Enhanced Rice Seed Classification Using Stacking and Bagging with Basic Features

Le Minh Dung, Le Van Giap, Bui Duc Vuong

July 1, 2024

Tóm tắt nội dung

This study aims to enhance the classification accuracy of rice seed images using stacking and bagging techniques with 52 basic features. Various machine learning methods were employed, achieving high accuracy rates: Random Forest at 96.03% and Support Vector Machine at 96.44%, etc. Additionally, an Artificial Neural Network (ANN) was implemented, achieving an accuracy of 97.45%. Our proposed approach, which combines these techniques through stacking, resulted in an average accuracy of 97.91%. This significant improvement demonstrates the effectiveness of stacking, bagging in enhancing rice seed classification performance.

1 Introduction

Rice is a staple food for more than half of the world's population, and the quality and purity of rice seeds are crucial for ensuring high yields and maintaining the genetic integrity of rice varieties. Traditionally, the classification of rice seeds has been performed manually by experts based on visual inspection. However, this method is labor-intensive, time-consuming, and subject to human error, which can lead to inconsistencies in seed quality assessment.

In recent years, advancements in computer vision, machine learning and also have opened new avenues for automating the classification of agricultural products, including rice seeds. Machine learning algorithms, Deep Learning combined with image processing techniques, have shown promise in accurately classifying rice seed varieties based on their morphological, color features, texture features and GLCM features. Despite these advancements, achieving high accuracy in rice seed classification remains challenging due to the subtle differences in visual characteristics among different varieties.

This study aims to enhance the accuracy of rice seed classification by leveraging stacking and bagging techniques with basic features. Stacking involves training multiple classifiers and using their predictions as inputs to a meta-classifier, while bagging involves training multiple instances of a classifier on different subsets of the data and averaging their predictions. By combining these ensemble methods, we aim to improve the robustness and accuracy of the classification system.

In this paper, we present a comprehensive approach to rice seed classification that utilizes a set of 54 basic features extracted from rice seed images. We compare the performance of individual machine learning algorithms, such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF), with the performance of ensemble methods using stacking and bagging. The results demonstrate a significant improvement in classification accuracy, highlighting the potential of ensemble learning in agricultural applications.

2 Materials and Methods

2.1 Rice Seed Samples

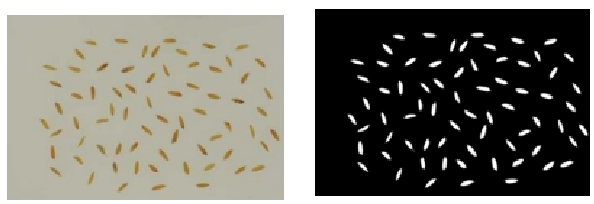
Six commonly cultivated rice seed varieties in Northern Vietnam—BC-15, Hương thơm 1, Nếp-87, Q-5, Thiên ưu-8, and Xi-23—were examined. These rice seeds were sourced from a production company where the varieties were cultivated and harvested under specific conditions to meet standard rice seed production requirements. The sampling was conducted in the Thaibinh and Hanoi regions in Northern Vietnam. [Hong et al., 2015]

2.2 Image Acquisition

We used a Nikon D300S CMOS color camera with a resolution of 640 x 480 pixels to capture images. The setup included a chamber with a white table serving as the background. Within a designated area of 10x16 cm^2, rice seeds were manually spread out. Each image taken by this system typically contained between 30 to 60 seeds. In total, we acquired 212 of these larger images. Our next step involves segmenting individual rice seed images from these acquired images. [Hong et al., 2015].

2.3 Image segmentation

To separate individual rice seed images from the acquired images, we implemented image segmentation. Given that the image background is consistent across all experiments, we selected a threshold method for background subtraction. We found that the blue channel of the images had an intensity that effectively distinguished between the background and the rice seeds. Consequently, we used a threshold method based on the intensity values in the blue channel. Specifically, in the blue channel, the intensity of rice seed pixels is always 90 or less, while the intensity of the background is always greater than 90. During the image segmentation process, pixels with a blue value greater than 90 were assigned a value of 0, and pixels with a blue value of 90 or less were assigned a value of 255. After generating the threshold image, we cropped the rice seed images based on the object contours (Fig. 1), ensuring each image contains only one rice seed within a minimal bounding box. Hereafter, any reference to rice seed images pertains to this processed set of images.[Hong et al., 2015]



a. A sample of acquired mage b. A thresholded image

Hình 1: An example image

2.4 Image description

After segmenting the image of a rice seed, the next step is to compute an image descriptor, which will serve as input for a classifier. An image descriptor encapsulates the characteristics of an image, its regions, or specific locations within the image. These characteristics are commonly referred to as "features."

The research into image description or feature extraction began in the 1960s, and since then, a vast number of image descriptors have been developed. These descriptors can be categorized using various criteria, such as global versus local features, intensity versus derivative, or spectral-based features. Generally, a robust feature should be invariant to changes in rotation, scale, illumination, and viewpoint.

In our study, we focus on several feature types that are representative of two main groups of features: global features (Morphological features, Color, Texture, GIST) and local features (SIFT, HOG, LBP, GLCM). Morphological features are the traditional descriptors used to capture the shape of objects in an image. Color and texture features are particularly useful for distinguishing objects with similar shapes. GIST is a comprehensive global feature calculated using a Gabor filter bank applied to the entire image, and it has proven highly effective for scene classification.

For local features, SIFT (Scale-Invariant Feature Transform), introduced by David Lowe, is renowned for its robustness. SIFT encompasses all the desirable properties of a good feature, including invariance to rotation, scale, and illumination, and it continues to be a prominent feature in the field. HOG (Histogram of Oriented Gradients) is another powerful local feature used to capture edge directions and has been widely used in object detection. LBP (Local Binary Patterns) is effective in texture analysis and describes local spatial patterns. GLCM (Gray-Level Co-occurrence Matrix) is a statistical method used to examine texture by considering the spatial relationship of pixels.

By leveraging these diverse features, our approach aims to create a comprehensive and reliable representation of the segmented rice seed images, ensuring accurate classification and analysis.

2.4.1 Basic descriptor

The basic descriptor is a combination of morphological features, color features, and texture features. We refer to it as the basic descriptor for reference. [Hong et al., 2015]
a) **Morphological descriptors**
Morphological features were extracted from the images of individual rice seeds. A morphological feature descriptor with 15 dimensions was calculated as follows:
1. **Area**: Number of pixels inside the contour
2. **Perimeter**: Arc length of the contour
3. **Length**: Major axis length of the minimum bounding rectangle
4. **Width**: Minor axis length of the minimum bounding rectangle
5. **Length to Width Ratio**:

$$\textbf{Length to Width Ratio} = \frac{\text{Length}}{\text{Width}} \tag{1}$$

6. **Major Axis Length**: Length of the majof axis of the fitted ellipse
7. M**inor Axis Length**: Length of the minor axis of the fitted ellipse
8. **Convex Hull Area**: Area of the convex hull around the contour
9. **Convex Hull Perimeter**: Perimeter of the convex hull around the contour
10. **Solidity** refers to the compactness and firmness of a rice grain

$$\textbf{Length to Width Ratio} = \frac{\text{Area}}{\text{Convex Hull Area}} \tag{2}$$

11. **Aspect Ratio**: The ratio of the width to the length of an object, used to describe its elongation or squashing.

$$\textbf{Length to Width Ratio} = \frac{\text{Width}}{\text{Length}} \tag{3}$$

12. **Extent**: The ratio of the area of an object's bounding box to its actual area, indicating how much it fills its bounding box.

$$\textbf{Extent} = \frac{\text{Bounding Box Area}}{\text{Area}} \tag{4}$$

13. **Circularity**: A measure of how closely an object's shape resembles a perfect circle, calculated using its area and perimeter.compactness and firmness of a rice grain

$$\textbf{Circularity} = 4 \times \pi \times \frac{\text{Area}}{\text{Perimeter}^2} \tag{5}$$

14. **Convexity**: he ratio of the perimeter of an object's convex hull (the smallest convex shape that can contain the object) to its original perimeter, indicating how much the object deviates from being convex.

$$\textbf{Convexity} = \frac{\text{Convex Hull Perimeter}}{\text{Perimeter}} \tag{6}$$

15. **Elongation**: refers to the extent to which an object is stretched or extended in shape

$$\textbf{Elongation} = \frac{\text{Major Axis Length}}{\text{Minor Axis Length}} \tag{7}$$

b) **Color descriptors**
The RGB components of all images were analyzed, and the mean values of the individual channels were computed. The color feature descriptor for rice seed image analysis consists of 26 dimensions, including:
1. **R_avg, G_avg, B_avg**: Average Red, Average Green, Average Blue
2. **Brightness**: Average brightness
3. **NRI, NGI, NBI** : Normalized Red Index, Normalized Green Index, Normalized Blue Index

4. **RS, GS, BS**: Red Squared, Green Squared, Blue Squared
5. **Mean_H, Mean_S, Mean_V**: Mean Hue, Mean Saturation, Mean Value
6. **Std_H, Std_S, Std_V**: Standard Deviation of Hue, Standard Deviation of Value
7. **Hue Variance**: Variance of the Hue
8. **Saturation Variance**: Variance of the Saturation
9. **R Skewness, G Skewness, B Skewness**: Skewness of Red, Skewness of Green, Skewness of Blue
10. **R Kurtosis, G Kurtosis, B Kurtosis**: Kurtosis of Red, Kurtosis of Green, Kurtosis of Blue
11. **Color Entropy**: Entropy of color distribution
12. **Dominant Color**: Most frequent color

c) **Texture descriptors**

Texture feature are calculated as:

- **Mean (m):**

$$m = \sum_{i=1}^{L-1} z_i p(z_i)$$

- **Standard deviation (σ):**

$$\sigma = \sqrt{\sum_{i=1}^{L-1} (z_i - m)^2 p(z_i)}$$

- **Uniformity:**

$$\text{Uniformity} = \sum_{t=0}^{L-1} p^2(z_i)$$

- **Third moment:**

$$\text{Third moment} = \sum_{i=1}^{L-1} (z_i - m)^3 p(z_i)$$

- **Energy:**

$$\text{Energy} = \sum_i \sum_j P(i,j)^2$$

- **Contrast:**

$$\text{Contrast} = \sum_i \sum_j (i-j)^2 \cdot P(i,j)$$

- **Homogeneity:**

$$\text{Homogeneity} = \sum_i \sum_j \frac{P(i,j)}{1 + (i-j)^2}$$

- **Entropy:**

$$\text{Entropy} = -\sum_i \sum_j P(i,j) \cdot \log(P(i,j))$$

- **Correlation:**

$$\text{Correlation} = \frac{\sum_i \sum_j [ij \cdot P(i,j)] - \mu_x \cdot \mu_y}{\sigma_x \cdot \sigma_y}$$

- **Skewness:**

$$\text{Skewness} = \frac{\sum_{i=1}^{L-1} (z_i - m)^3 p(z_i)}{\sigma^3}$$

- **Kurtosis:**

$$\text{Kurtosis} = \frac{\sum_{i=1}^{L-1} (z_i - m)^4 p(z_i)}{\sigma^4}$$

Where z_i is the gray-scale intensity, $p(z_i)$ is the probability density function representing the ratio of the number of pixels that have the intensity z_i to the total number of pixels in the image, and $P(i, j)$ is the gray-level co-occurrence matrix (GLCM) [Kurniati et al., 2024] representing the joint probability of two pixels with specific intensities i and j being spatially adjacent in the image. The texture feature set has 11 components: mean (m), standard deviation (σ), uniformity, third moment, energy, contrast, homogeneity, entropy, correlation, skewness, and kurtosis. These components collectively describe various statistical properties of the pixel intensity distribution and spatial relationships within the image.

Finally, we combine these component descriptors (morphological, color, texture) to obtain a descriptor of 52 dimensions.

2.4.2 GIST descriptor

Oliva and Torralba introduced the GIST descriptor for scene classification. This descriptor captures the overall shape of the scene, the relationships between surfaces and their properties while ignoring specific local objects and their relationships. The primary aim of this method is to develop a low-dimensional representation of the scene that doesn't require any form of segmentation. The structure of the scene is represented through a set of perceptual dimensions: naturalness, openness, roughness, expansion, and ruggedness.

[Hong et al., 2015]To compute the GIST descriptor, an original image is first converted to a grayscale image $I(x, y)$ and normalized. Pre-filtering is applied to $I(x, y)$ to reduce illumination effects and prevent local regions from dominating the energy spectrum. The filtered image $I(x, y)$ is then decomposed using a set of Gabor filters. A 2-D Gabor filter is defined as follows:

$$h(x, y) = e^{-\frac{1}{2}\left(\frac{x^2}{\delta_x^2}+\frac{y^2}{\delta_y^2}\right)} e^{-j2\pi(u_0 x + v_0 y)}$$

The Gabor filters are configured with 4 spatial scales and 8 orientations. At each scale (δ_x, δ_y), the image $I(x, y)$ is passed through a Gabor filter $h(x, y)$, isolating components with energies concentrated near the spatial frequency point (u_0, v_0). Therefore, the GIST vector is computed using an energy spectrum of 32 responses. To reduce the feature vector's dimensionality, each response is averaged over a 4x4 grid, resulting in a GIST feature vector of 512 dimensions [**?**].

2.4.3 SIFT Descriptor

Lowe proposed the scale-invariant feature transform (SIFT), which is resilient to image scaling, translation, rotation, and partially resilient to illumination changes. [Hong et al., 2015]The computation of SIFT features involves four main steps: (1) extraction of scale-space extrema of the Laplacian of Gaussian (LoG); (2) keypoint localization; (3) assignment of canonical orientation; and (4) keypoint description. Initially, local extrema of the Laplacian in scale space are extracted. This is efficiently achieved by constructing a Gaussian pyramid and detecting local extrema of the difference of Gaussians (DoG), making the keypoints invariant to scale changes. These detected points are then re-localized to enhance localization precision. Each point is assigned a canonical orientation to ensure that the subsequent keypoint description is rotation invariant. The keypoints are described by constructing an array of histograms of gradient orientations, resulting in a more compact and significantly more discriminative description than the original image signal. To describe an image using SIFT features, state-of-the-art methods typically employ the Bag of Words (BoW) technique. The size of the descriptor is determined by the predefined vocabulary size in the BoW model, which was set to 200 in our experiments.

2.4.4 HOG Descriptor

The histograms of oriented gradient (HOG) descriptor are extensively utilized in object detection and classification tasks, particularly in person detection, as initially proposed by Dalal and Triggs. Before computing the HOG descriptor, several preprocessing steps are taken to reduce noise and enhance performance. The gradient magnitude $M(x, y)$ and the gradient angle $\alpha(x, y)$ at each pixel are computed within a cell size of 16×16 pixels, a step also referred to as gradient computation. The gradient at pixel coordinates (x, y) is computed as follows:

$$\Delta_x = |G(x-1, y) - G(x+1, y)|$$

$$\Delta_y = |G(x, y-1) - G(x, y+1)|$$

$$M(x, y) = \sqrt{\Delta_x^2 + \Delta_y^2}$$

$$\alpha(x, y) = \arctan\left(\frac{\Delta_y}{\Delta_x}\right)$$

Here, the grayscale value at coordinate (x, y) is denoted as $G(x, y)$, with Δ_x and Δ_y representing the horizontal and vertical gradients, respectively. The dimension of the HOG feature vector depends on the cell size and the number of bin orientations used to create the gradient angle intervals. For our rice seed classification, we use 6 gradient orientations. These gradients are divided into 6 bins, each representing a range of 30 degrees from 0 to 180 degrees for unsigned gradients.

Each gradient feature's magnitude is added to the corresponding bin in the histogram. The histograms from all blocks, where each block contains 2×2 cells with 50% overlap, are then normalized and combined into a feature vector.

2.4.5 LBP Descriptor

Local Binary Patterns (LBP) is a simple yet efficient texture descriptor that labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. It was first introduced by Ojala et al. (1996) and has since been widely used in various applications due to its computational efficiency and robustness to monotonic illumination changes.

The basic LBP operator works in a 3x3 pixel block of an image [Ojala et al., 1996]. The center pixel value is used as a threshold, and the 8 surrounding pixel values are compared to this threshold. If the surrounding pixel value is greater than or equal to the center pixel value, it is set to 1; otherwise, it is set to 0. This results in an 8-bit binary number (or 256 different labels) which is usually converted to a decimal value for convenience.

Formally, the LBP operator can be expressed as:

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p$$

where g_c is the gray value of the center pixel (x_c, y_c), g_p is the gray value of the p-th surrounding pixel, P is the number of neighbors, and $s(x)$ is the sign function defined as:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

The histogram of these LBP values computed over an image or a region of an image is used as a texture descriptor. This histogram effectively captures the distribution of local micro-patterns, making it a powerful feature for texture classification.

To extend the basic LBP to a more robust version, various extensions have been proposed. One common extension is the Circular LBP, where the neighborhood is defined in a circular manner and bilinear interpolation is used to compute the values of neighbors that do not fall exactly on pixel coordinates. Another extension is the Uniform LBP, which reduces the number of possible patterns by considering only patterns that have at most two bitwise transitions from 0 to 1 or vice versa in the circular binary representation.

In our study, we employ the uniform LBP with a circular neighborhood of 8 pixels (P=8) and a radius of 1 (R=1). This approach results in 59 possible patterns (including uniform and non-uniform patterns), which significantly reduces the feature dimensionality while retaining discriminative power. The LBP histogram is computed for each segmented rice seed image and is then used as an input feature for the classifier.

By utilizing the LBP descriptor, we can capture fine texture details of the rice seed surface, which are essential for distinguishing between different rice varieties. The combination of LBP with other global and local features ensures a comprehensive representation of the rice seed images, leading to improved classification accuracy.

2.4.6 GLCM Descriptor

Grey Level Co-occurrence Matrix (GLCM) textures as originally described by Haralick and others in 1973.

The Gray Level Co-occurrence Matrix (GLCM) descriptor is a powerful tool for texture analysis and has been applied effectively in the classification of rice seeds. The GLCM is constructed by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image.

Correlation: This feature measures the correlation between pixel pairs. It provides information about the linear dependency of gray levels in the co-occurrence matrix. High correlation values indicate that the pixel intensities are highly correlated.

Entropy: Entropy is a measure of randomness in the image texture. It quantifies the complexity or disorder present in the image. Higher entropy values indicate more complexity.

Homogeneity: Homogeneity assesses the similarity of pixel values. It is higher when the distribution of elements in the GLCM is concentrated near the diagonal, indicating that the image has uniform texture.

Energy: Also known as angular second moment, energy measures the uniformity of the image texture. Higher energy values indicate that the gray level distribution has either constant or periodic patterns.

Contrast: Contrast measures the intensity difference between a pixel and its neighbor over the entire image. It indicates the amount of local variations present in the image. Higher contrast values signify more variation in intensity.

The formula to calculate these 5 features are above on part c) Texture descriptors.

3 Image classification

After extracting features, a classifier or an ensemble classifier is trained to distinguish between different rice varieties. Below, we discuss some notable classification models:

3.1 Support vector machine

The fundamental concept of support vector machine (SVM) is to find an optimal hyper-plane that separates linearly separable patterns in a high-dimensional feature space. The goal is to identify the hyper-plane that maximizes the margin around the separating hyper-plane. This is determined by the support vectors, which are the data points closest to the decision surface and influence the position of the hyper-plane. SVMs can also classify patterns that are not linearly separable by transforming the original data into a higher-dimensional space using a kernel function, making the classes linearly separable in that space. SVMs are powerful and widely used classifiers.

3.2 K-Nearest Neighbors

K-nearest neighbor (KNN) is a classification method that classifies a new sample based on the majority class of its **k** nearest neighbors. This technique is popular in classification problems due to its simplicity, effectiveness, and non-parametric nature.

3.3 Random-Forest

Random Forest (RF) is a classification technique introduced by Breiman, which constructs an ensemble of decision trees. For each tree, RF uses a different bootstrap sample of the response variable and modifies the tree construction process: each node is split using the best predictor from a random subset of predictors chosen at that node, and the tree is grown to its maximum extent without pruning. For new data prediction, RF aggregates the outputs of all the trees. RF is efficient, handles large datasets well, and performs competitively compared to many other classifiers, including discriminant analysis, SVMs, and neural networks, while being robust against overfitting.

3.4 Logistic Regression

Logistic Regression is a statistical method used for binary classification problems. It models the probability that a given input belongs to a particular class. The logistic function (sigmoid function) is used to map predicted values to probabilities. Despite its simplicity, Logistic Regression is effective for linearly separable data and serves as a baseline for more complex models. It can also be extended to multiclass classification through techniques like one-vs-rest and softmax regression.

3.5 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) are inspired by the biological neural networks that constitute animal brains. ANNs consist of interconnected nodes (neurons) organized in layers: an input layer, one or more hidden layers, and an output layer. Each connection has an associated weight that is adjusted during the training process to minimize the error in predictions. ANNs are highly flexible and can model complex relationships in data. They have been successfully applied in various domains, including image and speech recognition, and are especially powerful in handling non-linear patterns.

4 Performance metrics

Formulas for performance metrics

Performance Metrics	Formula
Accuracy	*((TP + TN)/n)*100*
Precision	*TP/(TP + FP)*
Recall	*TP/(TP + FN)*
F-1 Score	*2TP/(2TP + FP + FN)*

The study utilized a confusion matrix to assess the performance of the classification models. From this matrix, commonly used performance metrics such as accuracy, precision, recall, and F-1 score were derived. These metrics rely on four key values within the confusion matrix: True Positive (TP) and True Negative (TN) represent correctly predicted positive and negative samples, respectively, while False Positive (FP) and False Negative (FN) denote incorrectly predicted positive and negative samples. Formulas provided in the Table below were employed to compute these performance metrics based on these values.

5 Proposed model

In this study, we propose an enhanced rice seed classification model that leverages ensemble learning techniques, specifically stacking and bagging, to improve classification accuracy

Bagging: An abbreviation for Bootstrap Aggregating, is a machine learning ensemble strategy for enhancing the reliability and precision of predictive models. It entails generating numerous subsets of the training data by employing random sampling with replacement. These subsets train multiple base learners, such as decision trees, neural networks, or other models. During prediction, the outputs of these base learners are aggregated, often by averaging (for regression tasks) or voting (for classification tasks), to produce the final prediction. Bagging helps to reduce overfitting by introducing diversity among the base learners and improves the overall performance by reducing variance and increasing robustness.

Stacking: Involves creating a new classification model by combining multiple classifiers. Regardless of the number of classifiers, learning is performed using the predictions from these classifiers along with the existing dataset to create a new model. This new model generates predictions by leveraging the combined strength of the individual classifiers, although it can sometimes yield lower classification results than its constituent classifiers. In our study, predictions from the ANN, SVM, KNN, RF, and LR models are used as inputs for the Stacking model.

Our approach begins with the extraction of basic features from rice seed images, encompassing morphological, color, and texture descriptors. The model training is conducted in two distinct phases:

- **Phase 1**: In this initial phase, we train the models using different machine learning algorithms and an Artificial Neural Network (ANN) with three variations in the basic feature set which is **22, 36, 52** features, global features and local features: The machine learning algorithms employed include K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF). These classifiers, along with the ANN, are trained separately using the specified feature sets.
- **Phase 2**: In the second phase, we implement ensemble learning techniques, specifically stacking and bagging, with the same three variations in the feature sets:
  - 22 features
  - 36 features
  - 52 features.

To implement stacking, we combine the predictions from the base classifiers trained in Phase 1. These base classifiers include KNN, SVM, RF, and ANN. The combined predictions are then used as input for a meta-classifier, which in our model is a Logistic Regression (LR) classifier, chosen for its ease of implementation and interpretability.

5.1 Experiment set up

To experimenting all the experiments, we used a computer with 64bit Window 11, core i5, CPU 4.00 GHz (1 CPUs) and 24 GB main memory.

To create a dataset for each variety of rice seeds, we selected all examples with positive labels. For negative labels, we chose five other rice seed images per positive example to balance the numbers. This ensured that the number of positive and negative labeled examples was roughly equal. To ensure a fair comparison of different classification methods, we kept the test and training sets fixed and used the Out-Of-Bag technique to estimate generalization error . Approximately 67% of the samples for each rice variety were randomly assigned to the training set, while the remaining samples were used for the test set in classification tasks.

Rice variety	Number of individual rice seed images
BC-15	3680
Hương thơm 1	4152
Nếp-87	2877
Q-5	3019
Thiên ưu-8	2011
Xi-23	4152

Bảng 1: Description of rice seed image dataset

Initially, to classify rice seeds using KNN, SVM, LR and RF with ANN methods, we began by extracting various types of features: global features include: Morphological features, Color, Texture, and GIST, as well as local features include: LBP, GLCM, HOG, SIFT. Following the training phase, these classification models were tested on the designated test datasets. The performance metrics and accuracy of these methods were reported in **Tables** 2, 3, 4, 5, 6, 7, 8, 9 and a notation is that **Table 2, 3, 4** represented performance metrics and accuracy of these methods with correspond number of basic features is 22, 36, 52 .

With the K-Nearest Neighbors (KNN) method, one of the most critical parameters is the choice of a suitable value for K. In our experiment, we tested different values of K, ranging from 1 to 55, and found that the KNN model achieved the best results with K set to 23.

For the **Support Vector Machine (SVM)**, we utilized both linear and radial basis function (RBF) kernels. The linear SVM provided a straightforward decision boundary, while the RBF kernel allowed for more complex, non-linear decision boundaries.

Regarding the **Random Forest (RF)** method, it is essential to specify two parameters to train the model effectively: the number of trees (ntree) to be constructed in the forest and the number of input variables randomly sampled as candidates at each node (mtry). In our experiments, we used ntree = 500 and adjusted

$$mtry = \sqrt{\text{number of features}} \tag{8}$$

accordingly for different feature sets.

For the **Artificial Neural Network (ANN)**, we configured a model with 100 hidden layers, which enabled it to capture complex patterns and interactions within the data.

In addition, we employed **Logistic Regression (LR)** with a regularization parameter C set to 10, allowing for effective handling of the trade-off between maximizing the likelihood and controlling model complexity.

5.2 Experiment result

5.2.1 Phase 1

Like we metion earlier, **phase 1** we will train the models using different machine learning algorithms and an Artificial Neural Network (ANN) with three variations in the basic feature set which is **22, 36, 52** features, global features and local features:

Bảng 2: Basic Feature Comparison of Various Models Using 22 Features

Seed	22 Basic Features									
	KNN		SVM		RF		Logistic Regression		ANN	
	F	Acc	F	Acc	F	Acc	F	Acc	F	Acc
BC-15	88.56%	87.50%	90.63%	89.95%	89.86%	89.70%	86.43%	86.24%	92.01%	92.01%
Hương thơm 1	91.23%	90.48%	92.66%	92.17%	91.54%	90.96%	88.12%	87.74%	93.87%	93.87%
Nếp-87	95.19%	95.30%	95.74%	97.12%	96.58%	96.70%	97.58%	97.58%	96.94%	96.94%
Q-5	92.97%	92.36%	94.68%	94.68%	94.69%	94.52%	91.65%	91.65%	95.87%	95.87%
Thiên ưu-8	95.65%	95.52%	96.81%	96.77%	95.31%	95.27%	97.01%	97.01%	97.73%	97.73%
Xi-23	92.74%	91.85%	95.96%	95.42%	94.16%	93.49%	89.77%	89.77%	96.20%	96.20%
Average	**92.72%**	**92.16%**	**94.68%**	**94.37%**	**93.69%**	**93.44%**	**91.66%**	**91.66%**	**95.44%**	**95.44%**

Bảng 3: Performance Comparison of Different Models Using 36 Basic Features

Seed	36 Basic Features									
	KNN		SVM		RF		Logistic Regression		ANN	
	F	**Acc**	**F**	**Acc**	**F**	**Acc**	**F**	**Acc**	**F**	**Acc**
BC-15	88.97%	88.06%	91.43%	91.43%	91.91%	91.93%	88.14%	88.14%	93.33%	93.33%
Hương thơm 1	94.56%	94.23%	97.23%	97.23%	96.12%	95.91%	96.74%	96.57%	97.74%	97.74%
Nếp-87	95.11%	95.05%	98.20%	98.31%	95.83%	96.10%	98.11%	98.10%	97.89%	97.89%
Q-5	93.73%	93.26%	97.00%	97.04%	94.93%	94.97%	92.76%	92.76%	97.59%	97.59%
Thiên ưu-8	95.73%	95.62%	97.28%	97.28%	98.06%	98.04%	97.76%	97.76%	98.04%	98.04%
Xi-23	93.87%	93.06%	96.34%	95.83%	95.61%	95.10%	92.25%	92.25%	95.98%	95.98%
Average	**93.66%**	**93.21%**	**96.06%**	**96.06%**	**95.41%**	**95.34%**	**94.26%**	**94.26%**	**96.76%**	**96.76%**

Bảng 4: Performance Comparison of Different Models Using 52 Basic Features

Seed	**KNN**		**SVM**		**RF**		**Logistic Regression**		**ANN**	
	F	**Acc**	**F**	**Acc**	**F**	**Acc**	**F**	**Acc**	**F**	**Acc**
BC-15	89.22%	88.22%	92.16%	91.93%	91.94%	91.93%	90.84%	90.61%	94.81%	94.81%
Hương thơm 1	95.60%	95.33%	97.78%	97.66%	97.44%	97.37%	97.42%	97.30%	98.54%	98.54%
Nếp-87	95.93%	95.89%	98.10%	98.10%	97.40%	97.47%	98.21%	98.21%	98.21%	98.21%
Q-5	93.45%	92.96%	96.99%	96.88%	95.09%	95.17%	94.98%	94.77%	98.29%	98.29%
Thiên ưu-8	96.04%	95.92%	98.05%	98.04%	97.50%	97.43%	98.65%	98.64%	98.04%	98.04%
Xi-23	94.01%	93.20%	96.45%	96.05%	96.96%	96.78%	94.22%	93.57%	96.78%	96.78%
Average	94.04%	93.59%	96.59%	96.44%	96.06%	96.03%	95.72%	95.52%	97.45%	97.45%

Phase 1: Analysis

- **Performance with Basic Features:**

Bảng 5: Performance Comparison of Different Models Using GLCM Features

Seed	KNN		SVM		RF		Logistic Regression		ANN	
	F	Acc	F	Acc	F	Acc	F	Acc	F	Acc
BC-15	72.33%	69.69%	72.79%	67.79%	80.46%	80.56%	62.85%	62.60%	72.45%	72.65%
Hương thơm 1	80.36%	79.20%	80.51%	78.76%	86.14%	85.77%	78.23%	76.35%	80.28%	80.29%
Nếp-87	92.15%	91.99%	93.53%	93.36%	83.53%	83.88%	93.63%	93.47%	93.78%	93.78%
Q-5	73.44%	71.33%	74.46%	70.52%	70.23%	69.72%	71.74%	70.12%	75.22%	75.55%
Thiên ưu-8	83.94%	83.23%	86.24%	85.20%	79.94%	78.85%	86.09%	85.80%	87.90%	87.92%
Xi-23	77.49%	74.34%	72.19%	68.57%	75.67%	73.39%	69.60%	63.67%	78.01%	78.14%
Average	**79.95%**	**78.30%**	**79.95%**	**77.37%**	**79.33%**	**78.70%**	**77.02%**	**75.34%**	**81.27%**	**81.39%**

Bảng 6: Performance metrics for different classifiers on various seeds using GIST features.

Seed	**KNN**		**SVM**		**RF**		**Logistic Regression**		**ANN**	
	F	**Acc**	**F**	**Acc**	**F**	**Acc**	**F**	**Acc**	**F**	**Acc**
BC-15	70.31%	72.24%	74.29%	74.63%	74.76%	74.46%	66.28%	66.64%	76.94%	76.94%
Hương thơm 1	86.10%	84.82%	88.15%	87.81%	84.80%	84.74%	79.94%	79.34%	89.92%	89.93%
Nếp-87	82.82%	80.19%	91.19%	91.04%	88.37%	88.41%	86.66%	86.41%	91.36%	91.36%
Q-5	78.25%	76.46%	78.74%	78.27%	76.05%	76.56%	71.04%	70.22%	79.64%	79.68%
Thiên ưu-8	90.91%	91.24%	93.65%	93.50%	91.05%	91.24%	91.29%	91.09%	93.80%	93.81%
Xi-23	79.46%	76.46%	83.06%	80.92%	80.08%	76.75%	75.35%	71.49%	85.69%	85.67%
Average	81.31%	80.24%	84.85%	84.36%	82.52%	82.03%	78.43%	77.53%	86.23%	86.23%

Bảng 7: Performance metrics for different classifiers on various seeds using HOG features.

Seed	**KNN**		**SVM**		**RF**		**Logistic Regression**		**ANN**	
	F	**Acc**	**F**	**Acc**	**F**	**Acc**	**F**	**Acc**	**F**	**Acc**
BC-15	77.95%	79.82%	87.76%	87.15%	85.90%	85.50%	82.50%	82.45%	86.40%	86.41%
Hương thơm 1	84.43%	81.75%	91.78%	91.53%	87.72%	87.59%	86.32%	85.84%	92.41%	92.41%
Nếp-87	85.93%	83.98%	92.29%	92.20%	87.77%	88.20%	90.09%	89.99%	92.62%	92.62%
Q-5	78.40%	77.67%	82.59%	81.09%	78.11%	76.66%	76.08%	75.45%	80.98%	80.99%
Thiên ưu-8	90.09%	90.33%	96.21%	96.07%	94.51%	94.26%	95.25%	95.17%	95.46%	95.47%
Xi-23	83.21%	79.97%	88.27%	86.77%	85.71%	84.36%	84.64%	82.75%	89.48%	89.47%
Average	83.34%	82.25%	89.82%	89.14%	86.62%	86.10%	85.81%	85.28%	89.56%	89.56%

Bảng 8: Performance metrics for different classifiers on various seeds using SIFT features.

Seed	**KNN**		**SVM**		**RF**		**Logistic Regression**		**ANN**	
	R	**Acc**	**R**	**Acc**	**R**	**Acc**	**R**	**Acc**	**R**	**Acc**
BC-15	56.20%	72.41%	75.20%	79.49%	79.77%	79.24%	72.62%	76.84%	76.77%	76.77%
Hương thơm 1	79.78%	81.68%	83.82%	86.93%	85.23%	85.18%	85.22%	85.63%	85.18%	85.18%
Nếp-87	93.08%	79.77%	83.65%	85.14%	84.73%	83.46%	83.65%	84.00%	83.98%	83.98%
Q-5	54.64%	63.88%	71.60%	71.73%	70.35%	70.42%	67.85%	69.49%	69.92%	69.92%
Thiên ưu-8	85.97%	76.13%	85.37%	80.97%	78.34%	77.19%	82.99%	79.32%	77.79%	77.79%
Xi-23	52.55%	65.86%	73.86%	73.83%	81.67%	74.63%	73.99%	75.27%	72.70%	72.66%
Average	70.37%	73.29%	78.92%	79.68%	80.02%	78.35%	77.72%	77.78%	77.72%	77.72%

Bảng 9: Performance metrics for different classifiers on various seeds using LBP features.

Seed	KNN		SVM		RF		Logistic Regression		ANN	
	R	Acc	R	Acc	R	Acc	R	Acc	R	Acc
BC-15	69.89%	67.38%	74.72%	71.50%	67.97%	67.46%	64.09%	66.64%	69.93%	69.93%
Hương thơm 1	77.27%	73.43%	76.43%	74.67%	71.45%	72.92%	67.64%	68.03%	74.74%	74.74%
Nếp-87	82.81%	86.20%	85.95%	88.20%	85.10%	87.67%	85.53%	87.46%	87.46%	87.46%
Q-5	71.60%	69.22%	77.12%	71.33%	69.86%	67.20%	68.44%	66.90%	73.04%	73.04%
Thiên ưu-8	83.88%	85.50%	88.06%	88.07%	83.49%	85.95%	88.36%	87.46%	86.71%	86.71%
Xi-23	83.01%	70.83%	85.62%	74.34%	78.01%	72.81%	75.16%	68.06%	72.88%	72.88%
Average	**78.08%**	**75.43%**	**81.32%**	**78.02%**	**75.98%**	**75.67%**	**74.87%**	**74.09%**	**77.46%**	**77.46%**

- Performance tends to increase with the number of basic features.
- **ANN** consistently performs well, showing high precision, recall, F1-score, and accuracy.
- **SVM and RF** also perform strongly, particularly as the number of features increases.

  - **KNN and Logistic Regression** show competitive results but are slightly behind SVM, RF, and ANN.

- **Texture and Other Features (GLCM, GIST, HOG, SIFT, LBP):**
  - Performance metrics vary significantly with different feature sets.
  - **ANN** tends to perform well across different feature sets.
  - **GLCM and LBP** features show decent performance but generally lower than basic features.
  - **GIST and SIFT** features show variability in performance, with SIFT generally performing lower than others.
  - **HOG** features also show strong performance, particularly with ANN and SVM.

- **Best Overall Performance:** ANN consistently shows the highest metrics across most feature sets, particularly with a higher number of basic features.

- **Strong Contenders:** SVM and RF also perform very well, often close to or matching ANN in many cases. Moderate Performers: KNN and Logistic Regression are generally competitive but tend to lag behind ANN, SVM, and RF.

- **Feature Impact:** Increasing the number of basic features improves model performance. Texture features like GLCM, HOG, and LBP provide valuable insights but generally offer lower performance compared to basic features.

5.2.2 Phase 2

Phase 2: In this phase, we implement ensemble learning techniques, specifically focusing on stacking and bagging, using the same three variations in the feature sets:

- 22 features
- 36 features
- 52 features.

For stacking, we combine the predictions from the base classifiers trained in Phase 1. These base classifiers include KNN, SVM, RF, and ANN. The combined predictions serve as input for a meta-classifier. In our model, we use a Logistic Regression (LR) classifier as the meta-classifier due to its ease of implementation and interpretability.

For bagging, we use the same base classifiers (KNN, SVM, RF, and ANN) but create multiple instances of these classifiers to form an ensemble. Specifically, we set the number of estimators (**n_estimators**) in the bagging technique to **5**, ensuring a robust aggregation of predictions from the ensemble.

Here is the result of **Bagging and Stacking after the same three variations in the basic feature sets** :

- With 22 Basic Features:
- With 36 Features
- With 52 Features

5.3 Phase 2 Analysis

- Table 10
  - **Bagging** achieves relatively high average performance with Precision at **94.93%**, Recall at **94.79%,** F1-Score at **94.82%**, and Accuracy at **94.85%**. The "Nếp-87"type shows the highest performance with P, R, F, and Accuracy all at **97.79%.** The lowest performance is observed in the "BC-15"type with P at 91.58%, R at 91.17%, F at **91.23%**, and Accuracy at **91.27%**.

Bảng 10: Performance Using Basic (22 features)

	Bagging				Stacking			
	P	**R**	**F**	**Acc**	**P**	**R**	**F**	**Acc**
BC-15	91.58%	91.17%	91.23%	91.27%	94.67%	94.65%	94.64%	94.65%
Hương thơm 1	93.25%	93.31%	93.27%	93.28%	98.69%	98.69%	98.69%	98.69%
Nếp-87	97.79%	97.79%	97.79%	97.79%	98.12%	98.10%	98.10%	98.10%
Q-5	94.75%	94.51%	94.55%	94.57%	98.10%	98.09%	98.09%	98.09%
Thiên ưu-8	97.29%	97.28%	97.28%	97.28%	98.34%	98.34%	98.34%	98.34%
Xi-23	94.93%	94.69%	94.80%	94.85%	96.71%	96.71%	96.71%	96.71%
Average	94.93%	94.79%	94.82%	94.85%	97.44%	97.43%	97.43%	97.43%

Bảng 11: Performance Using 36 Features

	Bagging				Stacking			
	P	**R**	**F**	**Acc**	**P**	**R**	**F**	**Acc**
BC-15	93.21%	93.02%	93.07%	93.08%	94.67%	94.65%	94.64%	94.65%
Hương thơm 1	98.34%	98.30%	98.32%	98.32%	98.69%	98.69%	98.69%	98.69%
Nếp-87	98.42%	98.42%	98.42%	98.42%	98.12%	98.10%	98.10%	98.10%
Q-5	97.09%	96.94%	96.98%	96.98%	98.10%	98.09%	98.09%	98.09%
Thiên ưu-8	98.04%	98.03%	98.04%	98.04%	98.34%	98.34%	98.34%	98.34%
Xi-23	96.26%	96.33%	96.30%	96.35%	96.71%	96.71%	96.71%	96.71%
Average	96.89%	96.84%	96.86%	96.87%	97.44%	97.43%	97.43%	97.43%

Bảng 12: Performance Metrics Using 52 Features

	Bagging (52 features)				Stacking (52 features)			
	P	**R**	**F**	**Acc**	**P**	**R**	**F**	**Acc**
BC-15	93.42%	93.28%	93.32%	93.33%	95.31%	95.30%	95.30%	95.30%
Hương thơm 1	98.17%	98.17%	98.17%	98.18%	98.61%	98.61%	98.61%	98.61%
Nếp-87	98.21%	98.21%	98.21%	98.21%	98.84%	98.84%	98.84%	98.84%
Q-5	97.76%	97.65%	97.68%	97.69%	98.69%	98.69%	98.69%	98.69%
Thiên ưu-8	98.34%	98.34%	98.34%	98.34%	98.64%	98.64%	98.64%	98.64%
Xi-23	96.51%	96.52%	96.52%	96.56%	97.37%	97.37%	97.37%	97.37%
Average	97.07%	97.03%	97.04%	97.05%	97.91%	97.91%	97.91%	97.91%

- **Stacking** shows higher average performance compared to Bagging with Precision at **97.44%**, Recall at **97.43%**, F1-Score at **97.43%**, and Accuracy at **97.43%**.
- The **"Hương thơm 1"** type achieves the highest performance with P, R, F, and Acc all at 98.69%. Similar to Bagging, the lowest performance in Stacking is for the "BC-15"type with P at 94.67%, R at 94.6%, F at 94.64%, and Accuracy at 94.65%.
- **Comparison between Bagging and Stacking:**
- Stacking consistently outperforms Bagging across all rice seed types and performance metrics.
- Notably, the largest performance gap is observed in the "Hương thơm 1"and "Nếp-87"types, indicating that Stacking has a superior classification capability compared to Bagging.
- **Conclusion:**
- The Stacking model demonstrates better performance than the Bagging model in classifying rice seeds using 22 basic features.
- With higher Precision, Recall, F1-Score, and Accuracy, Stacking is the preferred choice for this classification task.

- The significant improvement seen with Stacking suggests that combining multiple models enhances the classification accuracy and system reliability.
- **Summary:**
- **Bagging**: Good performance but inferior to Stacking.
- **Stacking**: Significantly better performance, especially with high-performing rice types like "Hương thơm 1"and "Nếp-87".

- Table 11
  - **Bagging**
    - **BC-15**: Achieved performance with Precision (P) of 93.21%, Recall (R) of 93.02%, F1-Score (F) of 93.07%, and Accuracy (Acc) of 93.08%. These metrics indicate stability but are not particularly high, only reaching an average level compared to other rice types.
    - **Hương thơm 1**: Precision 98.34%, Recall 98.30%, F1-Score 98.32%, and Accuracy 98.32%, showing very high and stable performance.
    - **Nếp-87**: Achieved Precision 98.42%, Recall 98.42%, F1-Score 98.42%, and Accuracy 98.42%, indicating high accuracy and reliability.
    - **Q-5**: Precision 97.09%, Recall 96.94%, F1-Score 96.98%, and Accuracy 96.98%, good metrics but not as high as Hương thơm 1 and Nếp-87.
    - **Thiên ưu-8**: Precision 98.04%, Recall 98.04%, F1-Score 98.04%, and Accuracy 98.04%, very high and stable.
    - **Xi-23**: Precision 96.26%, Recall 96.39%, F1-Score 96.30%, and Accuracy 96.35%, slightly lower than other rice types.
    - **Average**: Precision 96.89%, Recall 96.84%, F1-Score 96.86%, and Accuracy 96.87%, indicating that Bagging maintains high and consistent performance across different rice types.
  - **Stacking**
    - **BC-15**: Precision 94.67%, Recall 94.65%, F1-Score 94.64%, and Accuracy 94.65%, higher than Bagging, showing significant improvement.
    - **Hương thơm 1**: Precision 98.69%, Recall 98.69%, F1-Score 98.69%, and Accuracy 98.69%, very high and consistent performance.
    - **Nếp-87**: Precision 98.12%, Recall 98.10%, F1-Score 98.10%, and Accuracy 98.10%, almost equivalent to Hương thơm 1.
    - **Q-5**: Precision 98.10%, Recall 98.09%, F1-Score 98.09%, and Accuracy 98.09%, higher than Bagging.
    - **Thiên ưu-8**: Precision 98.34%, Recall 98.34%, F1-Score 98.34%, and Accuracy 98.34%, very high and stable.
    - **Xi-23**: Precision 96.71%, Recall 96.71%, F1-Score 96.71%, and Accuracy 96.71%, higher than Bagging.
    - **Average**: Precision 97.44%, Recall 97.43%, F1-Score 97.43%, and Accuracy 97.43%, indicating that Stacking not only maintains high performance but also significantly improves over Bagging.
      - **Conclusion**
      - **Bagging**: This method shows very high and consistent performance across most rice types, with average metrics all above 96%.
      - **Stacking**: This method performs even better than Bagging, with average metrics all above 97%, showing a clear improvement in performance.

- **Table 11**
  - **Bagging**:

    * **BC-15**: Achieved performance with Precision (P) of 93.42%, Recall (R) of 93.28%, F1-Score (F) of 93.32%, and Accuracy (Acc) of 93.33%. These metrics indicate stability but are not particularly high, only reaching an average level compared to other rice types.
    * **Hương thơm 1**: Precision 98.17%, Recall 98.17%, F1-Score 98.17%, and Accuracy 98.18%, showing very high and stable performance.
    * **Nếp-87**: Achieved Precision 98.21%, Recall 98.21%, F1-Score 98.21%, and Accuracy 98.21%, indicating high accuracy and reliability.
    * **Q-5**: Precision 97.76%, Recall 97.65%, F1-Score 97.68%, and Accuracy 97.69%, good metrics but not as high as Hương thơm 1 and Nếp-87.
    * **Thiên ưu-8**: Precision 98.34%, Recall 98.34%, F1-Score 98.34%, and Accuracy 98.34%, very high and stable.
    * **Xi-23**: Precision 96.51%, Recall 96.52%, F1-Score 96.52%, and Accuracy 96.56%, slightly lower than other rice types.
    * **Average**: Precision 97.07%, Recall 97.03%, F1-Score 97.04%, and Accuracy 97.05%, indicating that Bagging maintains high and consistent performance across different rice types.
  - **Stacking**:
    * **BC-15**: Precision 95.31%, Recall 95.30%, F1-Score 95.30%, and Accuracy 95.30%, higher than Bagging, showing significant improvement.
    * **Hương thơm 1**: Precision 98.61%, Recall 98.61%, F1-Score 98.61%, and Accuracy 98.61%, very high and consistent performance.
    * **Nếp-87**: Precision 98.84%, Recall 98.84%, F1-Score 98.84%, and Accuracy 98.84%, almost equivalent to Hương thơm 1.
    * **Q-5**: Precision 98.69%, Recall 98.69%, F1-Score 98.69%, and Accuracy 98.69%, higher than Bagging.
    * **Thiên ưu-8**: Precision 98.64%, Recall 98.64%, F1-Score 98.64%, and Accuracy 98.64%, very high and stable.
    * **Xi-23**: Precision 97.37%, Recall 97.37%, F1-Score 97.37%, and Accuracy 97.37%, higher than Bagging.
    * **Average**: Precision 97.91%, Recall 97.91%, F1-Score 97.91%, and Accuracy 97.91%, indicating that Stacking not only maintains high performance but also significantly improves over Bagging.

Summary

- **Bagging**: This method shows very high and consistent performance across most rice types, with average metrics all above 96%.
- **Stacking**: This method performs even better than Bagging, with average metrics all above 97%, showing a clear improvement in performance.

6 CONCLUSION AND FUTURE WORKS

In this study, we focused on analyzing visual features of rice seed images, including color, shape, texture, GIST, SIFT, LBP, GLCM, and HOG. Various classification models were applied to these features, demonstrating that image processing techniques can be effectively combined with classification techniques such as KNN, SVM, and RF to identify rice seeds in mixed samples. The RF method, using simple features, exhibited the highest classification capability and accuracy, achieving an average of 90.27% in precision and 90.54% in recall.

Our analysis showed that performance tends to increase with the number of basic features. Specifically, ANN consistently performed well across all metrics, while SVM and RF also showed strong performance, particularly with more features. KNN and Logistic Regression were competitive but slightly behind SVM, RF, and ANN.

When examining texture and other features (GLCM, GIST, HOG, SIFT, LBP), the performance varied significantly. ANN tended to perform well across different feature sets, while GLCM and LBP features showed decent performance, albeit generally lower than basic features. GIST and SIFT features exhibited variability, with SIFT performing lower than others. HOG features demonstrated strong performance, especially with ANN and SVM.

The performance of both Bagging and Stacking models improved with an increased number of features. For instance, Bagging's average accuracy rose from 94.85% to 96.87% when increasing from 22 to 36 features, and to 97.05% with 52 features. Similarly, Stacking achieved the highest average accuracy of 97.91% with 52 features, up from 97.43% with fewer features. These findings suggest that incorporating a larger number of features can significantly enhance the accuracy and stability of classification models.

Our work has the potential to be deployed in rice seed production plants in Vietnam, aiding in the quality assessment of rice seeds through improved classification accuracy and efficiency. Further investigation into additional feature types and classification models could continue to improve performance.

Tài liệu

[Hong et al., 2015] Hong, P. T. T., Hai, T. T. T., Hoang, V. T., Hai, V., Nguyen, T. T., et al. (2015). Comparative study on vision based rice seed varieties identification.

[Kurniati et al., 2024] Kurniati, F. T., Manongga, D. H., Sediyono, E., Prasetyo, S. Y. J., and Huizen, R. R. (2024). Glcm-based feature combination for extraction model optimization in object detection using machine learning. *arXiv preprint arXiv:2404.04578*.

[Ojala et al., 1996] Ojala, T., Pietikainen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59.