Statistical Analysis of Normalization Impact on GLCM-Based Coral Classification

Adi Pandu Rahmat Nababan[1], Stendy Budi Hartono Sakur[2*], Miske Silangen[3], Toto Haryanto[4], and Sony Hartono Wijaya[5]
[1−3]Information Systems Study Program, Department of Informatics Technology, Politeknik Negeri Nusa Utara
Sulawesi Utara, Indonesia 95812
[4−5]Computer Science Study Program, School of Data Science, Mathematics and Informatics, IPB University
Jawa Barat, Indonesia 16680
Email: [1]adiprnababan@gmail.com, [2]sakur.stendy@gmail.com, [3]miskesilangen10@gmail.com, [4]totoharyanto@apps.ipb.ac.id, [5]sony@apps.ipb.ac.id

***Abstract*—To the best of the authors' knowledge, this research provides part of the first statistically validated and comprehensive comparison of multiple normalization methods for GLCM-based coral image classification across different classical classifiers using ANOVA and Tukey's HSD tests. The process addresses the lack of evidence-based normalization analyses in previous studies and offers practical guidance for preprocessing design in coral reef monitoring systems. The research investigates the statistical impact of five normalization methods including MinMax Scaling, Z-Score Standardization, Robust Scaler, Power Transformer, and L2 Normalization on the performance of GLCM-based coral image classification using Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest (RF) classifiers. A dataset of 2,511 coral images, with healthy, unhealthy, and dead classes, is analyzed using 8 GLCM texture features. Five-fold cross-validation is applied, and classification performance is evaluated using accuracy, precision, recall, and F1-score metrics. One-way Analysis of Variance (ANOVA) and Tukey's Honestly Significant Difference (HSD) tests assess performance differences among normalization methods. Results show that normalization significantly impacts classification performance across all classifiers ($p < 0.05$). L2 Normalization produces the lowest accuracy and F1-score particularly for SVM ($\approx$ 0.397) due to the suppression of essential magnitude information in GLCM features. Z-Score Standardization, Robust Scaler, and Power Transformer achieve the highest and statistically comparable performance. Z-Score Standardization has the best results for SVM with an estimated accuracy of 0.913, and MinMax Scaling performs competitively for KNN. The results confirm normalization as a critical preprocessing step in texture-based coral image classification.**

***Index Terms*—Statistical Analysis, Data Normalization, Gray Level Co-occurrence Matrix (GLCM), Coral Image Classification**

I. Introduction

CORAL reef is part of the most essential components of marine ecosystems but their health is increasingly threatened by climate change and anthropogenic activities [1]. The trend shows the importance of accurate, consistent, and large-scale monitoring of coral conditions to support effective conservation strategies [2]. However, traditional monitoring methods depend on direct human diver surveys which are often inefficient in terms of time, cost, and repeatability [3]. The manual surveys are also inherently limited in spatial coverage, temporal frequency, and consistency, leading to unsuitability for long-term and large-scale coral reef monitoring programs. The condition has led to the introduction of technology-based monitoring systems particularly those utilizing digital image analysis and Machine Learning (ML) as promising alternatives in computational oceanography studies [4].

Coral image dataset adopted in this research is categorized into three health conditions including healthy, unhealthy, and dead. The dataset also represents secondary data from a previous coral image classification study [5]. Coral health identification primarily depends on texture-based feature extraction in the research because the variations in the condition are often manifested through spatial texture patterns on the surfaces [6]. Gray Level Co-occurrence Matrix (GLCM) is identified as highly effective for capturing the spatial characteristics compared to several other texture descriptors [5].

Received: Dec. 03, 2025; received in revised form: March 07, 2026; accepted: March 09, 2026; available online: Sep. 08, 2026.
*Corresponding Author

GLCM features used in this dataset are extracted using specific technical parameters defined in a previous study [5]. Coral images are subjected to several preprocessing steps in the original dataset preparation process which includes object cropping, background removal, image sharpening, and resizing before conversion into grayscale images for texture analysis. These preprocessing operations are intended to enhance texture visibility and standardize the input images before GLCM computation. GLCM matrices are calculated using three-pixel distances (d) of 1, 2, and 3 as well as four angular directions of 0°, 45°, 90°, and 135° to produce eight representative texture features including contrast, Angular Second Moment (ASM), energy, homogeneity, dissimilarity, correlation, entropy, and autocorrelation. These features have previously achieved a maximum classification accuracy of 91.85% [5]. However, the performance is highly dependent on preprocessing and feature extraction settings which are often underreported or selected heuristically in coral image classification studies.

The analysis from a broader perspective shows that previous studies on coral image classification have used a wide range of feature representations and learning paradigms including handcrafted texture, color, and shape descriptors combined with classical machine learning algorithms, as well as deep learning methods for end-to-end feature learning. Recent studies have reported the ability of Convolutional Neural Networks (CNNs), such as MobileNetV2 and other modern deep architectures, to achieve strong performance for coral reef and coral health classification under complex underwater conditions [6, 7]. Deep learning models often achieve high predictive accuracy but typically require large annotated datasets, substantial computational resources, and provide limited model interpretability. These constraints pose significant challenges for underwater monitoring systems deployed in resource-limited, real-time, or long-term operational environments. The trend leads to the wide adoption of GLCM-based texture features combined with classical classifiers due to its interpretability, lower computational cost, and robustness when training data are limited.

GLCM feature vectors have strong discriminative capability but often present substantial variability in scale and value range across feature dimensions. An example is the possibility of contrast features spanning orders of magnitude larger than correlation features as confirmed by the descriptive statistical analysis conducted in this research. The disparities have the capacity to introduce bias in classification models particularly in distance-based or similarity-based algorithms. This further leads to unstable predictions when applied in real coral monitoring systems under varying acquisition conditions or sensor configurations. Therefore, data normalization is considered important as a preprocessing step to standardize feature ranges and ensure balanced feature contributions during model training [8–10].

Previous studies on normalization methods in machine learning have reported that there is no single method considered universally optimal across all datasets and applications. Comparative analysis shows the possibility of MinMax Scaling performing effectively under specific conditions [8]. Z-Score Standardization is also effective when features follow approximately normal distributions [6], and Robust Scaler is advantageous in the presence of outliers [10]. Normalization is reported to have a significant impact on model reliability and generalizability in other application domains such as radiomics [11]. However, normalization strategies have not been systematically examined in the context of GLCM-based coral image classification. This situation leads to a lack of evidence-based guidance regarding the impact on model stability, robustness, and interpretability which is the reason for the current study.

The research has three main contributions. First, it is the provision of a comprehensive comparison of five widely used normalization methods for GLCM-based coral image classification. Second, it is to statistically validate the impact of normalization preferences using one-way Analysis of Variance (ANOVA) and Tukey's Honestly Significant Difference (HSD) tests across multiple performance metrics. Last, the analysis of model-specific sensitivity to normalization across Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest (RF) classifiers is conducted to support robust and evidence-based preprocessing design for coral image classification systems.

The results are expected both fundamental and practically relevant. The fundamental aspect focuses on advancing applied machine learning studies by quantitatively showing the critical role of preprocessing in texture-based classification models derived from coral imagery. Practically, the results will reduce the trial-and-error typically experienced by scholars and practitioners developing underwater monitoring systems. The identification of statistically supported normalization strategies allows the research to provide grounded guidance empirically for designing more stable, accurate, and reliable coral image classification models. They are essential for delivering trustworthy information to marine conservation stakeholders and policymakers. Therefore, the aim is to rigorously assess the statistical impact of five normalization methods on GLCM-based coral image classification perfor-

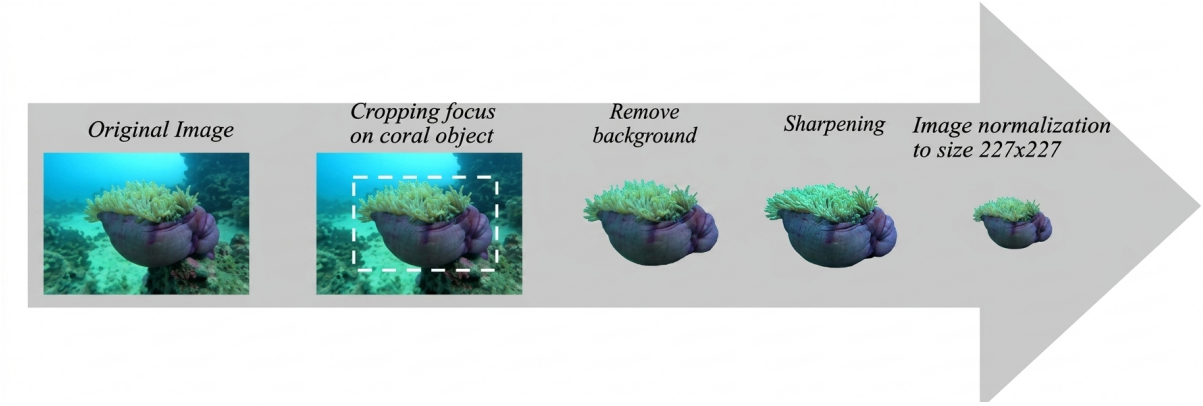

Fig. 1. Coral image preprocessing pipeline before Gray Level Co-occurrence Matrix (GLCM) feature extraction

mance. Advanced statistical analyses, such as one-way ANOVA and Tukey's HSD tests, are adopted to evaluate performance differences across the three classifiers including SVM, KNN, and RF. The analysis also determines the most suitable normalization strategy for enhancing classification reliability in coral reef monitoring applications.

II. Research Method

A. Data Acquisition and Exploration

The dataset used in the research consists of 2,511 coral image samples obtained from previous research [5]. The data are categorized into three classes (healthy, unhealthy, and dead) to represent the different conditions observed in underwater environments. The labeled samples provide the foundation for texture-based classification using statistical image features.

Several preprocessing steps are applied to the original dataset [5] to standardize coral images and enhance texture information before feature extraction. The pipeline includes coral object cropping to focus on the region of interest and background removal to eliminate irrelevant environmental regions. The others are image sharpening to enhance texture details and resizing to a fixed dimension of 227 $\times$ 227 pixels. The operations ensure that coral region becomes the dominant information source for subsequent texture analysis and improves the consistency of the extracted texture features. The overall preprocessing workflow used to prepare coral images before feature extraction is presented in Fig. 1.

The overall workflow used to generate GLCM features is summarized in Algorithm 1 to provide a clearer description of the preprocessing and feature extraction procedure. The algorithm outlines the sequential steps which start from loading coral images to performing preprocessing operations and subsequent conversion into grayscale representations. It also describes the computation of GLCM matrices using multiple pixel distances and orientations followed by the extraction of texture features used for classification.

Algorithm 1 Image preprocessing and Gray Level Co-occurrence Matrix (GLCM) feature extraction.

Input: Coral image dataset
Output: GLCM texture feature vectors

1. Load coral images from the dataset
2. Crop images to focus on coral object
3. Remove background regions
4. Apply image sharpening
5. Resize image to 227×227 pixels
6. Convert image to grayscale
7. Quantize grayscale levels
8. For each distance $d = \{1, 2, 3\}$
9. For each orientation $\theta = \{0°, 45°, 90°, 135°\}$
10. Compute GLCM matrix
11. Extract texture features
12. Store feature vectors
13. Use features as input for classification models

The preprocessing step is followed by the extraction of texture features using GLCM method. All coral images are converted into grayscale format before GLCM computation to simplify the representation of pixel intensity values used in texture analysis. The grayscale images are quantized to a fixed number of gray levels before constructing the co-occurrence matrices to standardize the intensity distribution used in the texture calculation.

GLCM matrices are computed using pixel distances of d = 1, 2, and 3, and 4 directional angles of 0°, 45°, 90°, and 135°. These parameter settings significantly affect the texture features produced by determining the spatial relationships between neighboring pixel intensities captured by the co-occurrence matrix. Moreover, the directional sensitivity is reduced by averaging the matrices obtained from the four orientations to produce

rotation-invariant texture representations.

Each averaged GLCM matrix produces eight standard texture features which included contrast, ASM, energy, homogeneity, dissimilarity, correlation, entropy, and autocorrelation. These statistical descriptors represent different characteristics of texture patterns such as intensity variation, uniformity, and spatial dependency. The extracted features are subsequently used as input variables for the classification models evaluated. The same feature extraction configuration is adopted to ensure consistency and comparability with the dataset used in the original study [5].

An in-depth exploratory data analysis is conducted before normalization to understand the characteristics of the feature [12]. Descriptive statistical analysis is subsequently conducted to identify extreme disparities in feature scales [13], which can lead to bias in distance-based models such as SVM and KNN. This process is followed by distribution visualization and skewness computation to confirm that the features have highly non-symmetric distributions [14]. The purpose is to justify the evaluation of both linear and non-linear normalization methods. Finally, a feature-level one-way ANOVA is conducted to assess whether statistically significant mean differences existed between coral classes. It is applied to validate the significance of mean differences across coral classes [15] to ensure that GLCM features used are discriminatively valid.

B. Data Preprocessing: Normalization Methods

The research comparatively evaluates five different normalization methods applied to GLCM feature matrix. MinMax Scaling in Eq. 1 often transforms numerical feature values into a fixed standard range [0, 1]. The transformation is considered important in preventing distance/gradient-based models from being biased toward large-magnitude features [10].

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{1}$$

where x is the original feature value while $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the feature. Z-Score Standardization in Eq. 2 transforms each feature to have an mean μ of zero and a standard deviation σ of one. The transformation is defined as follows:

$$x' = \frac{x - \mu}{\sigma}. \tag{2}$$

where x represents the original feature value, μ is the mean of the feature across all samples, σ is the standard deviation of the feature, and x' is the standardized value (Z-score). Z-score reflects the number of standard deviations of the original value x from the mean μ for the purpose of comparing the features with different scales on a common standardized scale. Moreover, the Robust Scaler in Eq. 3 is designed to address sensitivity to outliers in conventional scaling methods [10]:

$$x'_i = \frac{x_i - Q_2(x)}{Q_3(x) - Q_1(x)}. \tag{3}$$

Robust Scaler centers the data around the median $Q_2(x)$ and scales it using the Interquartile Range (IQR) which is defined as the difference between the third quartile $Q_3(x)$ and the first quartile $Q_1(x)$. The process ensures the scaled output becomes more robust and less impacted by extreme values.

L2 Normalization in Eq. 4 rescales each feature vector to ensure L2 norm becomes one. The process is also referred to as Euclidean normalization [16] and the transformation is defined as follows:

$$x' = \frac{x}{\sqrt{\sum_{i=1}^{n} x_i^2}}, \tag{4}$$

where $\mathbf{x}$ represents the original feature vector, x_i is the i-th component of the feature vector, n is the total number of features, and $\mathbf{x}'$ is the normalized feature vector. The denominator corresponds to the L_2 norm of the vector which is computed as the square root of the sum of the squared feature values. The aim is to scale each feature vector towards having a unit length with the aim of preserving the directional information while suppressing the absolute magnitude differences among the samples. The Yeo–Johnson Transformation in Eq. 5 is a flexible generalization that can be applied to data containing zero, positive, or negative values [17]:

$$y = \begin{cases} \dfrac{(x+1)^{\lambda} - 1}{\lambda}, & x \geq 0, \lambda \neq 0 \\ \dfrac{-(-x+1)^{2-\lambda} - 1}{2-\lambda}, & x < 0, \lambda \neq 0 \end{cases}, \tag{5}$$

The transformation function $f(x, \lambda)$ depends on the original data and a tuning parameter λ which are estimated from the dataset. This transformation benefits algorithms that assume approximate normality by improving data validity and predictive performance.

C. Classification Models

SVM is a supervised learning algorithm that determines an optimal hyperplane with a maximum margin to separate data into distinct classes [18]. The inclusion of distance calculations and feature space geometry in the optimization process leads to the high sensitivity of the performance to feature scaling. The attribute causes the status of SVM as a key model, and the objective

function is presented in the following Eq. 6:

$$\min_{w,b} \frac{1}{2}\|w\|^2. \tag{6}$$

Then, the linear optimal hyperplane is defined in Eq. 7 as follows:

$$w^T x + b = 0, \tag{7}$$

where w is the weight vector considered orthogonal or normal to the hyperplane, x is the input feature vector, and b is the bias term that determines the offset of the hyperplane from the origin. It describes the decision boundary that separates the data points into different classes in the feature space. The sign of the expression $w^T x + b$ shows the side of the hyperplane on which a given sample lies to form the basis for linear classification in the SVM models.

Linear Kernel in Eq. 8 is the simplest kernel function used for linearly separable data or high-dimensional feature spaces [19]. The kernel function is defined as follows:

$$K(x_i, x_j) = x_i^T x_j. \tag{8}$$

where x_i and x_j represent two feature vectors in the input space and the superscript T is the transpose. The kernel value corresponds to the inner (dot) product between the two vectors and quantifies the similarity in the original feature space. The direct determination of the similarity without mapping data into a higher-dimensional space allows the linear kernel to offer computational efficiency. It is also considered effectively suitable for problems where classes can be separated by a linear decision boundary.

Radial Basis Function (RBF) or Gaussian kernel is particularly effective for modelling nonlinear data patterns [20]. Its formulation is presented in the following Eq. 9:

$$\exp\left(-\frac{1}{2\sigma^2}\|x_i - x_j\|^2\right), \tag{9}$$

where $\|x_i - x_j\|$ is the Euclidean distance between these vectors, and σ is the kernel width parameter controlling the spread of Gaussian function. A small value of σ produces broader decision boundaries and risks underfitting. Meanwhile, a large value leads to tighter decision boundaries which can cause overfitting.

The polynomial kernel maps input data into higher-dimensional feature spaces using polynomial functions [21]. The formulation is presented in the following Eq. 10:

$$(\gamma x_i^T x_j + r)^p, \tag{10}$$

where $x_i^T x_j$ is the inner (dot) product, γ is a scaling parameter controlling the impact of the input features, r is the kernel coefficient (constant term) that balances the contribution of lower-order versus higher-order terms, and p is the degree of the polynomial. A higher value of p allows the model to capture more complex nonlinear relationships but increases the risk of overfitting.

KNN is an instance-based non-parametric algorithm considered highly dependent on Euclidean distance which leads to high sensitivity to feature scaling [22]. The Euclidean distance is defined through the following Eq. 11:

$$d(x, x_i) = \sqrt{\sum_{j=1}^{n}(x_j - x_{ij})^2}, \tag{11}$$

where x and x_i represent the feature vectors of the test and training samples respectively, x_j and x_{ij} are the values of the j-th feature, and n is the total number of features. Classification is further performed by majority vote between the k nearest neighbor $N_k(x)$ [23] as expressed in Eq. 12:

$$\hat{y} = \text{mode}\{y_i \mid x_i \in N_k(x)\}, \tag{12}$$

where $\hat{y}$ represents the predicted class label, y_i is the class label of the i-th nearest neighbor, $N_k(x)$ is the set of the k nearest neighbors of the sample x, and k is a user-defined parameter that impacts the classification performance.

RF is an ensemble algorithm composed of multiple decision trees [24]. Decision-tree models are generally insensitive to feature scaling but RF serves as a stable control model to assess the extent of improvement contributed by normalization relative to scale-sensitive classifiers in Eq. 13 [25]:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_T(x)\}. \tag{13}$$

The $h_t(x)$ is the prediction of the t-th tree, and T is the total number of trees.

Hyperparameters are optimized within the training data of each cross-validation fold for each classifier using a grid-search strategy. The selected hyperparameters correspond to the configuration that maximizes the mean F1-score on the training folds. These optimal parameters are subsequently fixed and applied to the corresponding validation fold to obtain unbiased performance estimates. The procedure ensures model selection, and performance evaluation are strictly separated to prevent information leakage from the validation data.

D. Evaluation and Statistical Analysis

Feature normalization is conducted independently in each fold within the five-fold cross-validation framework. Normalization parameters for every fold are

learned exclusively from the training subset and subsequently applied to the corresponding validation subset using the same parameters. This fold-wise normalization strategy is adopted to prevent data leakage and to ensure an unbiased and fair evaluation of classification performance.

The model performance was evaluated using four primary metrics derived from the confusion matrix under five-fold cross-validation [26] including accuracy, precision, recall, and F1-score. These metrics are selected to enable a comprehensive comparison of the classification performance across different normalization methods and classifiers. The use of cross-validation ensures that the reported results reflect the generalization ability of the models while the aggregation of metric values across folds provide a stable basis for subsequent statistical analysis. Accuracy measures the proportion of correctly classified samples among all the samples. Accuracy represents the proportion of correctly classified samples in Eq. 14 [27]:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}, \tag{14}$$

where TP (True Positive) represents the number of correctly classified positive samples, TN (True Negative) is the number of correctly classified negative samples, FP (False Positive) is the number of negative samples incorrectly classified as positive, and FN (False Negative) is the number of positive samples incorrectly classified as negative.

Meanwhile, precision focuses on the reliability of positive predictions by measuring the proportion of correctly predicted positive instances in the test set. This metric is particularly relevant in situations where FP errors can lead to misleading conclusions or increased operational costs. Therefore, precision provides complementary information to accuracy by emphasizing prediction quality rather than overall correctness. The precision presented in Eq. 15 evaluates how often positive predictions are correct [28] and is considered important when the cost of FP is high.

$$\text{Precision} = \frac{TP}{FP + FP}. \tag{15}$$

Recall presented in Eq. 16 measures the ability of the model to identify all positive instances [29]. The emphasis is on the capability of the model to capture all relevant positive instances in the dataset. The equation is as follows:

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{16}$$

A high recall reflects the ability of the model to effectively minimize missed positive samples. It is important in applications where overlooking positive cases is undesirable. Moreover, recall complements precision by showing the trade-off between missed detections and false alarms.

F1-score is the harmonic mean of precision and recall as presented in Eq. 17 [30]. Both metrics are integrated into a single measure with the aim of balancing the trade-offs. The specific usefulness of F1-score is when the class distribution is uneven or a single evaluation value is required for comparison across different models and experimental settings. The equation is as follows:

$$\text{F1-Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}. \tag{17}$$

The statistical assessment of performance differences in terms of accuracy and F1-score is significant across the feature normalization methods [31]. Therefore, one-way ANOVA is conducted independently for each classifier [32]. The analysis enables the comparison of the mean performance values obtained from different normalization methods and determines the possibility of the variations observed occurring by chance. ANOVA is applied to the averaged results from the five-fold cross-validation to ensure a stable estimation of the model performance. F-statistic used in this analysis is defined in Eq. 18.

$$F = \frac{\text{MeanSquareBtwGroups}(MS_B)}{\text{MeanSquareBtwGroups}(MS_W)}, \tag{18}$$

where MS_B represents the mean square variance between normalization method groups, and MS_W is the mean square variance within the groups. A larger F-value reflects greater variability between groups relative to the within-group trend.

A p-value less than the significance level α = 0.05 leads to the rejection of the null hypothesis (H_0) that there is no statistically significant difference in the model performance among normalization methods [33]. It shows that the feature normalization method selected has a statistically significant impact on the classification performance of the corresponding classifier.

The pairs of normalization methods with significant difference are further identified through the application of Tukey's HSD test [34]. HSD value is determined using Eq. 19:

$$\text{HSD} = Q_{\alpha,k,N-k} \cdot \sqrt{\frac{MS_W}{n}}, \tag{19}$$

where $Q_{\alpha,k,N-k}$ is the critical value from the studentized range distribution, MS_W is the within-group variance from ANOVA [35], and n is the number of samples per group. Any absolute mean difference between the two normalization methods that exceeds

Cite this article as: A. P. R. Nababan, S. B. H. Sakur, M. Silangen, T. Haryanto, and S. H. Wijaya, "Statistical analysis of normalization impact on GLCM-based coral classification", CommIT Journal 20(2), 379–399, 2026.

TABLE I
DESCRIPTIVE STATISTICS OF GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) TEXTURE FEATURES EXTRACTED FROM 2,511 CORAL IMAGES PRIOR TO NORMALIZATION.

Feature	Mean	Std.	Min,	25%	50%	75%	Max.
Contrast	937.049600	579.310800	41.312530	475.198000	875.644300	1278.779000	4166.432000
Dissimilarity	18.787280	7.073014	4.336873	13.361720	18.839410	23.682050	46.718980
Homogeneity	0.171254	0.140001	0.024541	0.068420	0.105433	0.249898	0.823834
Energy	0.101960	0.147221	0.006369	0.010265	0.015193	0.194289	0.815972
Correlation	0.793194	0.114372	0.294986	0.724641	0.811269	0.881768	0.986745
Entropy	12.407550	1.832318	3.096369	11.495660	12.948540	13.826870	14.801910
Angular Second Moment (ASM)	0.032061	0.069939	4.06E-05	0.000105	0.000231	0.037748	0.665810
Autocorrelation	0.793194	0.114372	0.294986	0.724641	0.811269	0.881768	0.986745

HSD value reflects a significant difference in performance at α = 0.05 [36].

Tukey's HSD analysis is specifically applied to F1-score metric in the research to provide a more balanced and informative evaluation of classifier performance than only accuracy. F1-score simultaneously incorporates precision and recall to enable the assessment to consider both FP and FN in a single unified measure. The property is particularly important for multiclass coral classification where class distributions possibly differ across healthy, unhealthy, and dead categories and misclassification costs are not inevitably uniform. Accuracy can be misleading in these scenarios due to the possibility of achieving high values by favoring majority classes while underperforming on minority or ecologically critical categories. However, F1-score more reliably reflects the ability of the classifier to discriminate among all coral conditions and also maintains a balanced predictive behavior.

The focus on the post hoc statistical analysis from F1-score allows the research to ensure that the detected performance differences among normalization methods are not motivated by only overall correctness but rather represent genuine improvements in class-level discrimination. This method ensures that Tukey's HSD test is able to identify normalization method pairs that significantly enhance the robustness and consistency of the classification performance across all the classes. Therefore, the statistical comparisons provide a more meaningful and operationally relevant evaluation to support reliable inferences regarding the effectiveness of the methods in practical coral reef monitoring. The experimental pipeline strictly follows a training-validation separation at every stage including normalization, hyperparameter tuning, and model evaluation to ensure that no information from the validation folds affects the training process.

III. RESULTS AND DISCUSSION

A. Initial Data Exploration

Initial exploratory data analysis is conducted to validate the characteristics of the dataset consisting of 2,511 coral images which are represented by 8 primary GLCM features and to justify the need for preprocessing. The descriptive statistics in Table I show extreme disparities in the feature scales and the presence of serious outliers. An example shows that the contrast feature reaches values in the thousands with the maximum $\approx$ 4166.43 while energy and ASM remain in very small ranges at a maximum $\approx$ 0.82 and $\approx$ 0.67, respectively. ASM and homogeneity also show signs of substantial outliers as observed in the large gaps between Q_{75} and Max. These scale disparities and potential extreme values can lead to bias in distance-based models such as SVM and KNN. Therefore, normalization of the features is important to balance contributions and ensure model stability.

The analysis of the discriminative power of GLCM features [37] leads to the generation of violin plots in Fig. 2 to visualize the distribution of each across the target classes including healthy, unhealthy, and dead. The violin plots combine boxplots and kernel density estimates to assess class separability. The visualization shows that the features with the highest discriminative ability are contrast, dissimilarity, correlation, autocorrelation, and entropy. The trend shows that the distributions for the dead class consistently shift away from the other two classes, contrast and dissimilarity, to higher values as a reflection of coarser textures. Meanwhile, correlation, autocorrelation, and entropy tend to lower values as a sign of weaker pixel dependency or more uniform texture. Homogeneity only presents moderate discriminative capability but energy and ASM show the weakest class distinction due to highly overlapping distributions concentrated near small values (< 0.2). The violin plots generally confirm distinct and consistent textural characteristics for the dead category compared to the other classes.

ANOVA is conducted to evaluate the statistical significance of the differences in the mean values of GLCM features across the three target classes. The results in Table II show that all GLCM features are statistically significant at P-value < 0.05. The results show that none can be ignored because each con-

Cite this article as: A. P. R. Nababan, S. B. H. Sakur, M. Silangen, T. Haryanto, and S. H. Wijaya, “Statistical analysis of normalization impact on GLCM-based coral classification”, CommIT Journal 20(2), 379–399, 2026.

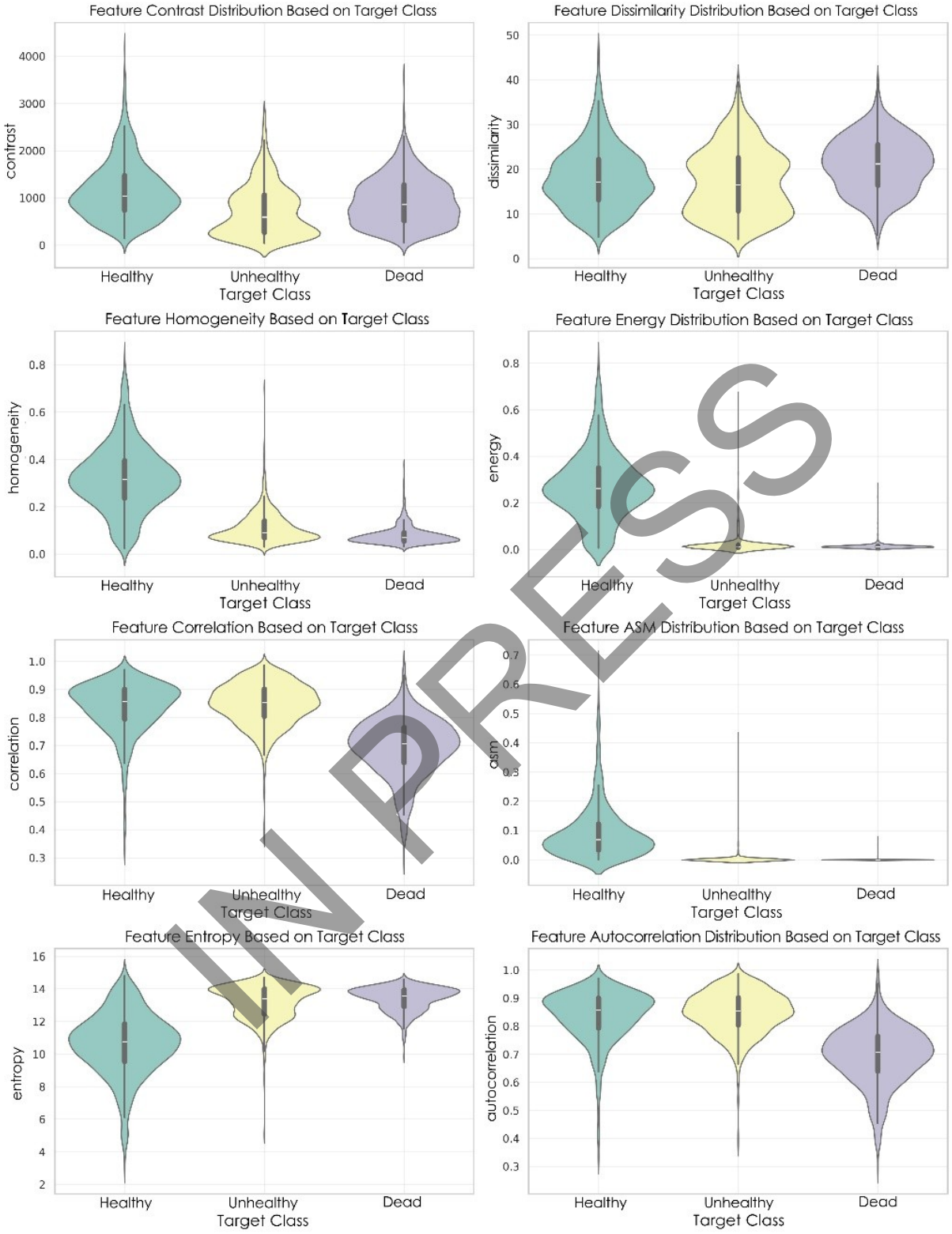

Fig. 2. Violin plots showing the distribution and discriminative power of GLCM texture features across three coral health classes including healthy, unhealthy, and dead for a dataset of 2,511 images.

TABLE II
ONE-WAY ANALYSIS OF VARIANCE (ANOVA) RESULTS FOR GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) FEATURES ACROSS HEALTH CLASSES TO VALIDATE FEATURE SIGNIFICANCE BEFORE NORMALIZATION.

No	Feature	F-Value	P-Value
3	Energy	2358.311426	0.000000e+00
2	Homogeneity	1750.149146	0.000000e+00
5	Entropy	1152.925560	0.000000e+00
6	Angular Second Moment (ASM)	787.289733	4.414233e-266
7	Autocorrelation	747.226825	2.751256e-255
4	Correlation	747.226824	2.751258e-255
0	Contrast	123.518041	6.872304e-52
1	Dissimilarity	74.233423	4.775378e-32

tributes relevant information for classification but the strength of the discriminative ability varies. Features, such as energy (F = 2358.31), homogeneity (F = 1750.15), and entropy (F = 1152.93), have the highest contributions followed by ASM, autocorrelation, and correlation. Meanwhile, contrast (F = 123.52) and dissimilarity (F = 74.23) have lower F-values as a sign of supporting rather than dominant role. The observation leads to the retention and subsequent normalization of all GLCM features to balance the scales and maximize classification performance.

B. Results of Data Normalization

The purpose of data normalization is to evaluate how different transformation methods impact the scale, distribution, and statistical stability of GLCM features and the subsequent effect of the changes on classifier performance. The substantial differences in the numerical range and variance among GLCM features show the need for normalization to prevent features with larger magnitudes from disproportionately impacting distance-based and kernel-based learning algorithms. Table III presents the descriptive statistics of normalized feature sets to reflect how each method reshapes the overall feature distribution. MinMax scaling compresses all feature values into the fixed range of [0, 1] to produce bounded and comparable scales, considered effectively suitable for distance-based classifiers. Z-Score Standardization also centers the data around zero with unit variance in line with its mathematical formulation and provision of standardized input distributions. The robust Scaler has a similar centering impact while reducing the effect of extreme values as reflected by the lower mean of the standard deviation.

L2 Normalization shows a significantly different distributional pattern characterized by an extremely narrow range and a very small standard deviation. This behavior signals normalization of the feature vectors primarily based on overall magnitude rather than individual feature distributions. The phenomenon leads to excessive compression of variability across features and a potential loss of discriminative texture information. Meanwhile, Power Transformer produces distributions, centered near zero with unit variance and expanded the interquartile range. It shows the effectiveness in reducing skewness and improving distribution symmetry. The statistics in Table III generally confirm that each normalization method induces distinct distributional characteristics to differentially impact the learning behavior and performance stability of the classification models in the subsequent analysis.

The boxplots in Fig. 3 show the impact of each normalization method on the contrast feature. MinMax Scaling produces a tight distribution between 0 and 1 to reflect strong range control. Meanwhile, Z-Score Standardization has a wider spread around zero with several high-end outliers which suggest the significant deviation of the contrast feature from the mean. Robust Scaler shows a similar pattern but with a narrower interquartile range as a sign of better concentration despite persistent outliers.

L2 Normalization records drastically different behavior by producing an extremely narrow distribution where approximately all contrast values cluster between 0.12 and 0.18. It is because the method scales the entire feature vector rather than the individual, leading to inapplicability for GLCM-based texture values. Meanwhile, Power Transformer designed to reduce skewness also produces a more symmetric distribution, centered around zero with fewer extreme values compared to Z-Score Standardization and Robust Scaler.

C. Results of Classification

GLCM feature dataset is normalized in the preprocessing stage using five methods which lead to five normalized datasets. This normalization step aims to examine how different feature scaling strategies impact classification performance. Each normalized dataset is subsequently classified using three algorithms including SVM, KNN, and RF.

The first evaluation phase focuses on independent subjection of each normalized dataset to hyperparameter optimization using GridSearchCV. This process aims to identify the best-performing configuration for each classifier-normalization combination based on internal validation splits. The optimal configurations obtained for SVM, KNN, and RF during this tuning stage are summarized in Tables IV–VI, respectively. In Table IV, parameter C acts as the regularization parameter for the SVM models, which critically controls the penalty hyperparameter to balance the trade-off between maximizing the decision boundary margin and

Cite this article as: A. P. R. Nababan, S. B. H. Sakur, M. Silangen, T. Haryanto, and S. H. Wijaya, "Statistical analysis of normalization impact on GLCM-based coral classification", CommIT Journal 20(2), 379–399, 2026.

TABLE III
DESCRIPTIVE STATISTICAL ANALYSIS OF GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) FEATURES ACROSS FIVE NORMALIZATION METHODS TO EVALUATE SCALE AND DISTRIBUTION STABILITY.

Normalization Method	Average Min	Average Max	Average Mean	Average Std.	Average 25%	Average 50%	Average 75%
MinMax Scaling	0.00000	1.00000	0.39295	0.15707	0.29221	0.37389	0.49253
Z-Score Standardization	-2.44274	4.09925	-3.68E-17	1.00019	-0.63450	-0.12526	0.60956
Robust Scaler	-1.71750	4.47708	0.16112	0.88468	-0.37363	0.00000	0.62636
L2 Normalization	0.12135	0.17728	0.13150	0.00592	0.12844	0.12969	0.13236
Power Transformer	-1.92144	2.37593	-2.12E-17	1.00019	-0.75770	-0.12128	0.89027

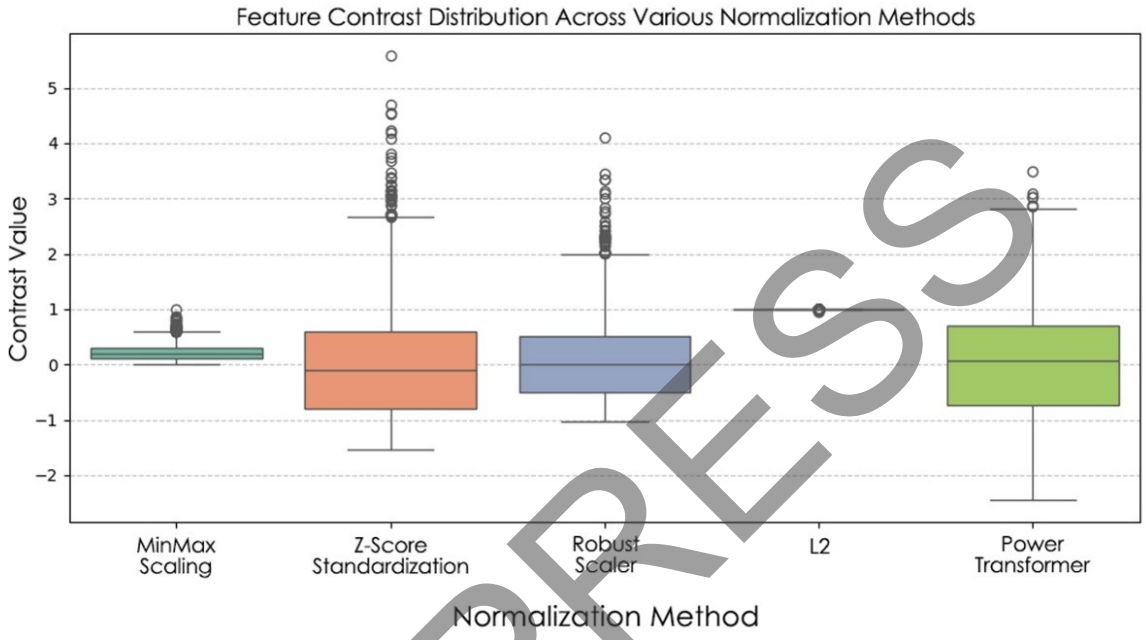

Fig. 3. Boxplots showing the distribution of the contrast feature across five normalization methods with an emphasis on the changes in scale, variance, and outlier presence (N = 2,511 images).

minimizing the training classification errors. A larger C value (e.g., C = 1000 for Linear kernels) enforces a stricter classification threshold, while a moderate value (e.g., C = 100 for RBF kernels) allows smoother decision surfaces to prevent overfitting. Meanwhile, in Table V, parameter k denotes the specific number of nearest neighbors cross-referenced by the KNN algorithm. This parameter determines the neighborhood size utilized to execute the majority voting mechanism for identifying the target coral health condition.

The results for SVM in Table IV shows that RBF kernel consistently achievs the best performance across most normalization methods. The highest accuracy is obtained using Z-Score Standardization with an accuracy of 0.9205 from parameters C = 100 and γ = 0.01. This result reflects strong compatibility between variance-based normalization and nonlinear decision boundaries. Polynomial kernels also show competitive performance particularly under MinMax Scaling and Power Transformer with both achieving an accuracy of 0.9185. Meanwhile, L2 Normalization has substantially lower results with a maximum accuracy of only 0.7316. It emphasizes the limited effectiveness for SVM-based GLCM feature classification.

KNN results in Table V show that the highest accuracy of 0.89341 is achieved under MinMax Scaling and Power Transformer using k = 9 and k = 7, respectively with Chebyshev distance metric. The trend suggests that distance-based classifiers benefit from bounded or well-distributed feature ranges. In comparison, Z-Score Standardization and Robust Scaler produce marginally lower but competitive accuracies of 0.89241 and 0.89192, respectively. It demonstrates that while variance-preserving techniques are highly stable, the strict bounding box of MinMax Scaling provides a slight advantage in neighborhood consistency for the KNN distance calculations.

The analysis of RF in Table VI shows that four

TABLE IV
OPTIMAL HYPERPARAMETER CONFIGURATIONS AND MAXIMUM ACCURACY FOR SUPPORT VECTOR MACHINE (SVM) MODEL ACROSS FIVE NORMALIZATION METHODS (N = 2,511 IMAGES).

Normalization Method	Kernel SVM	Accuracy	C	Parameter		
				Gamma	Degree	Coef0
MinMax Scaling	Linear	0.9066	1000	-	-	-
	Radial Basis Function (RBF)	0.9165	100	1	-	-
	Polynomial A	0.9185	10	1	3	-
	Polynomial B	0.9165	10	1	3	1
Z-Score Standardization	Linear	0.9085	1000	-	-	-
	Radial Basis Function (RBF)	0.9205	100	0.01	-	-
	Polynomial A	0.9125	10	1	3	-
	Polynomial B	0.9185	100	0.01	3	1
Robust Scaler	Linear	0.9085	1000	-	-	-
	Radial Basis Function (RBF)	0.9205	100	0.01	-	-
	Polynomial A	0.9026	100	1	2	-
	Polynomial B	0.9145	100	0.01	5	1
L2 Normalization	Linear	0.6859	1000	-	-	-
	Radial Basis Function (RBF)	0.6203	100	1	-	-
	Polynomial A	0.6481	100	1	5	-
	Polynomial B	0.7316	100	1	5	1
Power Transformer	Linear	0.8986	100	-	-	-
	Radial Basis Function (RBF)	0.9205	100	0.01	-	-
	Polynomial A	0.8986	10	1	2	-
	Polynomial B	0.9185	100	0.01	5	1

TABLE V
OPTIMAL HYPERPARAMETER CONFIGURATIONS AND MAXIMUM ACCURACY FOR K-NEAREST NEIGHBOR (KNN) MODEL ACROSS FIVE NORMALIZATION METHODS (N = 2,511 IMAGES).

Normalization Method	Accuracy	Parameter		
		k	Weight	Metric
MinMax Scaling	0.89341	9	Uniform	Chebyshev
Z-Score Standardization	0.89241	9	Uniform	Chebyshev
Robust Scaler	0.89192	11	Uniform	Chebyshev
L2 Normalization	0.79132	5	Uniform	Manhattan
Power Transformer	0.89341	7	Uniform	Chebyshev

TABLE VI
OPTIMAL HYPERPARAMETER CONFIGURATIONS AND MAXIMUM ACCURACY FOR RANDOM FOREST (RF) MODEL ACROSS FIVE NORMALIZATION METHODS (N = 2,511 IMAGES).

Normalization Method	Accuracy	Parameter				
		N Estimator	Max Depth	Max Features	Min Samples Leaf	Min Samples Split
MinMax Scaling	0.89541	100	None	$\log_2$	4	2
Z-Score Standardization	0.89541	100	None	$\log_2$	4	2
Robust Scaler	0.89541	100	None	$\log_2$	4	2
L2 Normalization	0.86453	300	10	$\log_2$	2	5
Power Transformer	0.89541	100	None	$\log_2$	4	2

normalization methods in the form of MinMax Scaling, Z-Score Standardization, Robust Scaler, and Power Transformer produce identical accuracies of 0.89541 with stable hyperparameter settings. They include 100 trees and $\log_2$ feature selection. However, L2 Normalization repeatedly has significantly lower performance with an accuracy of 0.86453 as a reflection of its limitations, extending to tree-based models.

Another important observation is that the optimal accuracies reported during the hyperparameter tuning stage in Tables IV–VI are not expected to exactly match the final cross-validation results in Table VII. The hyperparameter tuning focuses on identifying promising model configurations using internal training splits. Meanwhile, the final evaluation applies fixed optimal configurations to a full five-fold cross-validation with different data partitions which reflect the inevitability of minor numerical differences.

The best models identified after determining the optimal configurations are reapplied uniformly across

Cite this article as: A. P. R. Nababan, S. B. H. Sakur, M. Silangen, T. Haryanto, and S. H. Wijaya, "Statistical analysis of normalization impact on GLCM-based coral classification", CommIT Journal 20(2), 379–399, 2026.

TABLE VII
DETAILED PERFORMANCE EVALUATION OF ACCURACY AND F1-SCORE ACROSS FIVE-FOLD CROSS-VALIDATION FOR SUPPORT VECTOR MACHINE (SVM), K-NEAREST NEIGHBOR (KNN), AND RANDOM FOREST (RF) MODELS UNDER DIFFERENT NORMALIZATION METHODS.

Normalization Method	Fold	SVM		KNN		RF	
		Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
MinMax Scaling	1	0.8831	0.8840	0.0955	0.9058	0.9104	0.9109
MinMax Scaling	2	0.8980	0.8980	0.9080	0.9080	0.0930	0.0933
MinMax Scaling	3	0.8905	0.8909	0.8955	0.8957	0.8955	0.8958
MinMax Scaling	4	0.8678	0.8685	0.8678	0.8682	0.8778	0.8780
MinMax Scaling	5	0.8778	0.8786	0.8903	0.8907	0.8903	0.8903
Z-Score Standardization	1	0.9254	0.9257	0.9080	0.9082	0.9104	0.9109
Z-Score Standardization	2	0.9129	0.9130	0.9080	0.9081	0.0930	0.0933
Z-Score Standardization	3	0.9229	0.9231	0.8955	0.8955	0.8955	0.8958
Z-Score Standardization	4	0.9027	0.9030	0.8653	0.8657	0.8778	0.8780
Z-Score Standardization	5	0.9027	0.9029	0.8853	0.8856	0.8903	0.8903
Robust Scaler	1	0.9279	0.9282	0.9104	0.9107	0.9104	0.9109
Robust Scaler	2	0.9129	0.9129	0.8955	0.8957	0.0930	0.0933
Robust Scaler	3	0.9154	0.9156	0.8930	0.8930	0.8955	0.8958
Robust Scaler	4	0.8978	0.8980	0.8653	0.8658	0.8778	0.8780
Robust Scaler	5	0.8978	0.8980	0.8928	0.8931	0.8903	0.8903
L2 Normalization	1	0.3930	0.2866	0.7015	0.7000	0.8806	0.8809
L2 Normalization	2	0.3831	0.2673	0.7313	0.7317	0.8706	0.8708
L2 Normalization	3	0.4005	0.2960	0.7239	0.7239	0.8682	0.8683
L2 Normalization	4	0.4015	0.2984	0.6908	0.6923	0.8304	0.8305
L2 Normalization	5	0.4090	0.3105	0.7082	0.7090	0.8653	0.8657
Power Transformer	1	0.9303	0.9307	0.9005	0.9010	0.9104	0.9109
Power Transformer	2	0.9104	0.9105	0.9030	0.9031	0.0930	0.0933
Power Transformer	3	0.9030	0.9032	0.9055	0.9058	0.8955	0.8958
Power Transformer	4	0.8953	0.8954	0.8678	0.8682	0.8778	0.8780
Power Transformer	5	0.9002	0.9004	0.8853	0.8855	0.8903	0.8903

all five normalized datasets to ensure a fair and unbiased comparison. The performance is subsequently evaluated using five-fold cross-validation with the complete results reported in Table VII. A total of four normalization methods, including MinMax Scaling, Z-Score Standardization, Robust Scaler, and Power Transformer, generally have stable and consistent performance across all classifiers as reflected by small variations in accuracy and F1-score across folds.

The analysis of SVM shows that Z-Score Standardization and Robust Scaler achieve the strongest performance in most folds. Meanwhile, MinMax Scaling and Power Transformer remain competitive alternatives. KNN records the most stable behavior under MinMax Scaling with Z-Score Standardization, Robust Scaler, and Power Transformer producing comparable but slightly more variable results. RF shows the highest consistency overall with nearly identical accuracy and F1-score across the four normalization methods to confirming its robustness to feature scaling. Meanwhile, L2 Normalization consistently produces the weakest and least stable performance across all classifiers.

The cross-validation results in Table VII show three distinct performance regimes. First, Z-Score Standardization, Robust Scaler, and Power Transformer consistently deliver the highest and most stable F1-scores for SVM. These results reflect that variance-preserving normalization best maintains discriminative GLCM texture information. Second, KNN achieves its most stable performance under MinMax Scaling and suggests that bounded feature ranges improves neighborhood consistency in high-dimensional texture spaces. Third, RF records minimal sensitivity to normalization which further confirms the inherent robustness to feature scaling.

The analysis from a domain perspective shows the ability of GLCM features to encode contrast and spatial variability reflecting coral surface complexity and health conditions. Normalization methods that preserve relative variance across features provide more reliable classification outcomes. Meanwhile, magnitude-flattening methods such as L2 Normalization tends to suppress absolute contrast information, considered critical for effective coral texture discrimination.

D. Comparative and Statistical Performance Analysis

A comparative evaluation of the three classifiers shows the critical but different roles of normalization across algorithms. The magnitude of its impact is strongly correlated with the dependency of each model on the distance metrics and geometric relationships within the feature space. SVM has the strongest dependency on normalization selection compared to the other

TABLE VIII
FINAL CLASSIFICATION PERFORMANCE OF SUPPORT VECTOR MACHINE (SVM) ACROSS NORMALIZATION METHODS.

Normalization Method	Accuracy		Precision		Recall		F1-Score	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
L2 Normalization	0.3974	0.0098	0.5456	0.0173	0.3976	0.0102	0.2918	0.0161
MinMax Scaling	0.8835	0.0116	0.8876	0.0117	0.8835	0.0116	0.8840	0.0113
Power Transformer	0.9079	0.0137	0.9092	0.0140	0.9079	0.0137	0.9080	0.0138
Robust Scaler	0.9103	0.0128	0.9121	0.0132	0.9104	0.0128	0.9105	0.0128
Z-Score Standardization	**0.9133**	**0.0107**	**0.9146**	**0.0109**	**0.9133**	**0.0107**	**0.9135**	**0.0108**

Note: Results are reported as Mean ± SD. SD is Standard Deviation. Best values for each metric are presented in bold text.

TABLE IX
FINAL CLASSIFICATION PERFORMANCE OF K-NEAREST NEIGHBOR (KNN) ACROSS NORMALIZATION METHODS.

Normalization Method	Accuracy		Precision		Recall		F1-Score	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
L2 Normalization	0.7111	0.0165	0.7129	0.0153	0.7112	0.0165	0.7114	0.0164
MinMax Scaling	**0.8934**	**0.0160**	**0.8952**	**0.0154**	**0.8934**	**0.0160**	**0.8937**	**0.0159**
Power Transformer	0.8924	0.0158	0.8937	0.0157	0.8924	0.0158	0.8927	0.0158
Robust	0.8914	0.0163	0.8929	0.0154	0.8914	0.0163	0.8916	0.0162
Z-Score	0.8924	0.0179	0.8938	0.0172	0.8924	0.0178	0.8926	0.0178

Note: Results are reported as Mean ± SD. SD is Standard Deviation. Best values for each metric are presented in bold text.

classifiers with the performance differences exceeding 50% between the best and worst methods. It shows the vulnerability to feature magnitude distortion. SVM also has the highest sensitivity to normalization. Table VIII shows that L2 Normalization produces the poorest performance with an mean accuracy estimated at only 0.3974 ± 0.0098 and an exceptionally low F1-score of 0.2918 ± 0.0161. This drastic degradation signals that the classifier disrupts the magnitude relationships within texture-based GLCM features, considered important for SVM kernel mapping. L2 Normalization forces every feature vector to unit length which leads to the partial loss of fundamental information in the feature magnitude and subsequently causes an ineffective decision boundary.

The other normalization methods produce significant performance improvements. Z-Score Standardization is the best-performing method with an accuracy of 0.9133 ± 0.0107 and F1-score of 0.9135 ± 0.0108. The relatively small standard deviations, except for L2 Normalization, reflect high stability across folds and confirm the suitability of Z-Score Standardization for kernel-based algorithms. Robust Scaler and Power Transformer also produce strong and consistent results which further show the capability of these methods to stabilize variance or reduce skewness. The process can provide favorable optimization conditions for SVM.

The behavior is closely related to the physical texture characteristics of coral surfaces. Dead and unhealthy corals are typically associated with coarser textures and higher contrast values in GLCM features. Meanwhile, the healthy category has more homogeneous patterns. The suppression of feature magnitude through L2 Normalization also removes ecologically meaningful texture cues and directly impairs the ability of the model to discriminate coral health conditions.

KNN has a narrower performance gap across most normalization methods compared to SVM. It reflects that the preservation of relative distances is more critical than the enforcement of distributional transformations. The status of KNN as a pure distance-based classifier shows the natural dependence on the relative scale of the features. Therefore, normalization is an important determinant of the model performance. Table IX shows that L2 Normalization repeatedly produces the lowest performance with an accuracy of 0.7111 ± 0.0165 and F1-score of 0.7114 ± 0.0164 despite less severe performance degradation. This moderate degradation suggests that L2 Normalization disrupts magnitude information but the dependence of KNN on feature-wise distance calculations allows the retention of a partial discriminative structure.

The remaining normalization methods generates comparable results with accuracies ranging from 0.89 to 0.894. MinMax Scaling is observed to have the highest numerical accuracy of 0.8934 ± 0.0160 which reflects the ability of range-based normalization to effectively preserve relative distances. The property is considered favorable for the neighborhood computations of KNN. Power Transformer, Robust Scaler, and Z-Score Standardization follow closely with differences, perceived as significantly small to show practical performance gaps.

RF shows near-invariant performance across most

Cite this article as: A. P. R. Nababan, S. B. H. Sakur, M. Silangen, T. Haryanto, and S. H. Wijaya, "Statistical analysis of normalization impact on GLCM-based coral classification", CommIT Journal 20(2), 379–399, 2026.

TABLE X
FINAL CLASSIFICATION PERFORMANCE OF RANDOM FOREST (RF) ACROSS NORMALIZATION METHODS.

Normalization Method	Accuracy		Precision		Recall		F1-Score	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
L2 Normalization	0.8630	0.0191	0.8641	0.0190	0.8630	0.0191	0.8632	0.0192
MinMax Scaling	**0.8954**	**0.0124**	**0.8967**	**0.0121**	**0.8954**	**0.0124**	**0.8957**	**0.0126**
Power Transformer	**0.8954**	**0.0124**	**0.8967**	**0.0121**	**0.8954**	**0.0124**	**0.8957**	**0.0126**
Robust	**0.8954**	**0.0124**	**0.8967**	**0.0121**	**0.8954**	**0.0124**	**0.8957**	**0.0126**
Z-score	**0.8954**	**0.0124**	**0.8967**	**0.0121**	**0.8954**	**0.0124**	**0.8957**	**0.0126**

Note: Results are reported as Mean ± SD. SD is Standard Deviation. Best values for each metric are presented in bold text.

Fig. 4. Mean (± standard deviation) accuracy and F1-score across normalization methods for Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Random Forest (RF) classifiers.

normalization methods as a confirmation of its theoretical robustness to feature scaling due to the threshold-based decision mechanism possessed. The tree-based ensemble method inherently performs split decisions based on thresholding rather than distance metrics. Consequently, RF is expected to be more robust against unscaled or unevenly scaled data. The evaluation results in Table X confirm this theoretical expectation with MinMax Scaling, Power Transformer, Robust Scaler, and Z-Score Standardization producing identical accuracy values of 0.8954 ± 0.0124 and F1-scores of 0.8957 ± 0.0126. This high level of consistency emphasizes the scale invariance and insensitivity of RF to normalization method selected. Meanwhile, L2 Normalization induces an identifiable performance reduction in RF with accuracy of 0.8630 ± 0.0191. However, the decline is considerably smaller than the trend observed for SVM and KNN classifiers.

The results show that tree-based models are inherently more tolerant to feature-scaling variations because the decision-making process depends on threshold-based splits rather than distance or kernel computations. However, the reduced performance suggests that L2 Normalization consistently distorts the foundational feature magnitude distributions. It can impact the placement of optimal split thresholds during the tree construction. In Table X, the remaining normalization methods (MinMax Scaling, Z-Score Standardization, Robust Scaler, and Power Transformer) have identical accuracy and F1-score values. It confirms the relative scale invariance and stability of RF classifier when appropriate normalization strategies are applied.

Figure 4 provides a compact visual summary of the classification performance across normalization methods for SVM, KNN, and RF. They are reported as mean ± standard deviation over five-fold cross-validation. Panels (a) and (b) respectively present accuracy and F1-score to enable an efficient comparison of normalization effects across models. A consistent trend is observed in which L2 Normalization produces the lowest performance particularly for SVM with both accuracy and F1-score deteriorating sharply. It confirms that the enforcement of unit-length feature vectors suppresses important magnitude information in GLCM features and is detrimental to distance-based classifiers. Meanwhile, Z-Score Standardization, Robust Scaler, and Power Transformer achieve the highest

TABLE XI
ONE-WAY ANALYSIS OF VARIANCE (ANOVA) RESULTS ON THE IMPACT OF NORMALIZATION METHODS ON CLASSIFICATION PERFORMANCE, INCLUDING ETA-SQUARED (η^2) EFFECT SIZE ANALYSIS.

Classification	Evaluation	Source of Variation	sum_sq	df	F	P-Value (PR(>F))	η^2
SVM	Accuracy	Normalization	1.028304551	4	1841.196933	2.33×10^{-25}	0.9972917299
		Residual (error)	0.002792489	20	-	-	-
	F1-Score	Normalization	1.028304551	4	1841.196933	4.26×10^{-26}	0.9977148218
		Residual (error)	0.002792489	20	-	-	-
KNN	Accuracy	Normalization	0.131447523	4	120.4590811	1.07×10^{-13}	0.9601463684
		Residual (error)	0.005456107	20	-	-	-
	F1-Score	Normalization	0.131463491	4	121.6187876	9.78×10^{-14}	0.9605113894
		Residual (error)	0.005404736	20	-	-	-
RF	Accuracy	Normalization	0.004192378	4	5.319344187	0.004389228	0.5154730853
		Residual (error)	0.003940691	20	-	-	-
	F1-Score	Normalization	0.004203257	4	5.261054943	0.004624262	0.5127206678
		Residual (error)	0.003994691	20	-	-	-

and most stable performance for SVM with Z-Score Standardization marginally outperforming the others. MinMax Scaling performs competitively for KNN and reflects the suitability of bounded feature ranges for Euclidean-distance-based classification. RF reports minimal sensitivity to normalization as observed in the nearly identical mean values and overlapping error bars across all methods. This behavior is in line with the theoretical property of tree-based models which are mostly invariant to feature scaling. The visual trends in

Figure 4 strongly supports the statistical results obtained from one-way ANOVA and effect-size analysis. It shows that normalization has a very large practical impact on distance-based classifiers and is moderate on RF. The observations generally emphasize normalization as a critical design process in GLCM-based coral image classification due to the direct impact on both performance and robustness rather than being only a preprocessing formality.

E. Comparison of Analysis of Variance (ANOVA)

One-way ANOVA is conducted to quantitatively assess the impact of the five normalization methods on the classification performance of the three learning algorithms. It uses both accuracy and F1-score as dependent variables. The primary objective is to evaluate the null hypothesis (H_0) that the mean performance across all normalization methods is equal as a reflection of no statistically meaningful differences.

The results presented in Table XI show exceptionally strong evidence of performance differences across normalization methods for SVM. F-statistic for accuracy reaches 1841.19 with an extremely small p-value of 2.33×10^{-25} which is substantially lower than the conventional significance threshold ($\alpha = 0.05$). A similarly strong effect is also observed for F1-score with $p = 4.26 \times 10^{-26}$. The results decisively reject the null hypothesis and confirm that normalization has a profound and statistically significant impact on SVM performance. The magnitude of F-statistic reflects a very large effect size which shows the extreme sensitivity of the algorithm to variations in feature scaling.

A similar pattern is identified for KNN where ANOVA results with an accuracy of $p = 1.07 \times 10^{-13}$ and F1-score of $p = 9.78 \times 10^{-14}$, leading to the rejection of the null hypothesis. These values show that normalization significantly shapes the classification outcome of KNN in line with the dependence of the algorithm on distance-based operations. F-statistics of $\approx$ 120–121 also reflect substantial differences among normalization groups even though the magnitude is significantly smaller than the trend observed in SVM. This signals the relatively lower sensitivity of KNN to scaling distortions.

RF is theoretically robust to feature scaling but ANOVA results show statistically significant differences across normalization methods. F-statistic for accuracy is 5.32 ($p = 0.0044$) and F1-score is 5.26 ($p = 0.0046$). The p-values are lower than the $\alpha = 0.05$ threshold. It reflects the rejection of the null hypothesis. Then, the effect size is considerably smaller relative to SVM and KNN. It is in line with the tree-based decision mechanism of RF which does not depend on geometric distances but can be affected by transformations altering the relative distributions among features.

The results of one-way ANOVA in Table XI confirm that the selection of normalization method has a statistically significant impact on classification performance across all models ($p < 0.01$). The effect size analysis conducted using eta squared (η^2) also emphasizes the practical magnitude of the impact in addition to the statistical significance. A very large effect size η^2 of 0.997 is recorded for SVM which reflects the

TABLE XII
TUKEY'S HONESTLY SIGNIFICANT DIFFERENCE (HSD) POST HOC PAIRWISE COMPARISONS OF NORMALIZATION METHODS FOR SUPPORT VECTOR MACHINE (SVM) WITH F1-SCORE.

Group1	Group2	Meandiff	P-Adj	Lower	Upper	Reject
L2 Normalization	MinMax Scaling	0.5922	0.0000	0.5674	0.6171	TRUE
L2 Normalization	Power Transformer	0.6163	0.0000	0.5915	0.6411	TRUE
L2 Normalization	Robust Scaler	0.6188	0.0000	0.5940	0.6436	TRUE
L2 Normalization	Z-Score Standardization	0.6218	0.0000	0.5970	0.6466	TRUE
MinMax Scaling	Power Transformer	0.0240	0.0605	-0.0008	0.0489	FALSE
MinMax Scaling	Robust Scaler	0.0266	0.0323	0.0017	0.0514	TRUE
MinMax Scaling	Z-Score Standardization	0.0295	0.0149	0.0047	0.0544	TRUE
Power Transformer	Robust Scaler	0.0025	0.9980	-0.0223	0.0273	FALSE
Power Transformer	Z-Score Standardization	0.0055	0.9620	-0.0193	0.0303	FALSE
Robust Scaler	Z-Score Standardization	0.0030	0.9961	-0.0218	0.0278	FALSE

majority of the observed variance in F1-score within the experimental setting associated with normalization method selected. Meanwhile, RF has a comparatively lower but substantial effect size η^2 of 0.512. The trend empirically supports the theoretical understanding that distance- and margin-based classifiers such as SVM and KNN are highly sensitive to feature scale disparities in GLCM features. However, tree-based models in the form of RF are comparatively more robust to normalization impacts.

F. Tukey's Honestly Significant Difference (HSD) Post-Hoc Analysis

Tukey's HSD test is used for pairwise comparisons of the five normalization methods to identify the pairs with statistically significant differences in classification performance. This post hoc analysis is applied after one-way ANOVA to further examine the specific sources of variation when the null hypothesis of equal means is rejected. The procedure enables a controlled comparison between all possible pairs of normalization methods while maintaining the general family-wise error rates. The significance threshold is set at $\alpha = 0.05$ for the research, and the results are used to determine the methods with significant differences in terms of classification performance across the evaluated models.

The table headers for Tukey's HSD analysis are defined as follows. The 'group1' and 'group2' represent the pair of normalization methods being compared. The 'meandiff' denotes the difference between the mean F1-scores of the two groups. Then, 'p-adj' indicates the adjusted p-value for multiple comparisons. Meanwhile, 'lower' and 'upper' specify the bounds of the 95% confidence interval for the mean difference. Last, 'reject' states whether the null hypothesis of equal means is rejected (TRUE) or not (FALSE) at a significance level of $\alpha = 0.05$.

The pairwise comparisons for SVM classifier in Table XII shows a highly distinct separation between L2 Normalization and the other four methods. The mean difference ranges from 0.592 to 0.622 with all p-values < 0.001. These results confirm that L2 Normalization produces significantly lower F1-score values than MinMax Scaling, Power Transformer, Robust Scaler, and Z-Score Standardization. The observation is in line with the severe performance degradation of SVM previously reported under L2 Normalization.

A contrasting trend is observed in the comparison of the four modern normalization methods as reflected through more nuanced differences. MinMax Scaling shows statistically significant differences when compared to Robust Scaler and Z-Score Standardization ($p < 0.05$) with an estimated mean of 0.026–0.030. However, the difference between MinMax Scaling and Power Transformer is not statistically significant ($p = 0.0605$). The pairwise comparisons among Power Transformer, Robust Scaler, and Z-Score Standardization also produce non-significant results ($p > 0.05$). These show that the methods perform equivalently in terms of F1-score for SVM.

Tukey's HSD post-hoc results for KNN in Table XIII show a clear and consistent pattern where L2 Normalization performs significantly worse than all other methods. The pairwise comparisons between L2 Normalization and the remaining methods also reflect large positive mean differences ranging from 0.1803 to 0.1823 with adjusted p-values equal to 0.000 ($p < 0.001$), leading to the rejection of the null hypothesis in all cases. These results show that L2 Normalization substantially degrades F1-score performance compared to MinMax Scaling, Power Transformer, Robust Scaler, and Z-Score Standardization.

A contrasting observation is the absence of any statistically significant difference between MinMax Scaling, Power Transformer, Robust Scaler, and Z-Score standardization. The mean differences between the methods are extremely small ranging from -0.0021 to 0.0010 with all adjusted p-values greater than 0.99 and confidence intervals at zero. The trend leads to the consideration of the four methods as statistically

Cite this article as: A. P. R. Nababan, S. B. H. Sakur, M. Silangen, T. Haryanto, and S. H. Wijaya, "Statistical analysis of normalization impact on GLCM-based coral classification", CommIT Journal 20(2), 379–399, 2026.

TABLE XIII
TUKEY'S HONESTLY SIGNIFICANT DIFFERENCE (HSD) POST HOC PAIRWISE COMPARISONS OF NORMALIZATION METHODS FOR K-NEAREST NEIGHBOR (KNN) WITH F1-SCORE.

Group1	Group2	Meandiff	P-Adj	Lower	Upper	Reject
L2 Normalization	MinMax Scaling	0.1823	0.0000	0.1512	0.2134	TRUE
L2 Normalization	Power Transformer	0.1813	0.0000	0.1502	0.2124	TRUE
L2 Normalization	Robust Scaler	0.1803	0.0000	0.1491	0.2114	TRUE
L2 Normalization	Z-Score Standardization	0.1812	0.0000	0.1501	0.2123	TRUE
MinMax Scaling	Power Transformer	-0.0010	1.0000	-0.0321	0.0301	FALSE
MinMax Scaling	Robust Scaler	-0.0021	0.9996	-0.0332	0.0290	FALSE
MinMax Scaling	Z-Score Standardization	-0.0011	1.0000	-0.0322	0.0300	FALSE
Power Transformer	Robust Scaler	-0.0011	1.0000	-0.0322	0.0300	FALSE
Power Transformer	Z-Score Standardization	-0.0001	1.0000	-0.0312	0.0310	FALSE
Robust Scaler	Z-Score Standardization	0.0010	1.0000	-0.0301	0.0321	FALSE

TABLE XIV
TUKEY'S HONESTLY SIGNIFICANT DIFFERENCE (HSD) POST HOC PAIRWISE COMPARISONS OF NORMALIZATION METHODS FOR RANDOM FOREST (RF) WITH F1-SCORE.

Group1	Group2	Meandiff	P-Adj	Lower	Upper	Reject
L2 Normalization	MinMax Scaling	0.0324	0.0129	0.0057	0.0592	TRUE
L2 Normalization	Power Transformer	0.0324	0.0129	0.0057	0.0592	TRUE
L2 Normalization	Robust Scaler	0.0324	0.0129	0.0057	0.0592	TRUE
L2 Normalization	Z-Score Standardization	0.0324	0.0129	0.0057	0.0592	TRUE
MinMax Scaling	Power Transformer	0.0000	1.0000	-0.0267	0.0267	FALSE
MinMax Scaling	Robust Scaler	0.0000	1.0000	-0.0267	0.0267	FALSE
MinMax Scaling	Z-Score Standardization	0.0000	1.0000	-0.0267	0.0267	FALSE
Power Transformer	Robust Scaler	0.0000	1.0000	-0.0267	0.0267	FALSE
Power Transformer	Z-Score Standardization	0.0000	1.0000	-0.0267	0.0267	FALSE
Robust Scaler	Z-Score Standardization	0.0000	1.0000	-0.0267	0.0267	FALSE

equivalent in terms of F1-score performance for KNN classifier.

Tukey's HSD post-hoc analysis for RF in Table XIV provides a more moderate pattern than KNN and SVM. Statistically significant differences are observed only in comparisons including L2 Normalization. It is specifically observed from the fact that L2 Normalization performs significantly worse than MinMax Scaling, Power Transformer, Robust Scaler, and Z-Score Standardization with a consistent mean difference of 0.0324, an adjusted p-value of 0.0129 ($p < 0.05$), and rejection of the null hypothesis across all corresponding pairs. The results signal that RF is relatively robust to feature scaling but L2 Normalization has a negative impact on its F1-score performance.

The other four methods do not show any statistically significant differences. All pairwise comparisons produce mean differences equal to zero, adjusted p-values of 1.000, and confidence intervals with zero. It shows that the methods have statistically indistinguishable impacts on F1-score of RF.

G. Key Findings and Scientific Interpretation

The results show a clear and statistically validate impact of normalization methods on the classification performance of GLCM-based coral image features. An example of the most striking and scientifically significant results, supported by Tukey's HSD test ($p \ll 0.05$) is the complete failure of L2 Normalization particularly in SVM model where accuracy reduces drastically to approximately 0.397. This outcome is theoretically rooted in the principle of L2 Normalization which constrains every feature vector to have a unit norm. The transformation removes the magnitude information contained in GLCM features (contrast, energy, homogeneity, and correlation), which are attributes required for the absolute numerical ranges to carry important textural information. The elimination of the relative magnitude variations across these features allows L2 Normalization to fundamentally disrupt its discriminative power. The phenomenon causes difficulty for distance-based and kernel-dependent algorithms to effectively separate complex coral textures.

The condition emphasizes the increased sensitivity of distance-based models such as KNN and kernel-based algorithms in the form of SVM to transformations that distort feature magnitude relationships. The Euclidean distance computations in KNN significantly depend on the absolute scale of features, and distortion of magnitude weakens the ability of the model to identify meaningful neighbors. SVM similarly depends on the feature distribution and variance to construct optimal hyperplanes through its kernel functions. The suppression of the magnitude information in L2 Normalization allows the feature space produced to be-

come less conducive to accurate boundary formation. Meanwhile, RF shows significantly greater robustness across normalization methods. The status as a tree-based model reflects non-dependence on distance computations, allowing maintenance of relatively stable performance despite variations in feature scaling. RF also shows a statistically significant decline when L2 Normalization is applied. It reflects that overly aggressive normalization can interfere with the ability of the model to determine effective feature split points even though the impact is significantly less severe than in SVM and KNN.

MinMax Scaling remains competitive particularly for KNN by achieving accuracy values of approximately 0.893. The transformation of all features into a consistent range of [0, 1] allows the method to effectively prevent any single feature from dominating distance calculations. The aim is to ensure more stable and reliable neighbor comparisons for models dependent on Euclidean distance. Furthermore, a significant and compelling aspect of the results is the strong consistency observed between the accuracy and F1-score across experiments as confirmed by Tukey's HSD analysis. The consistency observed in F1-score considering both precision and recall shows that the most effective normalization methods increase the overall number of correct predictions and also maintain a balanced classification performance across all coral categories. Therefore, methods, such as Z-Score Standardization, Robust Scaler, and Power Transformer, enhance global model accuracy and promote fairness and predictive stability across both minority and majority classes. It is a sign that the observed performance improvements are robust and not an artefact of class imbalance.

The evidence summarily reflects the critical role of normalization in the classification of GLCM-based coral images. The selection of the appropriate normalization method is important particularly for algorithms such as SVM and KNN that significantly depend on distance and variance relationships. Meanwhile, L2 Normalization needs to be avoided due to its detrimental impact on the feature magnitude which leads to substantial performance degradation. Z-Score Standardization is generally proven to be the most effective method particularly for highly scale-sensitive models. The strong performance confirms that the transformation of GLCM features into a standardized distribution is key to maximizing the separability of decision boundaries and achieving optimal classification performance.

IV. Conclusion

In conclusion, the research provides a comprehensive evaluation of how different data normalization methods impact the performance of GLCM-based coral image classification using SVM, KNN, and RF. The results show that normalization is not only a supporting preprocessing step but also a decisive factor directly determining the effectiveness of texture-based classification models. Statistical analysis conducted using one-way ANOVA and Tukey's HSD consistently confirms that normalization method selected leads to significant differences in performance across all classifiers ($p < 0.05$).

Then, L2 Normalization has the worst performance compared to the other methods by causing substantial degradation in classification accuracy and F1-score particularly in SVM and KNN. The failure is theoretically grounded in the suppression of the feature magnitude which is an important component of GLCM texture descriptors. This result emphasizes the vulnerability of distance- and kernel-based algorithms to normalization methods that can distort the magnitude relationships within features. However, Z-Score Standardization, Robust Scaler, and Power Transformer are the most effective normalization strategies by frequently achieving statistically indistinguishable performance. The highest accuracy is consistently delivered by Z-Score Standardization in SVM at $\approx$0.913 $\pm$0.0107. It shows that the centering and scaling of features into a standardized distribution maximize kernel separability. Robust Scaler further proves effective by mitigating the impact of outliers, and Power Transformer enhances the performance through the reduction of skewness and improvement of feature distribution symmetry. Moreover, MinMax Scaling remains competitive particularly for KNN to reflect the sufficiency of bounding features within a fixed range for stabilizing distance-based computations.

The results practically translate into a clear guideline where Z-Score Standardization should be used as the default normalization method for GLCM-based coral image classification. Robust Scaler or Power Transformer is also preferable when strong outliers or distribution skewness dominate. Meanwhile, L2 Normalization should be avoided for GLCM texture features. The quantitative analysis shows the possibility of normalization method selected leading to performance differences exceeding 0.60 in F1-score for SVM when comparing the best (Z-Score Standardization) and worst (L2 Normalization). This result reflects the substantial practical impact of preprocessing decisions.

The critical role of data normalization in determining the effectiveness of GLCM-based coral image classification models is emphasized. The results show that the optimal normalization strategy depends both on the characteristics of the dataset and the sensitivity of the classification algorithm. Z-Score Standardization

provides the most reliable performance across SVM, KNN, and RF while L2 Normalization consistently degrades feature discrimination. The trend establishes a clear methodological reference for selecting normalization methods in texture-based coral classification and supports the development of more accurate and scalable machine-learning-driven coral monitoring systems.

However, several limitations should be acknowledged. The analysis is conducted using a single coral image dataset and focuses exclusively on GLCM texture features which can restrict the generalizability of the results to different datasets, imaging conditions, or feature representations. The evaluation is also limited to classical machine learning classifiers while alternative learning paradigms, such as deep learning, can have different interactions with normalization methods. It shows the need for future studies to consider multi-dataset validation, the integration of complementary features such as color and shape descriptors, and a systematic investigation of normalization impacts in deep neural network frameworks. Adaptive or hybrid normalization strategies should also be assessed to further improve robustness in real-world coral reef monitoring applications.

Acknowledgement

The research was supported by the Research Center of Politeknik Negeri Nusa Utara. The authors were grateful to Politeknik Negeri Nusa Utara for providing the facilities and technical support. The authors were also grateful to providers of coral reef image dataset.

Author Contribution

Conceived and designed the analysis, A. P. R. N., S. B. H. S., and T. H.; Collected the data, A. P. R. N.; Contributed data or analysis tools, A. P. R. N., S. B. H. S., M. S., and S. H. W.; Performed the analysis, A. P. R. N., S. B. H. S., M. S., T. H., and S. H. W.; Wrote the paper, A. P. R. N. and S. B. H. S.; Retrieved and prepared the secondary coral dataset from previous research, performed the GLCM analysis, and drafted the manuscript, A. P. R. N.; Supervised the research framework, managed the project administration, and reviewed the final manuscript, S. B. H. S.; Assisted in verifying the statistical analysis and validation of the normalization results, M. S.; Designed the machine learning classification models and evaluated the statistical impact of normalization technique, T. H.; and Provided optimization tools for features analysis and validated the performance metrics of the classifier, S. H. W.

Data Availability

The data that support the findings of the research are available from the corresponding author, Stendy Budi Hartono Sakur, upon reasonable request. The data are currently reserved for further exploratory studies and methodological evaluations by the authors.

References

[1] S. Najeeb, R. A. A. Khan, X. Deng, and C. Wu, "Drivers and consequences of degradation in tropical reef island ecosystems: Strategies for restoration and conservation," *Frontiers in Marine Science*, vol. 12, pp. 1–15, 2025.

[2] M. H. Yuan, K. T. Lin, S. Y. Pan, and C. K. Yang, "Exploring coral reef benefits: A systematic SEEA-driven review," *Science of the Total Environment*, vol. 950, pp. 1–14, 2024.

[3] A. Apprill *et al.*, "Toward a new era of coral reef monitoring," *Environmental Science & Technology*, vol. 57, no. 13, pp. 5117–5124, 2023.

[4] A. Pourzangbar, M. Jalali, and M. Brocchini, "Machine learning application in modelling marine and coastal phenomena: A critical review," *Frontiers in Environmental Engineering*, vol. 2, pp. 1–27, 2023.

[5] A. P. R. Nababan, T. Haryanto, and S. H. Wijaya, "Classification of coral images using support vector machine with gray level co-occurrence matrix feature extraction," *JOIV: International Journal on Informatics Visualization*, vol. 9, no. 3, pp. 936–946, 2025.

[6] S. Anwarul and R. Tanwar, "CoralClassify: Advancing coral health monitoring using deep learning," *Procedia Computer Science*, vol. 259, pp. 88–97, 2025.

[7] D. Menaga, S. L. Deborah, and M. Makizhna, "Deep learning models for coral reef health classification: MobileNetV2, VGG16 and ResNet," in *International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 2024)*. Chennai, India: Atlantis Press, Dec. 30–31, 2025, pp. 1241–1251.

[8] Y. S. Kim, M. K. Kim, N. Fu, J. Liu, J. Wang, and J. Srebric, "Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models," *Sustainable Cities and Society*, vol. 118, pp. 1–12, 2025.

[9] J. M. H. Pinheiro *et al.*, "The impact of feature scaling in machine learning: Effects on regression and classification tasks," *IEEE Access*, vol. 13, pp. 199 903–199 931, 2025.

[10] L. B. V. De Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," *Applied Soft Computing*, vol. 133, 2023.

[11] A. Demircioğlu, "The effect of feature normalization methods in radiomics," *Insights Into Imaging*, vol. 15, pp. 1–11, 2024.

[12] U. Stańczyk, B. Zielosko, and G. Baron, "Importance of characteristic features and their form for data exploration," *Entropy*, vol. 26, no. 5, pp. 1–32, 2024.

[13] G. Kotronoulas *et al.*, "An overview of the fundamentals of data management, analysis, and interpretation in quantitative research," *Seminars in Oncology Nursing*, vol. 39, no. 2, pp. 1–9, 2023.

[14] T. N. Abebe and A. T. Goshu, "Asymmetric generalized error distribution with its properties and applications," *Frontiers in Applied Mathematics and Statistics*, vol. 10, pp. 1–14, 2024.

[15] T. Burger, "Controlling for false discoveries subsequently to large scale one-way ANOVA testing in proteomics: Practical considerations," *Proteomics*, vol. 23, no. 18, pp. 1–12, 2023.

[16] Z. Li, J. Smith-Colin, and J. Wang, "A unified multi-criteria equity metric for data-driven decision-making: Application to tract-level electric vehicle charging station planning," *Engineering Applications of Artificial Intelligence*, vol. 144, pp. 1–19, 2025.

[17] B. H. Liu, L. W. Zhang, Y. Q. Wei, and C. Chen, "Dual power transformation and yeo–johnson techniques for static and dynamic reliability assessments," *Buildings*, vol. 14, no. 11, pp. 1–21, 2024.

[18] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An overview on the advancements of support vector machine models in healthcare applications: A review," *Information*, vol. 15, no. 4, pp. 1–36, 2024.

[19] N. Amaya-Tejera, M. Gamarra, J. I. Vélez, and E. Zurek, "A distance-based kernel for classification via support vector machines," *Frontiers in Artificial Intelligence*, vol. 7, pp. 1–15, 2024.

[20] R. Yang, H. Li, and H. Huang, "A Gaussian kernel similarity approach to multisource information fusion considering the weight of focal element beliefs," 2023. [Online]. Available: https://www.qeios.com/read/N4DNZI

[21] J. Zhang, C. L. Liu, and X. Jiang, "Polynomial kernel learning for interpolation kernel machines with application to graph classification," *Pattern Recognition Letters*, vol. 186, pp. 7–13, 2024.

[22] B. Ahmadi, M. Gholamalifard, S. M. Ghasempouri, and T. Kutser, "Comparative analysis of k-nearest neighbors distance metrics for retrieving coastal water quality based on concurrent in situ and satellite observations," *Marine Pollution Bulletin*, vol. 214, 2025.

[23] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing k-nearest neighbor algorithm: A comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, pp. 1–55, 2024.

[24] Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, "An improved random forest based on the classification accuracy and correlation measurement of decision trees," *Expert Systems with Applications*, vol. 237, Part B, 2024.

[25] H. N. Noura, Z. Allal, O. Salman, and K. Chahine, "An optimized tree-based model with feature selection for efficient fault detection and diagnosis in diesel engine systems," *Results in Engineering*, vol. 27, pp. 1–20, 2025.

[26] C. P. J. Chen, R. R. White, and R. Wright, "Common pitfalls in evaluating model performance and strategies for avoidance in agricultural studies," *Computers and Electronics in Agriculture*, vol. 234, pp. 1–21, 2025.

[27] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Scientific Reports*, vol. 12, pp. 1–9, 2022.

[28] S. Farhadpour, T. A. Warner, and A. E. Maxwell, "Selecting and interpreting multiclass loss and accuracy assessment metrics for classifications with class imbalance: Guidance and best practices," *Remote Sensing*, vol. 16, no. 3, pp. 1–22, 2024.

[29] J. Kozak, B. Probierz, K. Kania, and P. Juszczuk, "Preference-driven classification measure," *Entropy*, vol. 24, no. 4, pp. 1–24, 2022.

[30] M. C. Hinojosa Lee, J. Braet, and J. Springael, "Performance metrics for multilabel emotion classification: Comparing micro, macro, and weighted F1-scores," *Applied Sciences*, vol. 14, no. 21, pp. 1–21, 2024.

[31] X. Sun, Y. Zhang, C. Liu, X. Zheng, and F. Zou, "Evaluating cross-platform normalization methods for integrated microarray and RNA-seq data analysis," 2026. [Online]. Available: https://www.biorxiv.org/content/10.1101/2024.09.30.615938v2

[32] N. A. Rozaini and Z. M. Khalid, "One-way and two-way Analysis of Variance (ANOVA) using parametric and non-parametric methods," *Proceedings of Science and Mathematics*, vol. 22, pp. 53–65, 2024.

[33] J. Rahnenführer *et al.*, "Statistical analysis of high-dimensional biomedical data: A gentle introduction to analytical goals, common approaches and challenges," *BMC Medicine*, vol. 21, pp. 1–54, 2023.

[34] C. E. Agbangba, E. S. Aide, H. Honfo, and R. G. Kakai, "On the use of post-hoc tests in environmental and biological sciences: Acritical review," *Heliyon*, vol. 10, no. 3, pp. 1–12, 2024.

[35] F. Strale Jr, "Partitioning for enhanced statistical power and noise reduction: Comparing one-way and repeated measures Analysis of Variance (ANOVA)," *Cureus*, vol. 16, no. 12, pp. 1–10, 2024.

[36] S. Kwak, "Are only p-values less than 0.05 significant? A p-value greater than 0.05 is also significant!" *Journal of Lipid and Atherosclerosis*, vol. 12, no. 2, pp. 89–95, 2023.

[37] M. H. Kabir *et al.*, "Exploring feature selection and classification techniques to improve the performance of an electroencephalography-based motor imagery brain–computer interface system," *Sensors*, vol. 24, no. 15, pp. 1–27, 2024.