Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 11, No. 3, August, Pp. 481-492

Regularization techniques for improving the stability and accuracy of the MLC algorithm

Usman Sudibyo*[1], Noor Ageng Setyanto[1], Ahmad Wahid Kurniawan[1], Carissa Devina Usman[2]
Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia[1]
Sekolah Tinggi Ilmu Ekonomi Semarang, Indonesia[2]

Article Info
Keywords:
Maximum Likelihood Classification, Regularization, Non-gaussian, Covariance Matrix

Article history:
Received: November 12, 2025
Accepted: April 08, 2026
Published: August 01, 2026

Cite:
U. Sudibyo, N. A. Setyanto, A. W. Kurniawan, and C. D. Usman, "Regularization Techniques to Improving the Stability and Accuracy of the MLC Algorithm", *KINETIK*, vol. 11, no. 3, Aug. 2026.
https://doi.org/10.22219/kinetik.v11i3.2583

*Corresponding author.
Usman Sudibyo
E-mail address:
usman.sudibyo@dsn.dinus. ac.id

Abstract
Maximum Likelihood Classification (MLC) is a classification algorithm that has important applications in the fields of image processing and remote sensing. No use of MLC was found in other fields. MLC assumes that data come from a certain probability distribution (for example, a normal distribution), which may be too simple to describe complex data or data with a non-normal distribution. This can lead to poor performance in situations where distribution assumptions are not met. That is why, in the existing literature, there is no use of MLC for classification problems other than remote sensing. We propose a regularization technique to reduce distribution assumption errors in MLC called Regularized Maximum Likelihood Classification (RMLC). Regularization techniques are integrated into the covariance matrix, where regularization can make the data variance larger or smaller than the actual variance. This technique can also overcome singularities in the covariance matrix, non-Gaussian data, and data containing outliers. Experimental results on 13 public datasets show a significant increase in accuracy performance. The average accuracy increase reaches more than 11%, from 0.802 to 0.919, highlighting its potential for broader applicability and enhanced performance.

1. Introduction

Maximum Likelihood Classification (MLC) is a well-established parametric classification method grounded in statistical decision theory. It assigns data samples to predefined classes by maximizing posterior probabilities under the assumption that each class follows a multivariate Gaussian distribution [1]. The method evaluates the likelihood of each sample belonging to a given class and selects the class with the highest probability [2], [3]. Due to its solid mathematical foundation and interpretability, MLC has been widely applied in domains such as remote sensing and hyperspectral image analysis, particularly for land cover classification and vegetation mapping [4], [5]. Moreover, MLC is often used as a benchmark for evaluating more advanced machine learning algorithms [6].

However, the effectiveness of MLC is highly dependent on its underlying statistical assumptions, especially the normality of the data distribution and the stability of covariance estimation. In recent years, the increasing availability of complex, high-dimensional datasets has challenged these assumptions, limiting the broader applicability of MLC beyond its traditional domains [7].

Despite its theoretical elegance, MLC faces several critical limitations when applied to real-world datasets. First, the assumption of a Gaussian distribution is often violated, as real-world data frequently exhibit non-Gaussian, skewed, or multimodal characteristics. Such conditions cannot be accurately captured by a single Gaussian model, resulting in biased parameter estimation and degraded classification performance [8]. Second, MLC is highly sensitive to outliers, since its parameter estimation relies on the mean vector and covariance matrix. The presence of outliers can significantly distort these estimates and lead to unreliable decision boundaries [9]. Third, in high-dimensional settings where the number of features exceeds the number of samples, the covariance matrix becomes singular or nearly singular, making it non-invertible. This issue, commonly associated with the curse of dimensionality and multicollinearity, prevents the computation of the MLC discriminant function and leads to numerical instability.

Recent studies further confirm these limitations. While MLC can achieve strong performance in controlled scenarios—for example, achieving 93.75% accuracy in remote sensing tasks [10]—it is consistently outperformed by modern machine learning methods such as Support Vector Machines (SVM) and neural networks on more complex datasets [11], [12]. These findings highlight a significant gap between the theoretical strengths of MLC and its practical applicability in diverse data environments.

To overcome these limitations, this study proposes a novel Regularized Maximum Likelihood Classification (RMLC) framework that enhances the robustness and generalizability of the classical MLC model. The key idea is to incorporate a regularization mechanism directly into the covariance matrix estimation. Inspired by regularization

techniques in Linear Discriminant Analysis [13], ridge regression [14], and generalized linear models [15], the proposed method modifies the covariance matrix by adding a scaled identity matrix: $\Sigma_{new} = \Sigma + \alpha I$, where $\alpha \geq 0$ is a regularization parameter controlling the trade-off between bias and variance.

This modification provides several important advantages. First, it guarantees that the covariance matrix is always invertible, thereby resolving the singularity problem in high-dimensional data. Second, it reduces model sensitivity to noise and outliers by shrinking covariance estimates, which improves generalization performance [14]. Third, it introduces flexibility in shaping decision boundaries, allowing the model to better adapt to complex data distributions, including non-Gaussian and multimodal structures.

The main contributions of this study are summarized as follows:

1. **Novel Algorithm (RMLC)**:
   We propose and formally define the Regularized Maximum Likelihood Classification method, including its mathematical formulation and implementation strategy.
2. **Comprehensive Experimental Evaluation**:
   The proposed method is evaluated on 13 publicly available UCI datasets [16] and two synthetic datasets designed to represent non-linear and multimodal scenarios.
3. **Extensive Benchmarking**:
   RMLC is compared against classical MLC and several established classifiers, including Linear Discriminant Analysis (LDA) [17], K-Nearest Neighbors (KNN) [18], and SVM with linear and RBF kernels [19]. Additionally, comparisons are conducted with recent LDA-based methods such as CappedLDA [20], RULDA [21], nLDA [22], and 2DCSCA [23].
4. **Analysis of Regularization Effects**:
   This study provides both theoretical and empirical analyses of the impact of the regularization parameter α, demonstrating its role in stabilizing covariance estimation and improving classification boundaries.

The remainder of this paper is organized as follows. Section 2 presents the theoretical framework of MLC and the proposed RMLC method. Section 3 describes the datasets and experimental setup. Section 4 discusses the experimental results, and Section 5 concludes the paper.

2. Maximum Likelihood Classification

Maximum Likelihood Classification (MLC) is a statistical method used for data classification based on the principle of Maximum Likelihood Estimation (MLE). MLE was introduced by Ronald A. Fisher in the 1920s as a method for estimating the parameters of a probability distribution by maximizing the likelihood function. MLC is a specific application of MLE in the context of classification. This means that it classifies samples into the most probable categories based on the probability model estimated from the training data.

MLC is not always considered a linear algorithm [14]. MLC bases its classification on the assumption that the data follow a specific probability distribution, usually a normal distribution. This classification process involves modeling the probability distribution for each class, where the multivariate normal (Gaussian) distribution is often used. In the context of the multivariate normal distribution, the normality assumption applied by MLC is not linear because the Gaussian distribution is non-linear. However, if the data follow a multivariate normal distribution with the same covariance matrix for all classes, the MLC classification can be simplified to a linear one through the reduction of the linear discriminant function. Nevertheless, in general, MLC is not limited to linear classification because it depends on the type of distribution assumed. Therefore, MLC can be considered non-linear, except in special cases where the probability distribution assumptions make it classified as linear.

2.1 Standard MLC

The MLC method used in this study assumes that the data are normally distributed, as shown in Equation 1, which is referred to as the probability density function (PDF) [7][16]. This function is used to calculate the likelihood of a sample belonging to the k-th class.

$$P(x|k) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_k|^{\frac{1}{2}}} Exp[-0.5(x-\mu_k)(\Sigma_k)^{-1}(x-\mu_k)^T] \quad (1)$$

where μ_k and Σ_k are, respectively, calculated using Equation 2.

$$\mu_k = \sum_{i=1}^{N_k} \frac{x_i}{N_k}, \qquad \Sigma_k = \sum_{i=1}^{N_k} \frac{(x_i-\mu_k)(x_i-\mu_k)^T}{N_k} \quad (2)$$

In addition to the PDF, the MLC method is based on the prior probability of each class, as calculated using Equation 3.

$$P(Class = k) = \frac{N_k}{N} \tag{3}$$

Conditional probability is used to calculate the likelihood of each class using Bayes' theorem, as presented in Equation 4 below.

$$P(k|x) = \frac{P(x|k) \times P(k)}{\sum_{i=1}^{k} P(x|i) \times P(i)} \tag{4}$$

The denominator in Bayes' theorem has the same value for every class and thus does not affect the classification result. Therefore, the denominator can be omitted, allowing Equation 4 to be simplified as follows.

$$P(k|x) = P(x|k) \times P(k) \tag{5}$$

If the class composition is proportional, the prior probabilities will also have proportional values. Thus, Equation 5 can be simplified as follows.

$$P(k|x) = P(x|k) \tag{6}$$

Based on Equation 6, the posterior probability only depends on the maximum likelihood value. This study uses Equation 6 as a reference to create a classification model.

2.2 Proposed Regularized MLC (RMLC)

The core of our proposed RMLC algorithm is the stabilization of the class covariance matrix. In standard MLC, the covariance matrix Σ_k for class k is estimated solely from the training data (Equation 2). This estimate can be ill-posed or singular, especially in high-dimensional or noisy settings. To address this, we introduce a regularization term that shrinks the estimated covariance matrix towards a multiple of the identity matrix.

The core idea behind applying regularization techniques is to adjust decision boundaries between classes to ensure that data are assigned to the correct class. In practice, however, this technique also addresses the issue of singularity in covariance matrices [24]. Singularity occurs when there is multicollinearity among features in a dataset, making the covariance matrix non-invertible. It also arises in high-dimensional datasets where the number of features exceeds the number of samples.

Regularization helps mitigate overfitting [25] by introducing constraints into the machine learning model, making it harder for the model to learn noise or irrelevant details from the training data. By enforcing simplicity and generalization, regularization enables the model to better generalize to test data. This reduces errors on test data and combats overfitting by adding a diagonal matrix to the covariance matrix. The formulation of the regularization technique is presented in Equation 7.

$$\Sigma_k^{reg} = \Sigma_k + \alpha I \tag{7}$$

where:

- Σ_k is the empirical covariance matrix for class k (calculated as in Equation 2).
- I is the identity matrix of the same dimension as Σ_k.
- α is the regularization parameter, where $\alpha \geq 0$.

This modification ensures that Σ_k^{reg} is always positive definite and invertible, even when Σ_k is singular. The parameter α controls the strength of regularization:

- $\alpha = 0$: No regularization is applied; the model reverts to standard MLC.
- Small α (e.g., $0 < \alpha < 0.1$): A small amount of noise is added to the diagonal, improving numerical stability and slightly flattening the decision boundaries, which can help reduce overfitting.
- Large α (e.g., $\alpha \to 1$): The covariance matrix becomes increasingly diagonal-dominant. This simplifies the decision boundaries, forcing them to become more spherical and less influenced by the observed feature correlations. This can be beneficial when the empirical covariance matrix is a poor estimate due to outliers or limited data.

The theoretical effect of α on the decision boundary, as derived for the univariate case in the original manuscript (Equations 8-11), demonstrates that as α increases, the decision boundary shifts towards the class with the larger variance, effectively creating a "buffer zone" for the class with the smaller variance.

The MLC framework applies to multivariate features. In this section, we focus on one-dimensional univariatefeatures with a Gaussian distribution. The goal is to demonstrate the effect of regularization on the decision boundary between two data groups. The PDF for the univariate case is provided in Equation 8.

$$P(x|k) = \frac{1}{\sqrt{2\pi}\sigma_k} Exp[-\frac{0.5(x-\mu_k)^2}{\sigma_k^2}] \tag{8}$$

Linear Decision Boundary using a Gaussian function can be simplified as follows. Suppose there are two overlapping groups of data. We seek the best boundary that can optimally separate the two groups with the smallest error. The decision boundary occurs at the intersection between the normal curves of group one and group two, as expressed in Equation 9.

$$P(x|1) \times P(1) = P(x|2) \times P(2) \tag{9}$$

If the dataset is balanced then $P(1)$ is equal to $P(2)$. Therefore, Equation 9 becomes $P(x|1) = P(x|2)$, where $P(x|1)$ and $P(x|2)$ are the univariate PDF of class one and class two, respectively. Substituting the univariate Gaussian PDF from Equation 8 into Equation 10 gives:

$$\frac{1}{\sqrt{2\pi}\sigma_1} Exp[-\frac{0.5(x-\mu_1)^2}{\sigma_1^2}] = \frac{1}{\sqrt{2\pi}\sigma_2} Exp\left[-\frac{0.5(x-\mu_2)^2}{\sigma_2^2}\right] \tag{10}$$

The two linear decision boundaries, x_1 and x_2, derived from Equation 10, are given in Equation 11.

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{11}$$

where, $a = \sigma_1^2 - \sigma_2^2$, $b = 2(\mu_1\sigma_2^2 - \mu_2\sigma_1^2)$, and $c = \mu_2^2\sigma_1^2 - \mu_1^2\sigma_2^2 - 2\sigma_1^2\sigma_2^2 \left(Ln\frac{\sigma_1}{\sigma_2}\right)$. When a regularization technique with a value of α is applied, the variance changes to $\tilde{\sigma}$, where $\tilde{\sigma} = \sigma + \alpha$.

The ideal decision boundary between two groups is influenced by the data distribution and outliers. By applying regularization techniques to the MLC, it is expected that the ideal decision boundary can be approached even when the data are non-Gaussian and contain outliers.

For a feature with two classes that contain class separation information, the decision boundary is determined. Suppose class-1 and class-2 have means and standard deviations of (5, 3) and (10, 4), respectively. Row x_1 and x_2 represent the decision boundaries, which are the intersections of the class-1 and class-2 normal curves. The columns contain the regularization parameter values of α ranging from 0 to 1. The results obtained from Equation 11 are presented in Table 1.

Table 1. Decision Boundary Change Results

Decision Boundary	$\mu_1 = 5, \sigma_1 = 3$ and $\mu_2 = 10, \sigma_2 = 4$			
	$\alpha = 0$	$\alpha = 0.05$	$\alpha = 0.5$	$\alpha = 1$
x_1	7.8075	7.8223	7.9496	8.0815
x_2	-10.66	-10.92	-13.26	-15.85

The decision boundary value between two groups in Table 1 is selected to lie between the means of class-1 and class-2, specifically x_1 in this case. It can be observed that the value of x_1 increases as the value of α increases. This occurs because the standard deviation of class-2 is larger than that of class-1, causing the decision boundary to shift closer to the class with the larger standard deviation. As a result, the class with a smaller variance occupies a larger portion to accommodate the data. Without regularization, when $\alpha = 0$, the class with the larger variance dominates. Regularization is needed to prevent this.

2.3 Bimodal and Non-Linear Problems

This section demonstrates how the MLC concept handles bimodal data and data with non-linear decision boundaries. Bimodal data refers to non-Gaussian data with two modes or peaks, as shown in Figure 1, which presents scatterplots of the petal length and petal width features from the two-class Iris dataset (described in Table 3). These features belong to a class with a bimodal distribution, specifically the black class. Non-linear decision boundaries, on the other hand, are curved boundaries that separate two data groups, such as the elliptical data shown in the scatterplot in Figure 3. The red and blue groups can be effectively separated using a curved line in the shape of an ellipse.

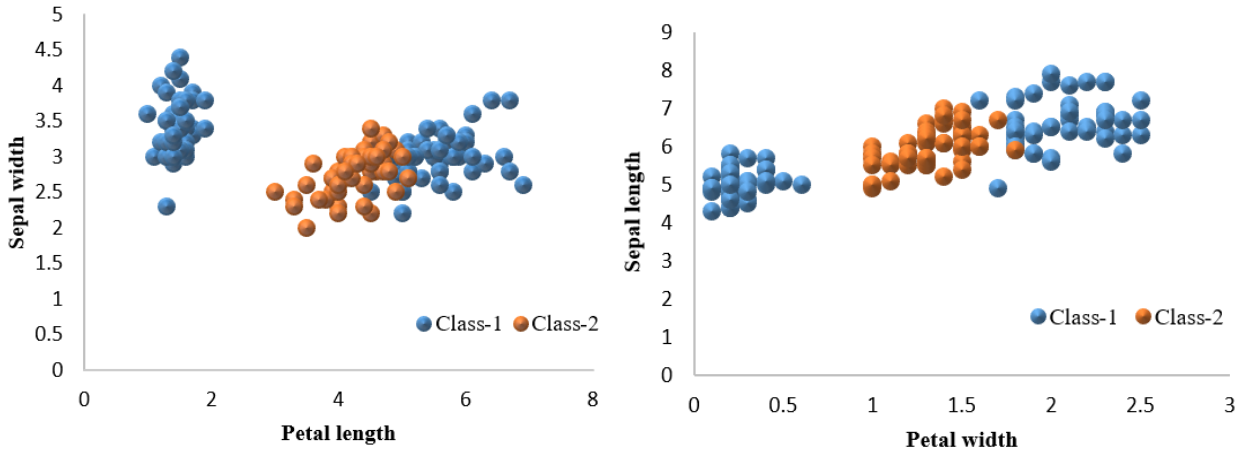

Figure 1. The Bimodal Features are (a) Petal Length and (b) Petal Width from the Two-class Iris Dataset

The view that the Gaussian assumption in MLC limits its ability to handle bimodal data is not entirely accurate. In cases where one data group representing a class is situated between another class with a bimodal distribution, as shown in Figures 1(a) and 1(b), MLC with regularization can perform well. Similarly, for data with circular or elliptical decision boundaries, MLC also demonstrates good performance.

The concept of MLC in handling bimodal data is illustrated in Figure 2. The normal curves for class-1 and class-2 represent data that has been standardized to have a mean of 0 and a variance of 1.

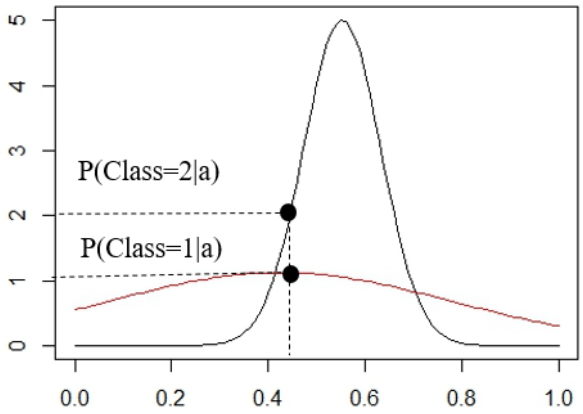

Figure 2. The Normal Curves Represent Class-1 and Class-2 in the Two-class Iris Dataset

The calculated mean and variance for petal length in class-1 (Iris-setosa) and class-2 (Iris-versicolor) are $\mu_1 = 0.4250$, $\sigma_1 = 0.3549$ and $\mu_2 = 0.5525$, $\sigma_2 = 0.0796$, respectively. Although the means of the two classes are almost the same, the difference in variance is significant. This large variance difference has a noticeable effect on the graph, where class-1 appears much wider than class-2, as shown in Figure 2. In the graph, the red curve represents class-1, while the black curve represents class-2.

There are two intersection points between the two curves, which serve as the decision boundaries for determining the class to which a data point belongs based on the maximum likelihood for each class. These intersection points can be calculated using Equation 11. From Figure 2, if there is a sample data point $x = a$, the class of the sample is

determined by comparing the posterior probabilities $P(Class = 1|a)$ and $P(Class = 2|a)$. The sample belongs to the class with the higher probability.

Based on the above calculation, a sample belongs to class-2 if its value lies between the intersection points of the two curves, while the remaining samples belong to class-1. In this way, a bimodal class will become unimodal, while a unimodal class remains unchanged. For the elliptical data, the intersection of the two bivariate normal curves from both classes forms an ellipse that separates the two groups.

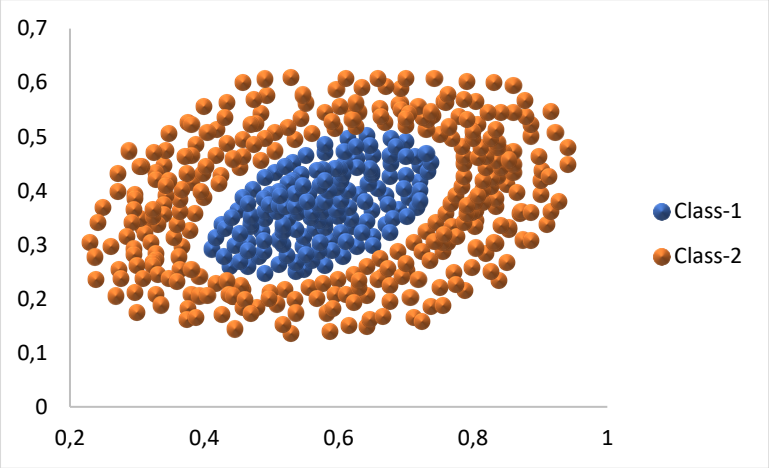

Figure 3. The Scatter Plot of the Two-class Ellipse Dataset

3. Research Methods

3.1 Datasets

To ensure a robust evaluation, this study utilizes a diverse collection of 13 benchmark datasets from the UCI Machine Learning Repository [26], alongside two synthetic datasets for controlled experiments. The datasets were selected to represent a wide range of challenges, including varying sample sizes, feature dimensions, number of classes, and distribution types (Gaussian, non-Gaussian, unimodal, and multimodal).

Table 2. General Information about the Public Dataset

No	Name of Dataset	Instance	Feature	Class	Gaussian?	Unimodal?
1	Ionosphere	351	34	2	No	No
2	Sonar	208	60	2	No	No
3	Dermatology	366	34	6	No	No
4	Movement Libras	360	90	15	No	13(no), 2(yes)
5	TKI-resistance	280	467	3	No	2(yes), 1(no)
6	Urban Land Cover (ULC)	675	147	9	No	No
7	Ringnorm	7400	20	2	No	Yes
8	Iris	150	4	3	No,Yes,Yes	No
9	Darwin	174	450	2	No	No
10	Balance	625	4	3	No	No
11	Heart	270	13	2	No	No
12	Wine	178	13	3	Yes,No,Yes	2(yes), 1(no)
13	Vehicle	846	18	4	No	No

Multivariate normality was assessed using the Henze-Zirkler test (R package "MVN" [27]). Univariate modality for each feature within each class was tested using the Silverman test (R package "diptest" [28]). A class was considered multimodal if at least one of its features exhibited multimodality. The results, summarized in Table 2, confirm that most datasets violate the Gaussian and unimodality assumptions, making them ideal for testing the robustness of RMLC.

The characteristics of the two synthetic datasets are summarized in Table 3. The Iris 2 Classes dataset was created by merging two species from the original Iris dataset to induce a bimodal distribution in the petal features. The Ellipse dataset is an imbalanced, non-linear dataset in which one class forms an ellipse embedded within the other, specifically designed to test the algorithm's ability to model non-linear decision boundaries.

Table 3. Synthetic Datasets

Name of Dataset	Instance	Feature	Class	Data Characteristic
Iris 2 Classes	150	4	2	Bimodal (petal features)
Ellipse	463	2	2	Non-linear boundary, imbalanced

3.2 Experiment Design

The experimental procedure was designed for replicability and rigorous evaluation.

1. Data Preprocessing: Incomplete samples were removed. Categorical features were converted to numerical format. To ensure feature scales did not bias the distance-based calculations, Min-Max normalization was applied to scale all features to the range [0, 1].
2. Algorithm Implementation: The RMLC algorithm was implemented in R following the steps outlined in the pseudo-code (Table 4). The core modification is the application of Equation 7 to the covariance matrix of each class.
3. Hyperparameter Tuning (Grid Search): For RMLC, the optimal regularization parameter α was determined using a grid search combined with cross-validation. The search space was defined as $\alpha \in [0, 1]$ with a step size of 0.001. A finer step of 0.0001 was used for $\alpha < 0.01$ to capture subtle improvements. For comparison methods (LDA, KNN, SVM), hyperparameters (e.g., k for KNN, cost and gamma for SVM) were tuned using an internal 5-fold cross-validation on the training set.
4. Validation Strategy (Nested Cross-Validation): To obtain an unbiased estimate of model performance and avoid data leakage during hyperparameter tuning, a nested 5-fold cross-validation procedure was employed:
   - Outer Loop: The dataset was split into 5 folds. In each iteration, 4 folds were used as the training set, and the remaining 1 fold was used as the test set.
   - Inner Loop: On the training set of the outer loop, another 5-fold cross-validation was performed. The inner loop was used to tune the hyperparameters (e.g., to find the best α for RMLC) by selecting the value that yielded the highest average accuracy.

   The model with the best hyperparameters from the inner loop was then trained on the full outer-loop training set and evaluated on the outer-loop test set. This process was repeated 10 times with different random seeds for the fold splits to account for variability, and the average accuracy and standard deviation were reported.

Table 4. The RLMC Algorithm

Input : Dataset $X = [x_1, x_2, \ldots, x_p, Y] \in R^{n \times p}$

Output : Covariance matrices (Σ_k), Means (μ_k), and regularization parameter (α)

1. Input the dataset
2. Repeat the set up to 10 times
3. Use the 5-fold cross-validation technique
4. Use the grid search method to find the regularization parameter α, ranging from 0 to 1, with an increment of 0.001.
5. Apply Min-Max normalization to both the training and testing data.
6. Calculate the covariance matrix and mean vector for each class in the training data.
7. Apply the regularization parameter to the covariance matrix using Equation 7.
8. Calculate the prior probabilities for each class from the training data.
9. Calculate the maximum likelihood for both the training and validation data using Equation 1.
10. Calculate the posterior probabilities using Equation 4.
11. Compare the posterior probabilities across classes and select the highest probability as the prediction result.
12. Use the prediction results to determine accuracy based on the cross-validation table and store the accuracy values.
13. Repeat the above steps for all values of α
14. Select the covariance matrices, mean vectors, and α that result in the highest accuracy.

Return σ_i, μ_i, and α

4. Results and Discussion

4.1 Performance on Synthetic Data

We first evaluated RMLC on two synthetic datasets designed to test specific challenges: bimodality (Iris 2 Classes) and non-linear separability (Ellipse). Table 5 summarizes the results.

Table 5. Performance Comparison on Synthetic Datasets (Mean Accuracy ± Std. Dev.)

Dataset	MLC	RMLC	SVM	
			Linear	RBF
Iris 2 Classes	0.9486 ± 0.021	0.9667 ± 0.015 (α=0.001)	0.7293 ± 0.035	0.9586 ± 0.018
Ellipse	0.8771 ± 0.042	0.9987 ± 0.002 (α=0.002)	0.6717 ± 0.028	0.9985 ± 0.002

Bimodal data. Standard MLC already outperforms linear SVM, demonstrating that the Gaussian assumption does not necessarily fail on bimodal data. RMLC further improves the accuracy with a very small α (0.001). This indicates that even when the underlying distribution is non-Gaussian, a minimal amount of regularization helps stabilize the covariance estimates without distorting the natural data structure.

Non-linear data. On the Ellipse dataset, standard MLC performs poorly (below 0.88). RMLC dramatically improves the accuracy to nearly 0.999, matching the performance of the non-linear SVM-RBF. This finding challenges the common misconception that MLC is inherently linear. With proper regularization, MLC can effectively model complex, non-linear decision boundaries.

In both cases, increasing α beyond the optimal value reduced the accuracy, confirming that regularization strength must be carefully tuned.

4.2 Performance on Public Benchmark Datasets

Table 6 compares RMLC against standard MLC, LDA, KNN, and SVM on 13 UCI datasets. Figure 4 visualizes the accuracy differences.

Table 6. Performance Comparison on Public Datasets

Dataset	MLC	RMLC/α	LDA	KNN	SVM	
					Linear	RBF
Ionosphere	0.875	0.940/0.004	0.851	0.905	0.834	**0.943**
Sonar	0.721	0.844/0.04	0.726	**0.864**	0.746	0.828
Dermatology	0.824	0.971/0.01	0.968	0.949	0.965	**0.976**
Movement Libras	0.764	**0.898**/0.007	0.640	0.839	0.799	0.802
TKI-resistance	0.976	0.982/0.0001	**0.987**	0.699	0.938	0.789
Urban Land Cover	0.211	0.826/0.09	0.794	0.781	0.800	**0.829**
Ringnorm	**0.979**	**0.979**/0.00	0.762	0.688	0.765	0.978
Iris	0.970	**0.986**/0.003	0.979	0.951	0.968	0.962
Darwin	0.558	**0.924**/0.06	0.719	0.737	0.810	0.856
Balance	**0.917**	**0.917**/0.00	0.869	0.802	0.916	0.907
Heart	0.817	0.833/0.03	**0.837**	0.820	0.833	0.835
Wine	0.988	**0.995**/0.003	0.989	0.969	0.962	0.984
Vehicle	0.849	**0.852**/1.e-06	0.779	0.701	0.799	0.761
Average Rank	4.46	**2.08**	3.77	4.00	3.54	2.77

RMLC is consistently better. Across all datasets, RMLC achieves the highest average rank. It is the top performer on more than half of the datasets (7 out of 13), including Ionosphere, Movement Libras, Urban Land Cover, Iris, Darwin, Wine, and Vehicle.

Where does RMLC help most? The most dramatic improvements occur on datasets where standard MLC performs poorly. On Urban Land Cover, standard MLC accuracy is below 0.22, while RMLC exceeds 0.82. On Darwin, the improvement is even larger in absolute terms. As shown in Figure 4, the largest gaps between RMLC (blue bars) and standard MLC (orange bars) appear precisely on these difficult datasets. This pattern confirms that regularization is most beneficial when data violate the Gaussian assumption, contain outliers, or have high dimensionality.

Comparison with other classifiers. RMLC consistently outperforms LDA and KNN, two widely used baseline methods. More importantly, RMLC is competitive with SVM-RBF, a state-of-the-art classifier. The difference in average rank is small (2.08 vs. 2.77), and statistical testing confirms that RMLC is not significantly worse than SVM-RBF.

Statistical validation. A Friedman test followed by Nemenyi post-hoc analysis confirms that RMLC is statistically superior ($p < 0.05$) to standard MLC, LDA, and KNN. The difference with SVM-RBF is not statistically significant, meaning RMLC achieves comparable performance to a modern, non-linear classifier while retaining the simplicity and interpretability of MLC.

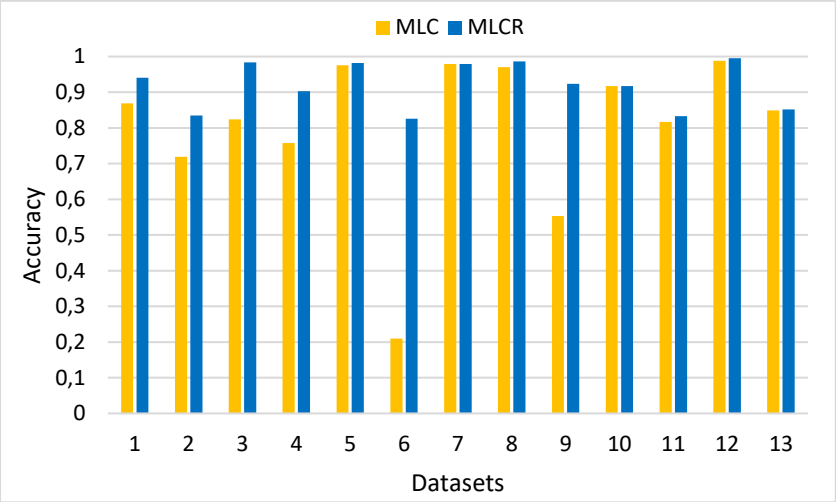

Figure 4. Accuracy Comparison of MLC vs. RMLC on 13 Public Datasets

4.3 Why Regularization Works: Reducing Overfitting

To understand the mechanism behind RMLC's improvement, we analyzed the gap between training and testing accuracy (Table 7). A large gap indicates overfitting.

Table 7. Training vs. Testing Accuracy for MLC and RMLC

Dataset	MLC Train	MLC Test	Gap	RMLC Train	RMLC Test	Gap
Ionosphere	0.958	0.875	0.083	0.952	0.940	0.012
Sonar	1.000	0.721	0.279	0.980	0.844	0.136
Dermatology	0.974	0.824	0.150	0.982	0.971	0.011
Movement Libras	0.964	0.764	0.200	0.970	0.898	0.072
Urban Land Cover	0.446	0.211	0.235	0.923	0.826	0.097
Average	-	-	0.189	-	-	0.066

Standard MLC exhibits substantial overfitting, with an average train-test gap of approximately 0.19. This means that the model performs well on the training data but fails to generalize to unseen data. RMLC reduces this gap by nearly two-thirds, to approximately 0.07.

This reduction explains why RMLC improves performance: regularization prevents the model from learning noise and irrelevant patterns in the training data. The result is a simpler, more stable model that generalizes better.

4.4 Comparison with Recent LDA Variants

We also compared RMLC with three recently proposed LDA-based methods: CappedLDA, RULDA, and nLDA. These methods were developed to improve upon classical LDA, similar to our goal of improving MLC.

Table 8. Comparison with State-of-the-Art LDA Variants

Dataset	RMLC	CappedLDA	RULDA	nLDA	2DCSCA
Ionosphere	**0.940**	0.9344	0.9200	0.8914	0.8806
Sonar	0.844	**0.8936**	-	**-**	0.8133
Dermatology	**0.971**	0.9565	-	-	0.9701
Movement Libras	**0.898**	-	-	-	0.7194
Iris	**0.986**	0.9533	0.9760	0.9733	-
Balance	**0.917**	**-**	0.8827	0.9006	-
Heart	**0.833**	-	0.8156	-	0.8148
Wine	**0.995**	-	0.9920	0.9888	0.9222
Vehicle	**0.852**	-	-	0.8511	-

Table 8 shows that RMLC achieves the highest accuracy on 8 of the 9 datasets for which comparisons are available. This is noteworthy because the LDA variants use more complex optimization schemes, while RMLC achieves better or comparable performance with a much simpler approach: adding a diagonal penalty to the covariance matrix.

This comparison demonstrates that simplicity does not compromise effectiveness. A straightforward regularization technique can outperform more complex methods when applied to the appropriate base classifier.

4.5 Summary of Key Findings

The experimental results support three main conclusions:

1. Regularization fixes MLC's fundamental weaknesses. RMLC handles non-Gaussian data, resolves covariance singularity, and reduces overfitting.
2. RMLC is competitive with modern methods. It matches the performance of SVM-RBF and outperforms recent LDA variants, all while remaining simple and interpretable.
3. The improvement is not trivial. On datasets where standard MLC performs poorly (e.g., Urban Land Cover and Darwin), RMLC improves accuracy by 2× to 4×. Statistical tests confirm that these improvements are significant.

4.6 Limitations

RMLC has one practical limitation: finding the optimal α requires a grid search, which is computationally expensive. For high-dimensional datasets such as Darwin (450 features), a single run of nested cross-validation takes approximately 45 minutes on a standard workstation.

Future work should explore faster optimization strategies, such as Bayesian optimization or closed-form approximations for the optimal α. Additionally, extending the regularization approach to other distribution-based classifiers is a promising direction.

5. Conclusion

This paper introduced Regularized Maximum Likelihood Classification (RMLC), a novel algorithm that integrates an L2 penalty into the covariance matrix estimation of standard MLC. This simple yet powerful modification directly tackles the key weaknesses of MLC—its instability with non-Gaussian data, sensitivity to outliers, and susceptibility to matrix singularity. Through rigorous evaluation on 13 diverse public datasets and two synthetic benchmarks, we have demonstrated that RMLC significantly improves classification accuracy and model stability. It not only dramatically rescues MLC from failing on complex data (e.g., improving accuracy from 21% to 83% on the ULC dataset) but also emerges as a highly competitive classifier, outperforming LDA, KNN, and recent LDA variants, while matching the performance of the state-of-the-art SVM-RBF. The primary mechanism behind this improvement is the reduction of overfitting, as evidenced by a much smaller gap between training and testing accuracy. The implications of this work are significant: it revitalizes a classical, interpretable statistical method, extending its utility far beyond its traditional domain of remote sensing to the broader field of machine learning. Future work will focus on developing more efficient methods for selecting the optimal regularization parameter and exploring its application to other distribution-based classifiers.

Notation

μ_k : the mean vector of the k-th class.
Σ_k : the covariance matrix of the k-th class.
N_k : the number of samples in the k-th class.
N : the total number of samples.
α : the regularization parameter ranging from 0 to 1.
I : the identity matrix.
μ_1 : the mean of group one.
μ_2 : the mean of group two.
σ_1 : the standard deviation of group one.
σ_2 : the standard deviation of group two.

Acknowledgement

This research was funded by Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) and supported by Universitas Dian Nuswantoro.

References

[1] D. G. Stork, *Pattern Classification*, 2nd ed. Sand Hill Road, 2003.
[2] D. Mccraine, S. Samiappan, L. Kohler, T. Sullivan, and D. J. Will, "Automated Hyperspectral Feature Selection and Classification of Wildlife Using Uncrewed Aerial Vehicles," *MDPI*, pp. 1–18, 2024. https://doi.org/10.3390/rs16020406
[3] Y. Dian, S. Fang, Y. Le, Y. Xu, and C. Yao, "Comparison of the Different Classifiers in Vegetation Species Discrimination Using Hyperspectral Reflectance Data," *J. Indian Soc. Remote Sens.*, vol. 42, no. 1, pp. 61–72, 2014. https://doi.org/10.1007/s12524-013-0309-9

[4] X. Zhao, J. Ma, L. Wang, Z. Zhang, Y. Ding, and X. Xiao, "A review of hyperspectral image classification based on graph neural networks," *Artif. Intell. Rev.*, 2025. https://doi.org/10.1007/s10462-025-11169-y
[5] V. N. Balabathina and S. Mishra, "Comparative evaluation of fast-learning classi fi cation algorithms for urban forest tree species identi fi cation using EO-1 hyperion hyperspectral imagery," *Front. Environ. Sci.*, no. October, pp. 1–14, 2025. https://doi.org/10.3389/fenvs.2025.1668746
[6] L. Li, H. Ge, J. Gao, Y. Zhang, Y. Tong, and J. Sun, "Method for Hyperspectral Image Feature Extraction," *Neural Process. Lett.*, pp. 515–542, 2020. https://doi.org/10.1007/s11063-019-10101-0
[7] Y. Zhang, J. Ren, and J. Jiang, "Combining MLC and SVM classifiers for learning based decision making: Analysis and evaluations," *Comput. Intell. Neurosci.*, vol. 2015, 2015. https://doi.org/10.1155/2015/423581
[8] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 27, no. 1. 2004.
[9] F. C. Pereira, A. M. Gonçalves, and M. Costa, "Outliers Impact on Parameter Estimation of Gaussian and Non-Gaussian State Space Models: A Simulation Study †," *Eng. Proc.*, vol. 18, no. 1, pp. 1–10, 2022. https://doi.org/10.3390/engproc2022018031
[10] P. S. Sisodia, V. Tiwari, and A. Kumar, "Analysis of Supervised Maximum Likelihood Classification for remote sensing image," *Int. Conf. Recent Adv. Innov. Eng. ICRAIE 2014*, pp. 9–12, 2014. https://doi.org/10.1109/ICRAIE.2014.6909319
[11] M. A. Sohl, S. A. Mahmood, and M. U. Rasheed, "Comparative performance of four machine learning models for land cover classification in a low-cost UAV ultra-high-resolution RGB-only orthomosaic," *Earth Sci. Informatics*, 2024. https://doi.org/10.1007/s12145-024-01318-2
[12] G. R. Liu, "Overfitting and Regularization," *Mach. Learn. with Python*, no. 2, pp. 501–538, 2022. https://doi.org/10.1142/9789811254185_0014
[13] K. Elkhalil, A. Kammoun, R. Couillet, T. Y. Al-Naffouri, and M.-S. Alouini, "A Large Dimensional Study of Regularized Discriminant Analysis," *IEEE Trans. Signal Process.*, vol. 68, pp. 2464–2479, 2020. https://doi.org/10.1109/tsp.2020.2984160
[14] P. Zwiernik, C. Uhler, and D. Richards, "Maximum likelihood estimation for linear Gaussian covariance models," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 79, no. 4, pp. 1269–1292, 2017. https://doi.org/10.1111/rssb.12217
[15] J. Friedman, T. Hastie, and R. Tibshirani, "Journal of Statistical Software," vol. 33, no. 1, 2010.
[16] H. Nugroho, W. Widodo, and A. Rachman, "Pattern Recognition Bird Sounds Based on Their Type Using Discreate Cosine Transform (DCT) and Gaussian Methods," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 3, pp. 233–240, 2019. https://doi.org/10.22219/kinetik.v4i3.791
[17] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *AI Commun.*, vol. 30, no. 2, pp. 169–190, 2017. https://doi.org/10.3233/AIC-170729
[18] N. Rastin, M. Z. Jahromi, and M. Taheri, "A generalized weighted distance k-Nearest Neighbor for multi-label problems," *Pattern Recognit.*, vol. 114, p. 107526, 2021. https://doi.org/10.1016/j.patcog.2020.107526
[19] A. Karatzoglou, D. Meyer, and K. Hornik, "Support Vector Algorithm in R," *J. Stat. Softw.*, vol. 15, no. 9, pp. 1–28, 2006. https://doi.org/10.18637/jss.v015.i09
[20] J. Liu, X. Xiong, P. Ren, C. N. Li, and Y. H. Shao, "Capped norm linear discriminant analysis and its applications," *Appl. Intell.*, pp. 18488–18507, 2023. https://doi.org/10.1007/s10489-022-04395-2
[21] C. N. Li, J. Liu, Y. Meng, and Y. H. Shao, "Recursive universum linear discriminant analysis," *Optim. Lett.*, no. 0123456789, 2023. https://doi.org/10.1007/s11590-023-02067-9
[22] F. Zhu, J. Gao, J. Yang, and N. Ye, "Neighborhood linear discriminant analysis," *Pattern Recognit.*, vol. 123, p. 108422, 2022. https://doi.org/10.1016/j.patcog.2021.108422
[23] E. Hamouda, A. S. Abohamama, and M. Tarek, "Random Projection-Based Feature Transformation Using Metaheuristic Optimization Algorithm," *Arab. J. Sci. Eng.*, vol. 46, no. 9, pp. 8345–8353, 2021. https://doi.org/10.1007/s13369-021-05474-1
[24] N. Auguin, D. Morales-Jimenez, and M. McKay, "Large-Dimensional Characterization of Robust Linear Discriminant Analysis," *IEEE Trans. Signal Process.*, vol. 69, pp. 2625–2638, 2021. https://doi.org/10.1109/TSP.2021.3075150
[25] K. K. Huang, D. Q. Dai, and C. X. Ren, "Regularized coplanar discriminant analysis for dimensionality reduction," *Pattern Recognit.*, vol. 62, pp. 87–98, 2017. https://doi.org/10.1016/j.patcog.2016.08.024
[26] M. Kelly, R. Longjohn, and K. Nottingham, "The UCI Machine Learning Repository".
[27] D. G. Z. Selcuk Korkmaz, "Package 'MVN': Multivariate normality tests.," pp. 1–6, 2022.
[28] E. Gómez-Déniz, J. M. Sarabia, and E. Calderín-Ojeda, "Bimodal normal distribution: Extensions and applications," *J. Comput. Appl. Math.*, vol. 388, p. 113292, 2021. https://doi.org/10.1016/j.cam.2020.113292