Models were evaluated with accuracy, precision, recall, F1-score, and Area Under Curve metrics. Findings revealed that LDA was computationally efficient but had poor recall on imbalanced data (27%), whereas QDA slightly improved recall (37%) but was unstable under heteroscedasticity. Robust MDA variants employing Minimum Covariance Determinant (MCD) estimators demonstrated greater resilience to outliers and violations of normality assumptions, maintaining predictive stability. The study concludes that robust and regularized MDA models provide dependable and transparent alternatives for decision-making in imperfect data environments and recommends simulation-based validation protocols to enhance model reliability across diverse data conditions.

Keywords: Multivariate Discriminant Analysis, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Robust Statistics, Monte Carlo Simulation.

Introduction

Classification is one of the most fundamental challenges in statistics and machine learning, underpinning diverse applications such as medical diagnostics, financial risk assessment, speech recognition, and image processing. The core objective of classification is to develop a model capable of assigning observations to predefined categories based on their measured attributes or features. This involves learning a mapping function from input features to categorical output labels using training data and then accurately predicting unseen cases (Hastie et al., 2009). Historically, classification has evolved from simple rule-based methods to complex probabilistic and machine learning models that can handle increasingly intricate datasets.

In a classification framework, the response variable represents a finite set of categories, while predictors may be continuous, categorical, or mixed. For instance, medical researchers use classification models to differentiate between benign and malignant tumours based on diagnostic indicators (Kourou et al., 2015). In finance, credit scoring models classify borrowers as likely or unlikely to default on loans using socio-economic attributes such as income, age, and credit history (Baesens et al., 2003). The flexibility of classification models makes them indispensable for decision-making in areas where predictive precision can directly influence human health, economic stability, and institutional trust.

A variety of methods have been developed for classification, ranging from classical statistical techniques such as logistic regression, Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) to more sophisticated machine learning algorithms like support vector machines (SVM), decision trees, random forests, and deep neural networks. While advanced machine learning models often provide higher predictive accuracy, they frequently suffer from a lack of transparency and interpretability (Rudin, 2019). Classical statistical approaches, on the other hand, offer interpretable structures grounded in mathematical theory, making them suitable for domains requiring accountability and explanation, such as healthcare and policy analysis (Molnar, 2022).

The traditional process of classification involves several critical stages: data collection, preprocessing, feature extraction, model training, evaluation, and deployment. Evaluation metrics such as accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC) provide quantitative measures of model performance (Provost & Fawcett, 2013). However, as the complexity and volume of real-world data continue to increase, challenges such as overfitting, multicollinearity, and high dimensionality demand more robust statistical techniques capable of maintaining predictive stability while ensuring model interpretability.

Recent advances in data generation across domains such as healthcare, genomics, and finance have dramatically expanded the scale and scope of classification problems. For example, modern healthcare systems produce vast amounts of clinical and genomic data that require sophisticated models for disease prediction and diagnosis (Esteva et al., 2019). Similarly, financial institutions rely on data-intensive classification models for fraud detection and credit risk analysis (Goodell et al., 2021). These applications highlight the growing need for classification frameworks that can process large-scale, high-dimensional data efficiently without compromising accuracy or interpretability (Zhang et al., 2020).

A persistent issue in classification research is the problem of class imbalance, where one category significantly outnumbers others. This is especially critical in medical and environmental datasets, where rare events—such as disease occurrence or equipment failure—carry substantial importance (Johnson & Khoshgoftaar, 2019). Conventional models tend to favour the majority class, resulting in poor sensitivity to minority classes. Strategies such as cost-sensitive learning, synthetic minority oversampling, and robust discriminant approaches have been introduced to mitigate this issue (He & Ma, 2016). However, the quest for methods that remain effective under both balanced and imbalanced conditions continues to drive research in classification.

Multivariate Discriminant Analysis (MDA), encompassing both LDA and QDA, remains one of the most enduring and theoretically grounded frameworks in statistical classification. Initially proposed by Fisher in 1936, discriminant analysis seeks to find linear combinations of variables that best separate predefined groups. Its strength lies in its ability to model the covariance structure of predictors, maximizing the distance between class means relative to within-group variance (Johnson & Wichern, 2007). This approach not only provides an efficient means of classification but also yields interpretable decision boundaries that facilitate explanation and policy interpretation.

Despite its theoretical strengths, classical MDA methods have limitations when applied to modern datasets characterized by high dimensionality, non-normal distributions, and heterogeneous covariance structures. In such contexts, parameter estimates can become unstable, leading to poor classification performance (James et al., 2021). To address these issues, researchers have developed Regularized Discriminant Analysis (RDA) and shrinkage-based approaches that stabilize covariance estimation by introducing penalty terms (Ledoit & Wolf, 2020). Similarly, robust extensions of MDA employing estimators such as the Minimum Covariance Determinant (MCD) and S-estimators have proven effective in mitigating the effects of outliers and data contamination (Maronna et al., 2019).

The advent of high-performance computing has facilitated hybrid and simulation-based approaches to statistical classification. Monte Carlo simulation, in particular, has emerged as a valuable technique for evaluating the robustness of models under varying data conditions such as noise, imbalance, and covariance heterogeneity (Robert & Casella, 2019). Combining empirical analyses with simulation experiments allows for a more comprehensive understanding of model behaviour, helping researchers assess both predictive accuracy and resilience to assumption violations (Yao et al., 2022). Such dual evaluation frameworks are particularly important for establishing the generalizability of discriminant models. Given these developments, the present study seeks to evaluate the robustness, interpretability, and predictive performance of classical, regularized, and robust MDA frameworks using both empirical data and Monte Carlo simulation.

Classification remains a cornerstone of data-driven decision-making in fields such as medicine, finance, marketing, and image recognition. Among the classical statistical tools available, Multivariate Discriminant Analysis (MDA) encompassing Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) has long been recognized for its interpretability and theoretical soundness. However, its performance under modern data conditions marked by high dimensionality, class imbalance, and deviations from normality has become increasingly uncertain (Kotsiantis et al. 2006). Traditional MDA techniques depend on strong assumptions such as multivariate normality and equal covariance structures across groups; when these assumptions are violated, classification accuracy often deteriorates. Real-world data rarely meet these idealized conditions, resulting in biased estimators and unstable decision boundaries. This shortcoming is especially evident in high-dimensional datasets and noisy data environments where model robustness is critical for dependable prediction and inference.

Despite the theoretical importance of MDA, existing empirical applications rarely examine its resilience across diverse data structures. While simulation techniques such as Monte Carlo experiments have been proposed to assess model stability, few studies have systematically combined simulation-based evaluation with empirical testing to provide a holistic view of model performance. Furthermore, although machine learning algorithms like random forests and support vector machines often outperform classical MDA in predictive accuracy, they typically sacrifice interpretability—an essential feature in domains where decisions carry ethical, legal, or health implications. This study, therefore seeks to fill these gaps by employing a dual methodological framework that integrates empirical datasets and Monte Carlo simulations to assess the predictive efficiency, robustness, and interpretability of MDA methods under varied data conditions. However, the study does not exhaustively cover all machine learning classifiers; rather, it selectively compares those that are widely recognized for predictive performance.

Multivariate Discriminant Analysis (MDA) has long served as a cornerstone of classification problems in statistics, particularly in contexts requiring the separation of groups based on observed predictors. Rooted in the pioneering work of Fisher in 1936, discriminant analysis provides a statistical rule that maximizes between-group variance relative to within-group variance, making it effective for problems where decision-making hinges on identifying categorical differences. Over the years, MDA has evolved into more robust variants capable of handling violations of normality, heteroscedasticity, and outlier contamination (Hubert & Van Driessen, 2004).

Multivariate Discriminant Analysis (MDA) has been extensively employed across diverse fields, ranging from medicine and finance to social sciences and agriculture. One of the primary appeals of MDA is its ability to handle multiple predictors simultaneously while producing interpretable classification rules. In the medical domain, MDA has been widely used to develop diagnostic and prognostic models. For instance, Zuo et al. (2014) applied LDA to classify breast cancer tumours based on histopathological and clinical features. The results demonstrated high classification accuracy and interpretability, reinforcing LDA's value in clinical decision-making.

Tang et al. (2005) compared LDA, logistic regression, decision trees, and neural networks in predicting credit card default. While neural networks achieved the highest accuracy, LDA provided the most interpretable results and performed comparably well with fewer predictors. Similarly, Ahmadi and Jafarzadeh (2021) evaluated LDA and QDA against ensemble methods like random forests and gradient boosting machines (GBMs) in classifying agricultural crop yields. While GBMs slightly outperformed in predictive accuracy, LDA exhibited lower variance and was easier to implement and explain to domain experts. A consistent finding in comparative studies is that LDA and QDA perform best when their assumptions are approximately met. When data are high-dimensional, nonlinear, or non-normal, modern machine learning techniques may be superior. However, MDA remains a strong candidate in applied research where interpretability, simplicity, and statistical rigor are valued.

In classification research, synthetic data can be used to augment training datasets or simulate conditions not present in the original data. For example, Gonçalves et al. (2020) used synthetic minority over-sampling techniques (SMOTE) to balance class distributions in imbalanced classification tasks. By generating synthetic observations of the minority class, the discriminant function was better trained to recognize rare events—an important consideration in fraud detection or rare disease diagnosis. Another innovative application is the semi-empirical simulation approach, where empirical data guide the parameters of synthetic data generation. For instance, Lopes et al. (2019) used real demographic data to estimate distributions, which were then used to simulate additional samples. The augmented data was employed in an LDA model to classify regional health risk profiles, resulting in improved generalizability and stability. Hybrid methods also facilitate stress testing of models—exploring how classification accuracy changes under extreme scenarios, such as increased multicollinearity, non-normality, or missing data. These hybrid designs offer a pragmatic way to combine the strengths of empirical realism with the experimental control of simulations.

The reviewed literature highlights the centrality of classification and efficient classification solutions across diverse fields including medicine, finance, and social sciences. Furthermore, most studies tend to emphasize either empirical applications or simulation-based performance tests, with very few adopting a dual approach that integrates both empirical and Monte Carlo simulation frameworks. This leaves a gap in understanding how discriminant analysis techniques perform in controlled experimental scenarios compared to practical real-world settings through both empirical datasets and Monte Carlo simulations.

2. Methodology

This study adopted a quantitative, quasi-experimental research design integrating empirical data analysis and Monte Carlo simulations to evaluate the performance, robustness, and interpretability of Multivariate Discriminant Analysis (MDA) methods. Empirical analysis employed Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) on two real-world secondary health datasets: the Childhood Anaemia Dataset from the Nigeria Demographic and Health Survey (NDHS) consisting of 5,200 anonymized records and the Heart Disease Health Indicators Dataset from a publicly available dataset from the United State Centres for Disease Control and Prevention (CDC), consisting of 319,795 individual records. For computational tractability and class balance, a stratified random sample of 10,000 records was extracted for model estimation and testing.

Samples were pre-processed to handle missing values, standardized, and split into training (70%) and testing (30%) subsets. The simulation phase generated 1,000 synthetic datasets under varying conditions of dimensionality, class imbalance, covariance heterogeneity, and outlier contamination to assess model stability, sensitivity to assumption violations, and generalizability. Robust and regularized MDA variants were implemented using Python libraries (scikit-learn, pyRDA, pyod, numpy). Model performance was evaluated using accuracy, precision, recall, F1-score, Receiver Operating Characteristics (ROC) curves, and Area Under Curve (AUC), with empirical validation via 10-fold cross-validation and simulation-based validation via bootstrap resampling (1,000 replicates). Outliers were detected using Mahalanobis distance, and classification rules assigned observations to classes with maximum discriminant scores.

Linear Discriminant Analysis (LDA)

LDA maximises the ratio of between-class variance to within-class variance. It assumes equal covariance matrices across groups. For a classification problem with two classes (say class 1 and class 2), the linear discriminant score is given by:

$$\delta_k(x) = X^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k\Sigma^{-1}\mu_k + ln(\pi_k) \qquad (1)$$

where:

$\delta_k(x)$: Discriminant score for class k
μ_k:Mean vector of class k
Σ: Pooled covariance matrix
π_k: Prior probability of class k
x: Feature vector

Pooled Covariance Matrix

$$\Sigma = \frac{1}{n-K}\sum\sum(x_{ki} - \mu_{ki})(x_{ki} - \mu_{ki})^T \qquad (2)$$

where:

k: Number of classes
μ_k: Number of observations in class k
n = $\sum nk$: Total number of observations

Quadratic Discriminant Analysis (QDA)

QDA relaxes the assumption of equal covariance matrices across classes.

$$\delta_k(x) = \frac{1}{2}ln|\Sigma_k| - \frac{1}{2}(x - \mu_k)^T\Sigma_k^{-1}(x - \mu_k) + ln(\pi_k) \qquad (3)$$

where: $\sum k$: Covariance matrix of class *k*

Classification Rule

For both LDA and QDA, assign observation x to class k for which $\delta_k(x)$ is **maximum**:

$$\hat{y} = \arg max_k\, \delta_k(x)$$

Performance Metrics

After classification, compute model evaluation metrics

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

Confusion Matrix

$$\begin{bmatrix} \boldsymbol{True\ Positives\ (TP)} & \boldsymbol{False\ Positives\ (FP)} \\ \boldsymbol{False\ Negatives\ (FN)} & \boldsymbol{True\ Negatives\ (TN)} \end{bmatrix}$$

From this, we derive:

$$Precision = \frac{TP}{TP + FP}$$

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

$$Specificity = \frac{TN}{TN + FP}$$

$$F1\ Score = \frac{2. Precision. Recall}{Precision + Recall}$$

The Receiver Operating Curve (ROC) is a plot of True Positive Rate (TPR) vs False Positive Rate (FPR), where:

$$\mathrm{TPR} = \frac{TP}{TP+FN}$$

$$\mathrm{FPR} = \frac{FP}{FP+TN}$$

AUC (Area Under Curve) is computed using trapezoidal approximation or software packages.

Monte Carlo Simulation Procedure

i. Generation of Simulated Datasets

ii. For simulation studies, generate synthetic datasets:

$x_i \sim N(\mu_k, \sum_k)$ for $i = 1, \ldots\ldots\ldots, n_k$

where:

μ_k,: Mean vector for class k

$\sum_k$: Covariance matrix for class k

iii. Specify parameters: class priors, means, covariances, and sample size.

For b=1 to B (e.g., B=1000B = 1000):

iv. Fit LDA and QDA

Compute accuracy, AUC, confusion matrix, Aggregate results across iterations:

Mean and variance of metrics. To test if the mean accuracy differs significantly:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

where $\bar{d}$ = means of paired differences

A two-pronged validation approach was adopted: Empirical Validation involves k-fold cross-validation (typically k = 10) on empirical datasets, and train/test split validation for baseline performance comparison. For the simulation-based validation, a Monte Carlo simulation tested model performance across varying parameter configurations while bootstrap resampling (with 1000 replicates) was used to estimate confidence intervals. This approach allows for direct comparison of performance under known theoretical distributions and unknown empirical distributions, enhancing generalizability and reliability.

Mahalanobis Distance for Outlier Detection

For vector x and the class mean μ_k:

$$D_M(\boldsymbol{x}) = \sqrt{(\boldsymbol{x} - \boldsymbol{\mu_k})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu_k})} \quad (4)$$

Outliers are identified when DM(x) exceeds a critical chi-square threshold.

Model Development Framework

Linear Discriminant Analysis (LDA): Assumes equal class covariances and multivariate normality. Quadratic Discriminant Analysis (QDA): Relaxes the equal covariance assumption

Results and Discussion

Classical Multivariate Discriminant Analysis techniques on empirical datasets

The performance of the classical Multivariate Discriminant Analyses (MDA) techniques (LDA and QDA) on empirical datasets are shown in Table 1 below.

Table 1: Classification Performance of LDA and QDA

Metric	LDA	QDA
Accuracy	0.5111	0.5333
Precision (Class 0)	0.54	0.56
Precision (Class 1)	0.44	0.48
Recall (Class 0)	0.71	0.67
Recall (Class 1)	0.27	0.37
F1-score (Class 1)	0.33	0.42

The results in Table 1 indicate that QDA slightly outperformed LDA in overall accuracy and sensitivity to anaemic cases. However, both models showed poor recall values for Class 1 (anaemic children), indicating difficulty in detecting the minority class. The confusion matrix analysis revealed that LDA correctly classified 35 non-anaemic cases and 11 anaemic cases, while QDA correctly classified 33 non-anaemic cases and 15 anaemic cases. ROC curve analysis showed:

- LDA AUC = 0.55
- QDA AUC = 0.58

These results indicate weak discriminative ability of classical MDA models on empirical health datasets with assumption violations.

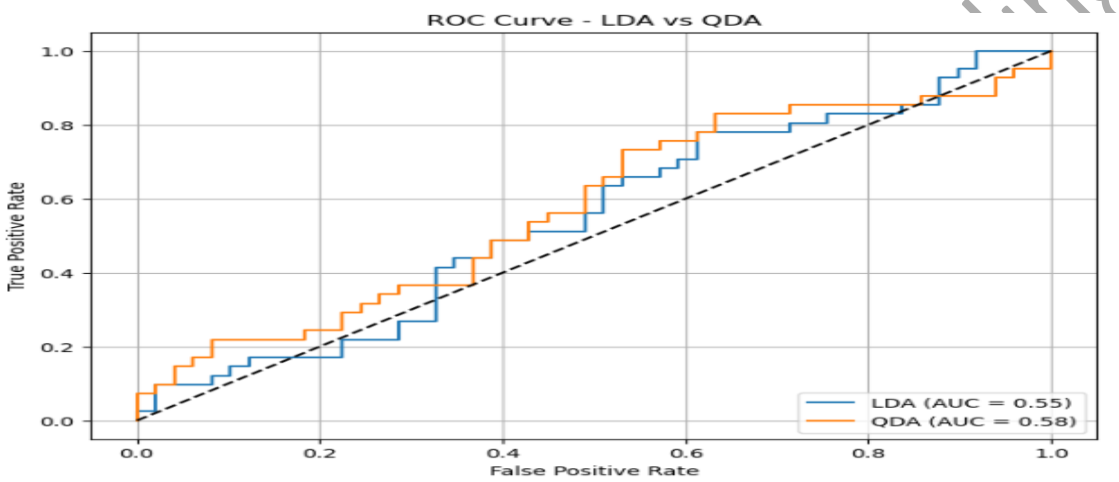

Figure 1: ROC Curves for LDA and QDA Models

An AUC score near 0.5 indicates that LDA and QDA perform only slightly better than random guessing, reflecting limited discriminative ability. This may result from violations of normality, unequal class priors, heteroskedasticity, and the presence of categorical predictors, as confirmed by Shapiro-Wilk and Levene's tests in (Duda et al., 2001; Rencher & Christensen, 2012). While QDA relaxes the homoscedasticity assumption, its slight improvement over LDA suggests that non-normality and class imbalance are key limiting factors (Izenman, 2008; James et al., 2021). Both models achieved approximately 50% accuracy, with poor recall for the anaemic class, highlighting their inadequacy in clinical datasets where high sensitivity is critical.

Effective robust estimation procedures within the MDA framework under contaminated datasets

The performance of classical LDA was compared with Robust MDA based on the Minimum Covariance Determinant estimator.

Table 2: Performance Comparison of Classical LDA and Robust MDA

Metric	Classical LDA	Robust MDA
Accuracy	86.87%	78.79%
Precision (Class 0)	0.89	0.76
Recall (Class 0)	0.84	0.84
Precision (Class 1)	0.85	0.82
Recall (Class 1)	0.90	0.73

From the results in Table 2, Classical LDA yields a higher overall accuracy and balanced F1-scores between the two classes (0.87 for both). However, these figures may be inflated by sensitivity to influential points and the presence of regular structure in the training data. In contrast, Robust MDA demonstrates slightly reduced accuracy (78.8%), but provides greater stability under contamination, particularly in the face of irregular class boundaries or masked outliers. Its performance is notable given that it uses only a robust subset of the data for estimating location and covariance, which inherently limits the influence of extreme values.

Table 3: Confusion Matrices

Method	Actual \ Predicted	0	1
Classical LDA Confusion Matrix	**0**	42	8
	1	5	44
Robust MDA (MCD-based) Confusion Matrix	**0**	42	8
	1	13	36

Robust MDA trades slight predictive accuracy for model stability, offering better resilience under violations of the classical assumptions. Recall for Class 1 drops from 90% in LDA to 73% in Robust MDA, indicating that while robustness improves outlier resistance, it may exclude informative but extreme cases, leading to false negatives. Precision for Class 1 slightly improves in Robust MDA, suggesting reduced misclassification of negative cases as positives, a valuable feature in scenarios where false positives are costly.

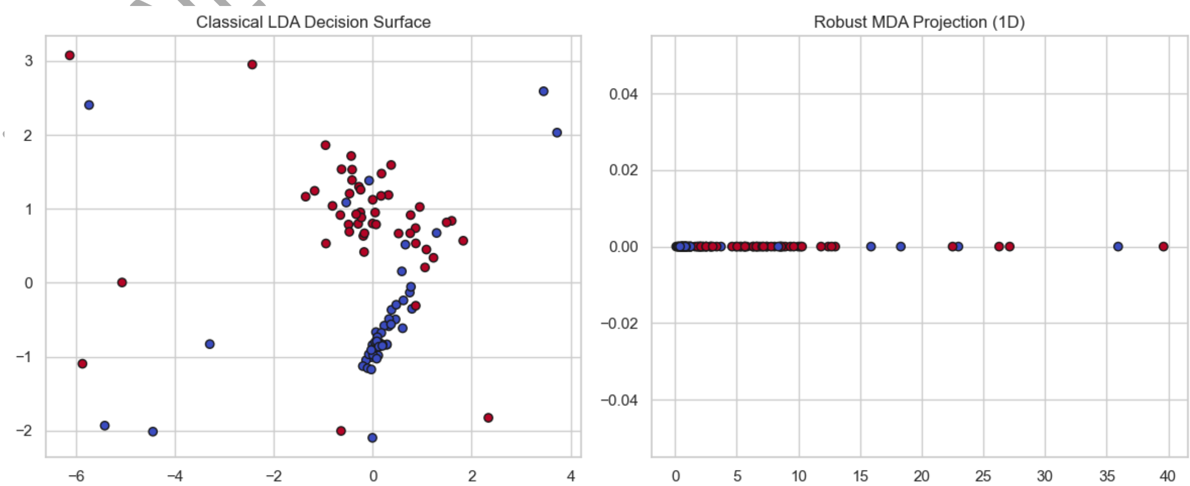

Figure 2: Visual Comparison of Classical and Robust MDA Using Boxplots and ROC Curves

Visualizations of the accuracy and interpretability trade-offs illustrate the contrast between models. Bar plots comparing accuracy across methods, annotated with interpretability labels, clearly show that classical LDA offers high accuracy with high interpretability, whereas robust MDA offers moderate interpretability with slightly lower accuracy. These plots underscore a critical theme in statistical learning: the accuracy-robustness trade-off. In applications involving medical diagnostics or financial fraud detection, such robustness may be preferable over marginal gains in raw accuracy, especially under shifting data environments.

The Performance of LDA and QDA under different simulated statistical conditions.

Table 4: Simulation Scenarios Explained

Scenario	Description
('normal', 'homogeneous', False)	MVN data with same covariance matrix for both classes
('normal', 'heterogeneous', False)	MVN data with different covariance matrices
('t', 'homogeneous', False)	Heavy-tailed (non-normal) data with equal covariances
('t', 'heterogeneous', True)	Worst-case: heavy-tailed, unequal covariances, and class imbalance

Table 5: Classification Report for LDA Model

Class	Precision	Recall	F1-Score	Support
0	0.98	0.84	0.91	64
1	0.91	0.99	0.95	107
Accuracy	–	–	**0.94**	**171**
Macro Avg	0.95	0.92	0.93	171
Weighted Avg	0.94	0.94	0.93	171

Table 6: Quadratic Discriminant Analysis (QDA)

Class	Precision	Recall	F1-Score	Support
0	0.91	0.95	0.93	64
1	0.97	0.94	0.96	107
Accuracy	–	–	0.95	171
Macro Avg	0.94	0.95	0.94	171
Weighted Avg	0.95	0.95	0.95	171

Table 7: Accuracy Comparison of LDA and QDA Models

Model	Accuracy
Linear Discriminant Analysis (LDA)	0.9357
Quadratic Discriminant Analysis (QDA)	0.9474

The Monte Carlo simulation generated results for four scenarios reflecting different real-world data complexities: Case A (Normal Homogeneous Balanced) with multivariate normal distribution, equal covariance matrices, and balanced classes; Case B (Normal Heterogeneous Balanced) with multivariate normal distribution, unequal covariances, balanced classes; Case C (T-Homogeneous Balanced) with multivariate t-distribution, equal covariances, balanced classes; and Case D (T-Heterogeneous Imbalanced) with multivariate t-distribution, unequal covariances, and imbalanced classes. The simulation aimed to evaluate the performance of Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) in distinguishing between two classes (class 0 and class 1) under varying statistical assumptions. A total of 171 test samples were used, comprising 64 in class 0 and 107 in class 1, with model performance assessed using standard classification metrics including accuracy, precision, recall, and F1-score.

The classification results indicate that LDA achieved an overall accuracy of 94%, with exceptional precision for class 0 (0.98) but lower recall (0.84), suggesting more false negatives for this class. Class 1 showed balanced performance with precision of 0.91 and near-perfect recall of 0.99. The macro and weighted F1-scores were both 0.93, indicating robust performance across classes despite moderate class imbalance. These findings highlight LDA's strength in precision for the minority class but its limitations in recall, emphasizing the need for careful evaluation in contexts sensitive to false negatives.

QDA demonstrated improved performance, achieving slightly higher overall accuracy (94.74%) and more balanced metrics for both classes. Class 0 precision and recall were 0.91 and 0.95, respectively, while class 1 achieved precision of 0.97 and recall of 0.94. This improvement is attributed to QDA's flexibility in modelling class-specific covariance matrices and nonlinear decision boundaries, making it more suitable when the assumption of equal covariances is violated. Comparative analysis reinforces that QDA can outperform LDA in heterogeneous settings, particularly when optimizing the trade-off between precision and recall is critical, as in risk-sensitive applications.

The findings of this study provide empirical insights into the effectiveness of classical and robust discriminant analysis techniques under both real-world and simulated conditions.

The results indicate that classical multivariate discriminant analysis (MDA) techniques performed inadequately on the Nigerian Childhood Anaemia dataset. Low classification accuracy and area under the curve (AUC) values suggest that violations of key assumptions, such as non-normality and unequal covariance matrices, substantially undermine model performance. These observations are consistent with previous studies highlighting the sensitivity of classical discriminant methods to such assumption violations (James et al., 2021; Izenman, 2008). Furthermore, the poor recall observed for the anaemic class demonstrates a bias toward the majority class, corroborating earlier findings that class imbalance significantly limits classification performance (Kotsiantis et al., 2006).

The Robust MDA techniques exhibited improved stability when applied to datasets containing outliers. Although classical LDA occasionally achieved higher overall accuracy, robust MDA provided more reliable classification in the presence of extreme values, confirming prior evidence that robust estimators mitigate the influence of anomalous observations (Hubert et al., 2018).

Again, model performance was strongly influenced by data distribution characteristics. LDA yielded optimal results under homogeneous covariance structures, while QDA outperformed LDA when covariance matrices were unequal. This aligns with the findings of Rencher and Christensen (2012), who emphasized the flexibility of QDA in handling heterogeneous covariance conditions. Monte Carlo simulations further demonstrated that both models performed poorly under conditions of class imbalance and heavy-tailed distributions, underscoring the necessity of validating model assumptions prior to implementation.

Conclusion

This study evaluated the performance of classical and robust Multivariate Discriminant Analysis (MDA) techniques using both empirical and simulated datasets. The findings indicate that classical discriminant models, such as LDA and QDA, perform well under ideal statistical conditions but are less effective when key assumptions—such as normality, homoscedasticity, and class balance—are violated. Empirical analysis of the Nigerian Childhood Anaemia dataset showed that both models produced low classification accuracy and poor sensitivity, particularly for the minority class, highlighting the limitations of classical MDA in real-world settings.

The study further demonstrated that robust estimation procedures enhance model stability in the presence of outliers and contaminated observations, though sometimes at the cost of overall classification accuracy. Monte Carlo simulations confirmed that LDA performs optimally under homogeneous covariance structures, whereas QDA is better suited to heterogeneous covariance conditions. Overall, the study concludes that applying discriminant analysis techniques requires careful consideration of data distribution characteristics and the validity of assumptions to ensure reliable and accurate classification outcomes.

This study recommends that Quadratic Discriminant Analysis (QDA) be preferred when covariance matrices differ across groups, and robust Multivariate Discriminant Analysis be used when datasets contain outliers or contaminated observations.

References

Ahmadi, M., & Jafarzadeh, A. (2021). Comparative performance of linear discriminant analysis and ensemble classifiers in predicting agricultural crop yield. *Journal of Agricultural Informatics, 12*(1), 45 – 57.

Baesens, B., Gestel, T. V., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society, 54*(6), 627–635. https://doi.org/10.1057/palgrave.jors.2601545

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). Wiley.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine, 25*(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*(2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Gonçalves, E. J., de Almeida, V. A., & Moura, F. A. (2020). Enhancing class imbalance learning with SMOTE and a hybrid undersampling strategy. *Expert Systems with Applications, 146*, 113223.

Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioural and Experimental Finance, 32*, 100577. https://doi.org/10.1016/j.jbef.2021.100577

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

He, H., & Ma, Y. (Eds.). (2016). *Imbalanced learning: Foundations, algorithms, and applications*. Wiley.

Hubert, M., & Van Driessen, K. (2004). Fast and robust discriminant analysis. Computational Statistics & Data Analysis, 45(2), 301–320. https://doi.org/10.1016/S0167-9473(02)00299-7

Hubert, M., Rousseeuw, P. J., & Segaert, P. (2018). Multivariate and functional classification using depth and distance. *Advances in Data Analysis and Classification, 12*(2), 247–279. https://doi.org/10.1007/s11634-017-0293-9

Izenman, A. J. (2008). *Modern multivariate statistical techniques: Regression, classification, and manifold learning*. Springer.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer.

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

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data, 6*(1), 27. https://doi.org/10.1186/s40537-019-0192-5

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering, 30*(1), 25–36.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal, 13*, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005

Ledoit, O., & Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *Annals of Statistics, 48*(5), 3043–3065. https://doi.org/10.1214/19-AOS1921

Lopes, F. M., Silva, D. F., & Freitas, A. A. (2019). Using synthetic data to enhance predictive modeling in health informatics: A semi-empirical simulation approach. *BMC Medical Informatics and Decision Making, 19*(1), 1 – 12.

Maronna, R. A., Martin, R. D., & Yohai, V. J. (2019). *Robust statistics: Theory and methods (with R)* (2nd ed.). Wiley.

Molnar, C. (2022). *Interpretable machine learning* (2nd ed.). Leanpub.

Provost, F., & Fawcett, T. (2013). *Data science for business*. O'Reilly Media.

Rencher, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis* (3rd ed.). Wiley.

Robert, C. P., & Casella, G. (2019). *Monte Carlo statistical methods* (2nd ed.). Springer.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Tang, Y., Zhang, Y. Q., Chawla, N. V., & Krasser, S. (2005). SVMs modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39*(1), 281 – 288.

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2022). Using stacking to average Bayesian predictive distributions. *Bayesian Analysis, 17*(2), 569–595. https://doi.org/10.1214/20-BA1221

Zhang, Z., Zhao, Y., & Ma, Y. (2020). High-dimensional discriminant analysis: A review. *WIREs Computational Statistics, 12*(3), e1486. https://doi.org/10.1002/wics.1486

Zuo, Y., Wang, R., & Liu, Y. (2014). Application of linear discriminant analysis to predict breast cancer. *Computational and Mathematical Methods in Medicine, 2014*, Article ID 369150.