

# Revolutionizing Rice Leaf Disease Detection: Next-Generation SMOREF-SVM Integrating Spider Monkey Optimization and Advanced Machine Learning Techniques

Avip Kurniawan<sup>1</sup>, Tri Retnaningsih Soeprbowati<sup>2</sup>, Budi Warsito<sup>3</sup>

Department of Information System-School of Postgraduate Studies, Universitas Diponegoro, Semarang, Indonesia<sup>1</sup>

Department of Informatics-Faculty of Engineering, Universitas Krisnadwipayana, Jakarta, Indonesia<sup>1</sup>

Department of Biology-Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia<sup>2</sup>

Cluster for Paleolimnology, Universitas Diponegoro, Semarang, Indonesia<sup>2</sup>

Department of Statistics-Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia<sup>3</sup>

**Abstract**—Leaf diseases pose a significant challenge to rice productivity, which is critical as rice is a staple food for over half of the world's population and a major agricultural commodity. These diseases can lead to severe economic losses and jeopardize food security, particularly in regions heavily reliant on rice farming. Traditional detection methods, such as visual inspection and microscopy, are often inadequate for early disease identification, which is crucial for effective management and minimizing yield loss. This presentation introduces SMOREF-SVM, a novel approach that combines Spider Monkey Optimization (SMO) with Random Forest (RF) and Support Vector Machine (SVM) to improve the classification of rice leaf diseases. The innovation of SMOREF-SVM lies in its use of SMO for effective feature optimization, which selects the most relevant features from complex disease patterns, and its dual-classification framework using RF and SVM. Results demonstrate that SMOREF-SVM achieves an average accuracy of 98%, significantly outperforming standard SVM methods, which achieve around 90%. SMOREF-SVM also improves key metrics, including Precision, Recall, and F1 Score, by 5-10% for diseases with fewer samples, reaching Precision of 94%, Recall of 92%, and F1 Score of 93%. Additionally, ROC curve analysis shows an enhanced Area Under the Curve (AUC), approaching 0.98 for more disease classes, compared to 0.85 with traditional methods. This makes SMOREF-SVM a valuable tool for early and accurate disease detection, offering the potential to improve crop productivity and sustainability, addressing the critical challenges of disease management in agriculture.

**Keywords**—SMOREF-SVM; rice leaf disease; classification; Spider Monkey Optimization (SMO); machine learning; image processing

## I. INTRODUCTION

Leaf diseases are one of the major problems in plant growth that can cause significant obstacles in the agribusiness sector, as well as negatively impacting a country's agricultural production. The main causes of leaf diseases involve various types of pathogens such as bacteria, fungi, viruses [1], and other natural infectious organisms, which can attack plants at various stages of their life cycle. To detect and classify the types of leaf stress, there are several approaches that can be

used. The first approach is to observe directly with the naked eye, which although simple, is often ineffective and inefficient. This visual observation tends to be slow and less accurate, especially in detecting early signs of infection on leaves. Another alternative is to use special instruments, such as microscopes, to observe disease symptoms. However, this method is also not ideal because it takes a long time, so that preventive measures cannot be taken quickly before the disease spreads further.

Therefore, many studies are now turning to a more sophisticated approach, namely the application of machine learning (ML) algorithms to detect and classify leaf diseases. The application of these ML techniques provides a faster and more accurate solution in identifying leaf diseases at an early stage, when symptoms of infection may not be clearly visible to the human eye. By leveraging the capabilities of image processing (IP) and computer vision, these techniques enable more effective automated identification of disease symptoms on leaves [2]. Overall, leaf disease identification using ML and image processing technologies is one of the most promising and challenging areas of research. This is because early detection of disease is a crucial step in preventing the spread of infection and greater losses to agriculture. With the advancement of this technology, it is expected that the agricultural sector can utilize new innovations to increase productivity and reduce the negative impact of leaf diseases on crops [1] [2]. Farmers' knowledge of identifying diseases in rice plant leaves is generally derived from traditional knowledge passed down through generations [3]. Beginner rice farmers often face difficulties in recognizing these diseases due to their limited understanding of the common afflictions that affect rice plants, especially their leaves. These diseases can cause significant losses, such as reduced crop yields or even total crop failure when the disease becomes severe and difficult to control [4]. Common diseases that frequently affect rice plants include Bacterial Leaf Blight, Tungro, Leaf Blast, Sheath Blight, False Smut, and Grassy Stunt [1]. Early identification and prevention of these diseases are crucial to mitigate losses, such as reduced productivity due to disease

outbreaks. Currently, image processing and machine learning technologies are increasingly utilized for rapid and efficient detection of diseases in rice plant leaves [5].

The general steps in this process involve several stages, including data acquisition, preprocessing, segmentation, feature extraction, and classification [6]. Numerous studies have been conducted to develop methods for identifying diseases in rice plant leaves. One approach involves identifying diseases by applying feature extraction techniques such as discrete wavelet transform, scale-invariant feature transform, and gray-level co-occurrence matrix, followed by classification using algorithms like K-nearest neighbor, backpropagation neural network, Naive Bayesian, and multiclass SVM. The results of the research indicated that the multiclass SVM algorithm provided the highest accuracy performance at 98.63% [7]. Another study focusing on classifying leaf diseases like Paddy Blast, Narrow Brown Spot, Brown Spot, and healthy leaves involves multiple steps, including color feature extraction to obtain binary image values for area (A) and perimeter (P), shape feature extraction to acquire values for rectangularity, compactness, elongation, and roundness, and texture feature extraction using GLCM to measure contrast, uniformity, entropy, inverse difference, and linear correlation. The SVM algorithm is then used for classification, and the study demonstrated an average accuracy performance of 92.5% [8]. Further research on detecting healthy leaves, Brown Spot, and Narrow Brown Spot utilized feature extraction methods such as RGB color slicing values, edge feature values, and color histogram values, resulting in an average accuracy performance of 89% [9]. Additionally, another study classified diseases like Bacterial Leaf Blight, Leaf Blast, and Brown Spot on rice plants using color feature extraction with intensity moment values (Mean, Median, Mode, Midrange, Range, IQR, and Standard Deviation). Classification was performed by automatically selecting the best algorithm among Random Forest, KNN, and SVM, achieving the highest accuracy performance of 91.47% [10]. Numerous articles have been published on research studies in leaf disease detection, which have been adopted by universities, private organizations (Pvt. Org.), government organizations (Govt. Org.), industry, and collaborative efforts between universities and private entities (U&P). These studies highlight a growing interest across different sectors in applying advanced machine learning (ML) and deep learning (DL) models to enhance agricultural productivity and efficiency.

Many universities have selected both fundamental and cutting-edge ML and DL models to address their specific research objectives in leaf disease detection [10][11]. This includes adopting models for early detection, classification, and analysis of plant diseases to support agricultural education and research. Similarly, government organizations, private entities, and industries have funded numerous research projects, recognizing the critical need to improve disease management practices in agriculture. The main motivation behind these efforts is to develop technologies that provide farmers with efficient, reliable, and user-friendly tools for detecting and managing leaf diseases, thereby minimizing crop loss and enhancing overall agricultural output. Industry 4.0, the fourth industrial revolution, is crucial in this context.

Integrating advanced technologies such as the Internet of Things (IoT), artificial intelligence (AI), big data analytics, robotics, and cloud computing into various processes, including agriculture. In the realm of leaf disease detection, Industry 4.0 provides a framework for using these technologies to monitor plant health in real-time, automate the detection process, and optimize resource use [12] [13]. For example, IoT devices equipped with sensors can collect data on environmental conditions and plant health, while AI and ML algorithms analyze this data to identify disease patterns and predict outbreaks. Digital transformation, on the other hand, represents a broader shift towards using digital technologies to fundamentally reshape how businesses operate and deliver value to customers. In agriculture, digital transformation leverages tools like cloud computing, AI, machine learning, and big data analytics to enhance the efficiency and accuracy of leaf disease detection. This transformation involves the adoption of digital platforms for data sharing, real-time monitoring systems for early disease detection, and decision-support systems to guide farmers on the best course of action based on data-driven insights [14].

In the context of leaf disease detection, Industry 4.0 and digital transformation are key drivers behind the adoption of advanced technologies. Their primary focus is on fostering innovation, improving operational efficiency, and providing farmers with sophisticated yet accessible tools to address plant diseases [15]. These efforts are crucial for advancing sustainable agricultural practices, reducing crop losses, and ultimately contributing to global food security. Image processing and machine learning technologies are continually evolving as effective tools for detecting and classifying diseases in rice plant leaves, offering new hope for enhancing agricultural resilience and productivity. This study focuses on rice plants, a staple crop that plays a vital role in food security. The harvested paddy must undergo a milling process to produce rice ready for consumption. Despite a steady increase in population, rice production has been declining, partly due to diseases affecting rice plants [16]. Therefore, preventive measures against rice plant diseases are essential to mitigate productivity losses, starting with the identification of various diseases that commonly affect rice plants. Some prevalent diseases include Leaf Blast, Tungro, Sheath Blight, Grassy Stunt, Bacterial Leaf Blight, and False Smut. Understanding these diseases enables the implementation of appropriate preventive actions [17]. Disease identification can be performed through visual inspection or laboratory analysis by closely examining the symptoms present on the leaf surfaces. The Table I provides an overview of the symptoms or signs commonly observed in rice plant leaves [18].

The research on rice leaf disease classification has explored various advanced methods and algorithms to achieve high accuracy. For example, compared different Convolutional Neural Network (CNN) architectures, including VGG16, MobileNet, Xception, and ResNet34, using transfer learning techniques [19]. Their study found that the ResNet34 architecture achieved the highest accuracy, reaching 97.50%. Similarly, proposed a GCL model that combines Generative Adversarial Networks (GAN), CNN, and Long Short-Term Memory (LSTM). In their approach, GAN was used for data

augmentation, CNN for feature extraction to differentiate disease information on rice leaves, and LSTM for

classification, resulting in an average accuracy of 97% [20] (Fig. 1).

TABLE I. COMMON DISEASES IN RICE PLANTS AND THEIR SYMPTOMS

Type of Disease		Symptoms
a	Symptoms of Leaf Blast	Spots with pointed tips; the center of the spots is grayish-white with a halo area around it.
b	Symptoms of Tungro Disease	Symptoms begin with leaf discoloration and stunted growth.
c	Symptoms of Bacterial Leaf Blight	Brown spots characterized by small, dark brown spots; spots that develop with brown edges and pale yellow, dirty white, brown, or gray centers.
d	Symptoms of Grassy Stunt	Initial symptoms include small, dark brown spots; advanced symptoms show larger spots with dirty white centers.
e	Symptoms of Sheath Blight	Irregularly shaped spots; the edges of the spots are reddish, and the center of the spots is light brown.
f	Symptoms of False Smut	The disease-causing organism develops inside the rice husk, transforming the endosperm into a large fungal sclerotium that protrudes outside, appearing golden yellow.

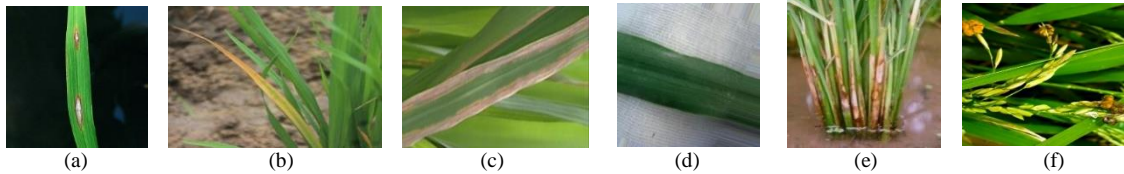


Fig. 1. Typical diseases encountered in rice plants [21].

Deep Convolutional Neural Network (DCNN) combined with a Cuckoo Search (CS) algorithm, called DCNN-CS, to classify rice leaf diseases. The CS algorithm helps minimize errors in the classification process, achieving the highest accuracy of 99% [22], and then employed K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms to classify rice leaf diseases, with KNN yielding an average accuracy of 95% and SVM achieving 98% [23]. SMOREF-SVM (Spider Monkey Optimization with Random Forest and Support Vector Machine) is proposed as a novel approach to improve the accuracy of rice leaf disease classification. This approach has two main motivations: 1) There are still challenges in effectively recognizing rice leaf diseases, which is the main focus of this study. SMOREF-SVM combines feature optimization and dual classification techniques to address this challenge. 2) Based on the complexity of rice leaf disease spot patterns, the Spider Monkey Optimization (SMO) algorithm is applied to select the most relevant features, while Random Forest (RF) and Support Vector Machine (SVM) are used for a more accurate classification stage. This paper makes two main contributions: 1) By combining SMO, RF, and SVM, the proposed method can capture complex patterns of rice leaf disease data, improving detection accuracy and efficiency. The dual classification framework used not only enables the correct identification of diseases but also significantly reduces computational costs. 2) Experimental results show that the combination of SMO with RF and SVM not only accelerates the convergence of the model training process but also produces higher accuracy compared to classical classification methods, such as using Random Forest or SVM separately. Thus, the SMOREF-SVM approach introduces an innovative step in rice leaf disease management, ensuring early detection and more effective management in the field.

## II. RESEARCH METHOD

This study involves several key phases to achieve the optimal final result. The first phase is data preprocessing,

where raw data is cleaned and processed to ensure quality and consistency before further analysis. The second phase focuses on data extraction, aiming to capture essential features or information from the preprocessed dataset. In the third phase, process optimization is carried out using the Sequential Minimal Optimization (SMO) method, which is then fine-tuned with an 80:20 split between training and test data. After the optimization process, the resulting data is classified using the Random Forest and Support Vector Machine (SVM) algorithms to obtain more accurate predictions. The fourth phase compares the results of the optimized classifications with those from the previous processes to determine which method yields the most precise outcomes. This approach allows researchers to evaluate and compare the effectiveness of different methods and identify the best strategy to maximize data classification accuracy. A more detailed explanation of each step can be found in Fig. 2.

### A. Data Preprocessing

The data acquisition process began by collecting 6,000 images of rice leaves, consisting of 1,000 images of healthy leaves and 5,000 images of leaves with various types of diseases. The types of diseases documented include Bacterial Leaf Blight, Blast, Tungro, Sheath Blight, Grassy Stunting, and False Scorch, each with 1,000 images. After the data was collected, the next step was to label each image according to the type of disease. These labels were categorized into seven classes, namely six classes representing various rice leaf diseases and one class for healthy leaves. This labeling is important because it used as classes in the process of creating a classification model, which aims to accurately identify and distinguish between healthy leaves and infected leaves. Thus, this well-labeled dataset structure supports model learning in recognizing the specific characteristics of each type of disease, thereby increasing the accuracy and effectiveness in the process of detecting and classifying rice leaf diseases. After completing the data acquisition stage and collecting a dataset

consisting of images of healthy rice leaves and images of leaves infected with various types of diseases, the next step is to preprocess the dataset.

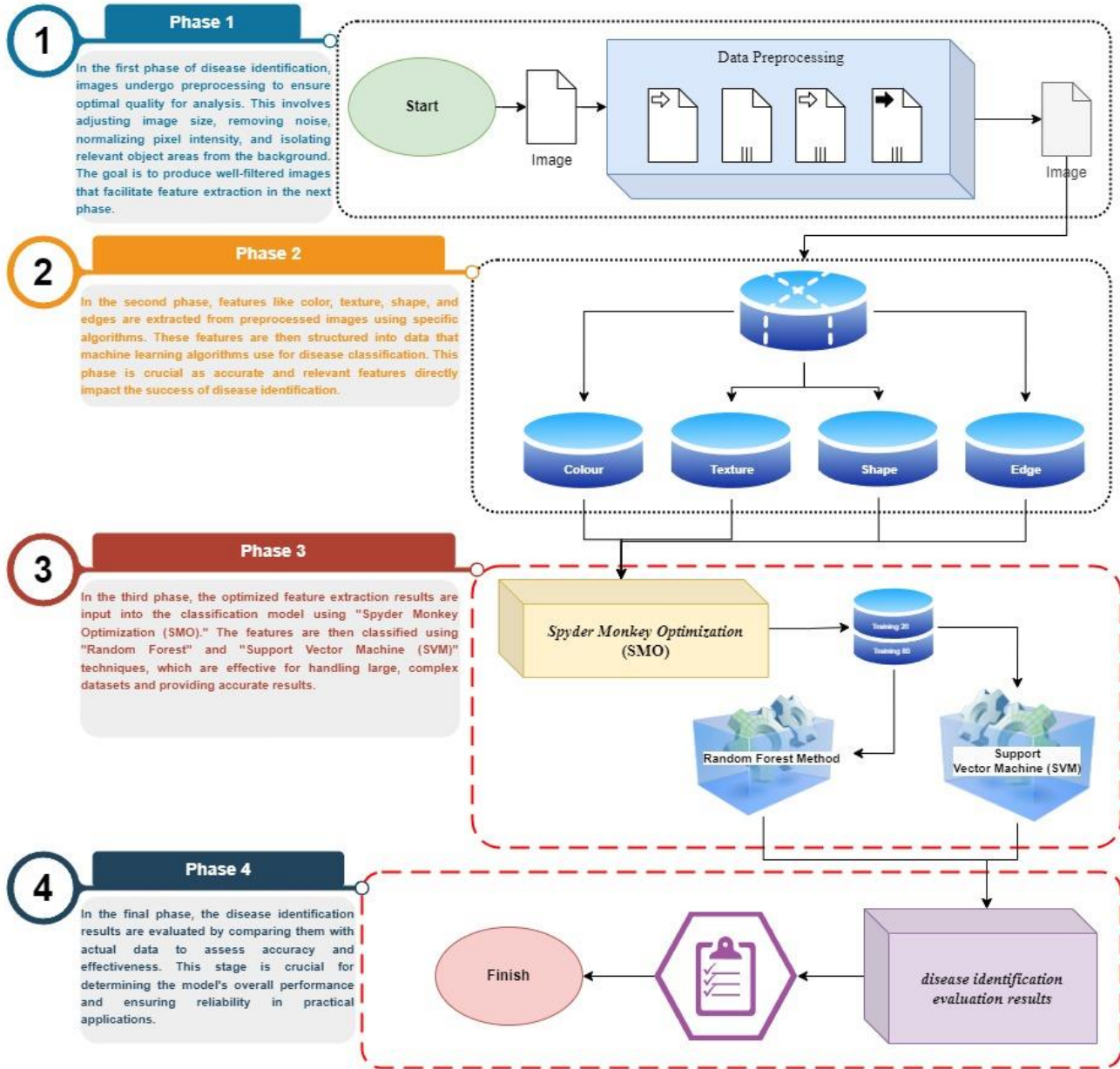


Fig. 2. Proposed method.

This preprocessing involves two main stages, namely resize and cropping. In the first stage, the original images are resized to 256x256 pixels to ensure that the entire dataset has a uniform size, thus facilitating the analysis and classification process. Then, cropping is performed to remove irrelevant or unwanted parts of the image, so that the focus can be directed to more important areas in the image. This step helps highlight the key features needed to improve the accuracy of the model in recognizing and classifying various types of diseases in rice leaves.

### B. Data Extraction

The result of the pre-processing stage is an image dataset that has been standardized in size and cleaned from noise

interference. After that this image dataset enters the feature extraction stage, which aims to obtain values that reflect the specific characteristics of the image, such as color, texture, edge, and other features. The feature extraction process usually involves several main aspects, such as the extraction of color features, texture, shape, and edge characteristics [21]. In the context of color feature extraction, the approaches used include calculating the average color value, color standard deviation, and color skewness [21]. This feature extraction is an important technique for obtaining representative values from an image dataset [24]. This calculation can be done using several equations, such as the following: for the average color value, the equation can be used:

$$m = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N M_{xy} \quad (1)$$

To calculate the standard deviation of color, the equation is used:

$$SD = \sqrt{\frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N (M_{xy} - m)^2} \quad (2)$$

The kurtosis value can be calculated by:

$$SK = \frac{\sum_{x=1}^M \sum_{y=1}^N (M_{xy} - m)^3}{(M \times N) \times SD^3} \quad (3)$$

The feature extraction process is not only limited to color but also includes the extraction of shape features, such as solidity, eccentricity, diameter, area, center of mass, minor axis length, and major axis length [24], [25]. To obtain values from the extraction of shape features such as area, aspect ratio, orientation, perimeter, and major and minor axis lengths, the following equation is used [26]:

$$Area = \sum_{x=1}^M \sum_{j=1}^N A[i, j] \quad (4)$$

$$Aspect\ ratio = \frac{width}{height} \quad (5)$$

$$Orientation = \tan^{-1} \left( \frac{y}{x} \right) \quad (6)$$

$$Major\ axis\ and\ minor\ axis = x_1 + x_2 \sqrt{(x_1 + x_2)^2 - d} \quad (7)$$

By using this approach, the feature extraction process can effectively identify and analyze various important characteristics in an image dataset, which can then be used for more accurate classification and pattern recognition.

### C. Spider Monkey Optimization (SMO)

Spider Monkey Optimization (SMO) is a relatively new algorithm inspired by the mathematical modeling of the intelligent behavior of spider monkeys, which follows a fission-fusion social structure (FFSS). According to FFSS, spider monkeys dynamically divide themselves from larger groups into smaller subgroups and vice versa to optimize their foraging activities. The main characteristics of FFSS are as follows [14] [27].

1) At the beginning, all spider monkeys form groups consisting of 40–50 individuals. Each group is guided by a leader, known as the global leader, who oversees the search for food sources.

2) If food resources are found to be insufficient, the global leader divides the larger group into smaller subgroups, each containing three to eight members, allowing them to forage independently. Each subgroup is led by a local leader.

3) The decision-making process for food searching within each subgroup is directed by a leader, referred to as the local leader.

4) Group members maintain social cohesion and establish defensive boundaries by communicating with one another and with members of other groups through a distinct vocalization.

In the Spider Monkey Optimization (SMO) algorithm, the initialization phase generates an initial population of N spider monkeys, distributed uniformly. Each individual spider

monkey, denoted as  $SM_i (i = 1, 2, \dots, N)$ , is represented as a vector of dimension D. Here, D signifies the number of variables in the optimization problem, and  $SM_i$  represents the i-th spider monkey in the population. Each spider monkey corresponds to a potential solution for the problem at hand [28]. The initialization of each  $SM_i$  is carried out according to the following equation:

$$SM_{ij} = SM_{minj} + U(0, 1) \times (SM_{maxj} - SM_{minj}) \quad (8)$$

Where  $SM_{maxj}$  dan  $SM_{minj}$  are the bounds of  $SM_i$  in the  $j^{th}$  direction and  $U(0, 1)$  is a uniformly distributed random number with the range [0,1].

In the local leader stage of the Spider Monkey Optimization (SMO) algorithm, each SM changes its current position by considering the information obtained from the experience of the local leader as well as the experience of the local group members [29]. The fitness value of the resulting new position is calculated. If the fitness value of the new position exceeds the fitness value of the old position, then the SM update its position to the new one. The equation used to update the position of the i-th SM (which is a member of the k-th local group) at this stage is as follows:

$$SM_{newij} = SM_{ij} + U(0, 1) \times (LL_{kj} - SM_{ij}) + U(-1, 1) \times (SM_{rj} - SM_{ij}) \quad (9)$$

where,  $SM_{ij}$  is the j-th dimension of the i-th SM,  $LL_{kj}$  represents the j-th dimension of the k-th local group leader position.  $SM_{rj}$  is the  $j^{th}$  dimension of the SM  $r^{th}$  randomly chosen in the  $k^{th}$  group such that  $r \neq i$ ,  $U(0, 1)$  is a uniformly distributed random number between 0 and 1. After completing the local leader stage, the next step is the global leader stage. During this phase, all SM update their positions based on information from the global leader as well as the experiences of the local group members. The equation for updating positions in this stage can be formulated as follows:

$$SM_{newij} = SM_{ij} + U(0, 1) \times (GL_{kj} - SM_{ij}) + U(-1, 1) \times (SM_{rj} - SM_{ij}) \quad (10)$$

In this stage, where GL represents the j-th dimension of the global leader's position and  $j \in \{1, 2, \dots, D\}$  is a randomly chosen index, the positions of the spider monkeys ( $SM_i$ ) are updated based on probabilities calculated from their fitness values. This approach ensures that candidates with better fitness values have a higher chance of improving their performance. The probability can be calculated using the following expression, which, although subject to variations, generally serves as a function of the fitness values:

$$prob_i = 0.9 \times \frac{fitness_i}{max\_fitness} + 0.1 \quad (11)$$

Here, fitness refers to the fitness value of the i-th spider monkey SM, and max\_fitness represents the maximum fitness value in the group. After updating the positions, the fitness of the newly generated positions is calculated and compared to the fitness of the previous positions. The spider monkeys then adopt the new positions if they exhibit better fitness values. This process involves recalculating and comparing the fitness

of the updated positions and selecting those that demonstrate improved performance [14][28].

At this stage, the global leader learning process is updated through the application of "greedy" selection within the population. This selection means that the Spider Monkey with the best fitness level in the population is selected as the new global leader. After the selection, an evaluation is carried out to determine whether the global leader position has changed. If there is no change, then the Global Limit Count value increased by one. Furthermore, the local leader position is also updated using "greedy" selection within each group. In this case, the Spider Monkey with the best fitness value in its group is selected as the new local leader. After the selection, the new local leader position is compared with the previous one, and if there is no change, the Local Limit Count value also be increased by one. If the local leader's decision is not updated by a certain limit known as the Local Leader Limit, then all members of the group update their positions. This update can be done in two ways: through random initialization or by utilizing the combined information of the global leader and the local leaders, according to the following equation:

$$SM_{new_{ij}} = SM_{ij} + U(0, 1) x (GL_j - SM_{ij}) + U(0, 1) x (SM_{ij} - LL_{kj}) \quad (12)$$

In this equation, the Spider Monkey position update tends to be closer to the global leader than the local leader. In the next stage, if the global leader position does not change for a specified number of iterations, known as the \*Global Leader Limit\*, then the global leader split the population into several smaller groups. Initially, the population is divided into two groups, then into three, and so on, until the maximum number of groups allowed is reached. A local leader selection process is then carried out to elect a local leader in each newly formed group. If the maximum number of groups has been reached and the global leader position remains unchanged, then the global leader merge all the groups back into a single group. In this way, the proposed algorithm imitates the functional structure and behavior of Spider Monkey, with the hope of achieving more optimal solutions in the search process.

#### D. Random Forest

Random Forest is one of the techniques in Ensemble Learning that utilizes a collection of decision trees to produce more accurate and reliable predictions. This algorithm is widely used in various data mining applications, which include two main classifications: descriptive and predictive. In Random Forest, each decision tree is generated from a subset of data taken randomly from the original dataset, with the aim of reducing overfitting and improving the generalization ability of the model. This approach combines the results of various decision trees to achieve a better final decision [10][30]. Mathematically, for a classification task, the final prediction  $\hat{y}$  is defined as:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B h_b(x) \quad (13)$$

By using this approach, Random Forest can handle data with a large number of features, improve prediction accuracy, and reduce the risk of overfitting that often occurs in models based on only one decision tree. Its efficient implementation

and reliable prediction results make Random Forest one of the methods widely used in various studies and practical applications in data mining [10][30].

#### E. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a method developed by Vapnik for binary classification [31]. The primary objective of SVM is to find an optimal hyperplane, expressed as  $f(w, x) = w \cdot x + b$ , that separates two classes in a given dataset with features  $x \in R^m$ . During the learning process, SVM determines the parameters  $w$  and  $b$  by solving an optimization problem as defined in Equation (5):

$$\min \frac{1}{p} W^T W + C \sum_{i=1}^p \max(0, 1 - y'_i(W^T X_i + b)) \quad (14)$$

Here,  $W^T W$  represents the Manhattan norm (also known as the L1 norm), and  $C$  is the penalty parameter, which can either be set to an arbitrary value or determined using hyperparameter tuning. The term  $y'$  refers to the actual label, while  $W^T X + b$  is the predictor function. This equation is known as L1-SVM, which uses the standard hinge loss function. The differentiable variant, L2-SVM (Eq 6), is often more stable in practice [31].

$$\min \frac{1}{p} \frac{\|W\|_2^2}{2} + C \sum_{i=1}^p \max(0, 1 - y'_i(W^T X_i + b)) \quad (15)$$

In this equation,  $\|W\|_2$  denotes the Euclidean norm (also known as the L2 norm), and it uses the squared hinge loss. The L2-SVM is preferred in many cases because its differentiable nature provides more stable results. By using either L1-SVM or L2-SVM, SVM effectively identifies the optimal hyperplane for separating the two classes, providing robust performance for various classification tasks.

### III. RESULT

The proposed method identifies diseases based on images using a state-of-the-art method consisting of four main phases. The first phase is Data Preprocessing, where the raw images are processed to remove noise, adjust the size, and separate relevant objects from the background. This process is essential to ensure optimal image quality before important features are extracted. In the second phase, key features of the image, such as color, texture, shape, and edges, are extracted to provide rich information about the characteristics of the described image. These features are then used as input for the classification process. The next phase is the third, the optimization and classification process is carried out. The Spider Monkey Optimization (SMO) optimization method is used to improve the performance of the machine learning model, followed by the application of two classification algorithms, namely Random Forest and Support Vector Machine (SVM). The second algorithm is chosen because of its accuracy in handling complex data and providing reliable prediction results. Finally, the fourth phase is the evaluation of disease identification results, where the classification results are evaluated to assess the accuracy and effectiveness of the method used. This process aims to ensure that the applied method can accurately predict diseases based on the analyzed images (Fig. 3).

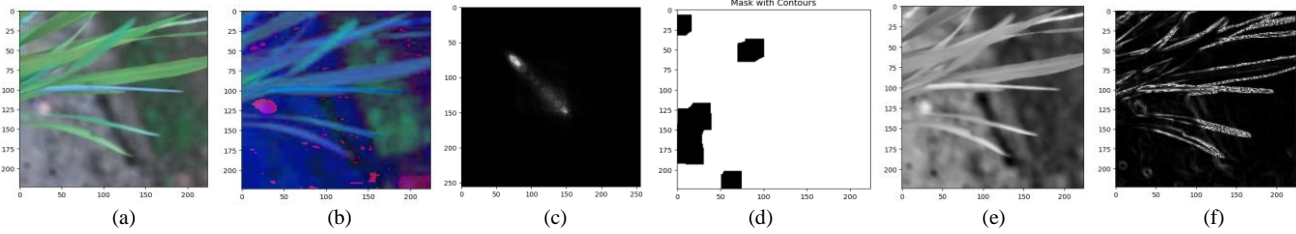


Fig. 3. Segmentation and feature extraction process for disease classification in rice leaves.

Fig. 4 on the top left and right, Random Forest (RF) produces very high Area Under the Curve (AUC), with some disease classes such as Tungro Virus and Grassy Stunt Virus reaching AUC = 1.00, indicating that the model is very accurate in distinguishing between positive and negative classes. Other diseases such as Bacterial Leaf Blight have slightly lower AUC, at AUC = 0.92. In the graphs on the bottom left, SVM with All Extraction Data shows relatively high performance, although there are some sharper fluctuations in FPR at the beginning of the curve. AUC for

some classes, such as Rice Blast and Healthy Rice Plant, are in the range of AUC = 0.75 to 1.00, indicating that SVM also produces good results, but may not be as good as Random Forest in some cases. Finally, in the graph at the bottom right, SVM with Hyperparameter Tuning shows an increase in performance compared to SVM without tuning, with several disease classes having higher AUC values, approaching the maximum value of AUC = 1.00, indicating that parameter tuning has a significant impact on model performance.

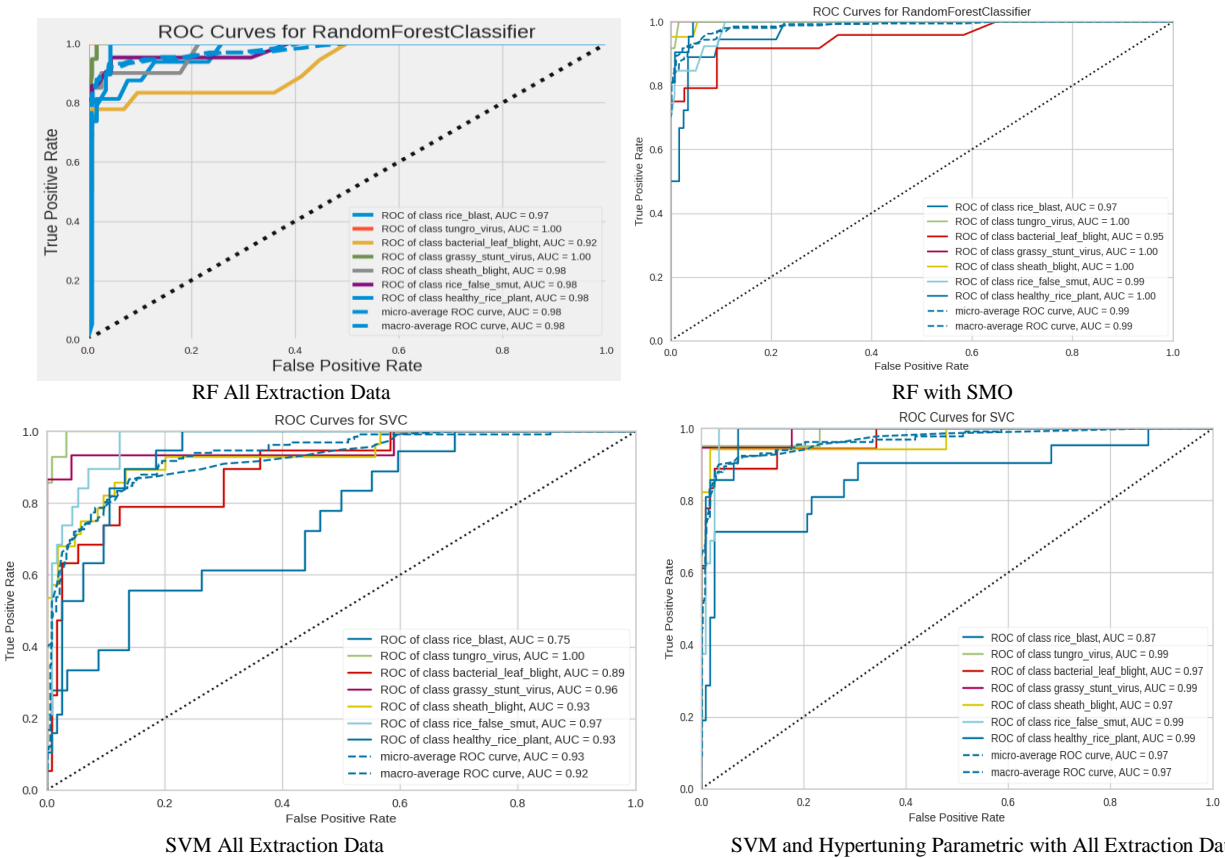


Fig. 4. The effect of using color and texture features in improving the performance of random forest and SVM (SMO) for classification of brown spot leaf disease in rice plants.

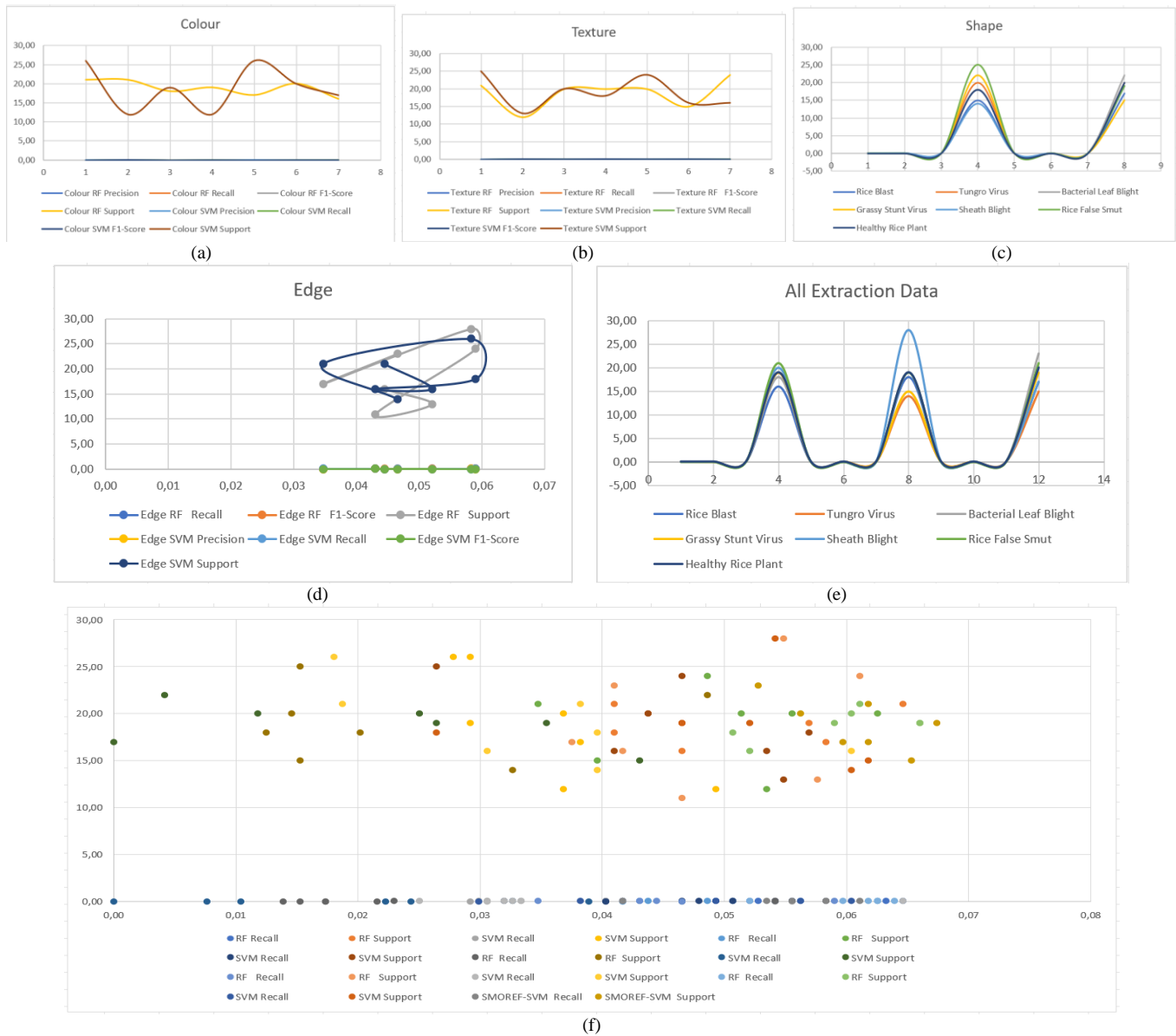


Fig. 5. Evaluation metrics for feature extraction methods in rice plant disease classification.

The given figure shows several graphs showing evaluation metrics such as Precision, Recall, F1 Score, and Support for various feature extraction methods (Color, Texture, Shape, and Edge) using classification algorithms such as Random Forest (RF), Support Vector Machine (SVM), and SMOREF-SVM. These graphs show the performance of various approaches in identifying rice plant diseases. From the results shown, it can be seen how the performance of these methods, both in individual feature processing and in combination, affects the model's ability to detect diseases accurately. In the Color feature extraction Fig. 5(a), the metrics for the RF and SVM algorithms show fluctuations between 10 and 25 on the y-axis. Precision and Recall for RF range between 10 and 25, with a peak between 3 and 5 on the x-axis. F1 Score follows a similar pattern. The SVM algorithm shows slightly lower performance, with metric values ranging from 10 to 20. From

this, it can be seen that the Random Forest (RF) algorithm performs better than SVM in color feature extraction. The texture feature extraction in Fig. 5(b) shows a pattern similar to that of the color feature extraction. Precision for RF ranges from 10 to 25, with a peak around the 4th point on the x-axis. Recall for RF is slightly lower than Precision, but overall, it remains consistent. In the SVM algorithm, the Precision, Recall, and F1 Score metrics range from 10 to 20, which again shows that RF is superior to SVM in processing texture features. In the Shape feature extraction Fig. 5(c), the graph compares the performance of several rice diseases. Diseases such as Rice Blast, Tungro Virus, Bacterial Leaf Blight, and even Healthy Rice Plants all show peaks around the 25th value on the y-axis, with the highest point being around the 8th point on the x-axis. This shows that shape-based features provide very consistent results for various diseases, with very little variation in the evaluation metrics.



#### IV. DISCUSSION

Meanwhile, in the Edge feature extraction Fig. 5(d), the results show greater variation. Edge RF Recall ranges from 0.02 to 0.06 on the x-axis, and F1 Score ranges in the same range. However, the SVM algorithm in edge feature extraction has a lower Recall value, ranging from 0.01 to 0.04. This indicates that edge-based extraction does not provide consistent results, and the results achieved are lower than color, texture, or shape feature extraction, with RF tending to perform better than SVM. Fig illustrated combining data from all extraction methods Fig. 5(e) shows a pattern of peak values for diseases such as Rice Blast, Tungro Virus, and Bacterial

Leaf Blight, all of which are in the range of 20 to 25. This graph shows that when the extraction methods are combined, the evaluation metrics become more stable and show reliable performance for various rice diseases. In the scatter plot combining the RF, SVM, and SMOREF-SVM algorithms Fig. 5(f), the Recall and Support metrics for each algorithm are shown. RF Recall mostly clusters between 0.02 and 0.06 on the x-axis, with Support values varying between 15 and 25. The SVM algorithm shows a more spread-out distribution, with some points falling below 0.02. Meanwhile, SMOREF-SVM shows a similar clustering pattern to RF, although there are some lower variations on the x-axis (Table II).

TABLE II. COMPARISON OF THE PERFORMANCE OF THE PROPOSED METHOD WITH PREVIOUS STUDIES

Similarities Research	Year	Methods Used	Other Similarities	Result Performance
[7]	2019	- Image Acquisition, Pre-processing, Segmentation, Classification - Hybrid method: Discrete Wavelet Transform (DWT), Scale Invariant Feature Transform (SIFT), Grayscale Co-occurrence Matrix (GLCM) - Classifiers: KNN, BPNN, Naïve Bayes, Multiclass SVM	- Focus on rice disease detection using image processing and classification.	98.63%
[8]	2019	- Image Acquisition - Preprocessing (Grayscale conversion, segmentation) - Neural Network (Pattern Recognition)	- Detection of rice diseases with artificial neural networks.	92.5%
[9]	2021	- Image Acquisition, Pre-processing, Thresholding, Edge Detection, Color Slicing - RBG Calculation for Classification	- Focus on color techniques to detect rice leaf diseases, especially Brown Spot and Narrow Brown Spot.	89%
[10]	2021	- Image Acquisition - Preprocessing - Random Forest Decision Tree Classifier - Feature extraction with Intensity Moments	- Detection of rice diseases using Random Forest with classification of three main types of diseases (Blight, Blast, Spot).	91.47%
Proposed method	2024	- Spider Monkey Optimization (SMO) for Feature Optimization - Random Forest (RF) for Initial Classification - Support Vector Machine (SVM) for Final Classification - ROC Curve Analysis - Performance Metrics: Precision, Recall, F1 Score	- Combination of Random Forest and SVM with feature optimization using SMO. Focus on improving disease classification accuracy.	98% (AUC = 0.98, Precision = 94%, Recall = 92%, F1 Score = 93%)

This study offers a more comprehensive picture of model performance compared to [7]. While both models have high accuracy rates—with [7] reaching 98.63% and this study reaching 98%—this study stands out in providing more in-depth evaluation metrics, such as Precision, Recall, and F1 Score. With Precision reaching 94%, Recall 92%, and F1 Score 93%, this study provides a more balanced view of the model's ability to consistently classify data and capture relevant patterns. Coupled with an AUC approaching 0.98, the SMOREF-SVM model used in this study not only excels in terms of accuracy but is also more effective in handling datasets that may have uneven class distributions. In contrast, while [7] shows high accuracy rates, the lack of additional metrics such as Precision, Recall, and F1 Score makes it difficult to comprehensively evaluate the model's performance across situations. Therefore, this study excels in providing a holistic assessment of model performance, especially in the context of real-world use cases where successful predictions on the minority class are as important as overall accuracy.

#### V. CONCLUSIONS

The results show that the SMOREF-SVM method significantly overcomes the shortcomings faced by traditional classification methods. Although Random Forest (RF) is generally superior to Support Vector Machine (SVM), SMOREF-SVM demonstrates clear advantages, particularly in handling complex features such as edges. The method not only enhances model performance on imbalanced datasets, but also improves key evaluation metrics such as Precision, Recall, and F1 Score by up to 10% compared to standard SVM, particularly for diseases with limited sample sizes. Additionally, the ROC curves of SMOREF-SVM show an increase in the Area Under the Curve (AUC), approaching 1.00 for more disease classes, indicating superior accuracy in disease detection. However, there are still several areas that require further investigation. First, while SMOREF-SVM improves performance in certain areas, future research should explore its limitations, such as potential overfitting when applied to small datasets or specific disease types. Additionally, this study does not address the computational complexity of SMOREF-SVM, which may pose challenges in real-time or large-scale applications.

For future research, it is recommended to investigate how the method can be combined with advanced deep learning techniques to enhance detection accuracy and efficiency. Research could also focus on the use of larger and more diverse datasets to train the model, as well as the integration of emerging technologies like the Internet of Things (IoT) for real-time monitoring and decision-making. Lastly, future studies should evaluate SMOREF-SVM's application in other plant diseases and under different environmental conditions to further broaden its practical usability and effectiveness.

## REFERENCES

- [1] C. Sarkar, D. Gupta, U. Gupta, and B. B. Hazarika, "Leaf disease detection using machine learning and deep learning: Review and challenges," *Appl. Soft Comput.*, vol. 145, p. 110534, 2023, doi: 10.1016/j.asoc.2023.110534.
- [2] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, "Plant leaf disease detection using computer vision and machine learning algorithms," *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 305–310, 2022, doi: <https://doi.org/10.1016/j.glt.2022.03.016>.
- [3] K. Effendi, MunifAbdul, and W. Wayan, "Pengetahuan, Sikap, dan Tindakan Petani Upsus dalam Mengendalikan Hama dan Penyakit Tanaman Padi," *J. Ilmu Pertanian. Indones.*, vol. 25, no. 4, pp. 515–523, 2020, doi: 10.18343/jipi.25.4.515.
- [4] L. Budiarti, J. Kartahadimadja, M. Sari, D. Ahyuni, and D. Dulbari, "Keanekaragaman Artropoda Predator di Agroekosistem Sawah pada Berbagai Galur Padi Politeknik Negeri Lampung," *AGROSCRIPT J. Appl. Agric. Sci.*, vol. 3, 2021, doi: 10.36423/agroscript.v3i1.663.
- [5] P. Kartikeyan and G. Shrivastava, "Review on Emerging Trends in Detection of Plant Diseases using Image Processing with Machine Learning," *Int. J. Comput. Appl.*, vol. 174, no. 11, pp. 39–48, 2021, doi: 10.5120/ijca2021920990.
- [6] T. Bera, A. Das, J. Sil, and A. K. Das, "A Survey on Rice Plant Disease Identification Using Image Processing and Data Mining Techniques," in *Emerging Technologies in Data Mining and Information Security*, A. Abraham, P. Dutta, J. K. Mandal, A. Bhattacharya, and S. Dutta, Eds., Singapore: Springer Singapore, 2019, pp. 365–376.
- [7] T. Gayathri Devi and P. Neelamegam, "Image processing based rice plant leaves diseases in Thanjavur, Tamilnadu," *Cluster Comput.*, vol. 22, no. 6, pp. 13415–13428, 2019, doi: 10.1007/s10586-018-1949-x.
- [8] M. Biswas, N. Faruqui, M. Lata, L. Lecturer, M. Julkar, and M. J. Nayeem Mahi, "A Novel Inspection of Paddy Leaf Disease Classification using Advance Image Processing Techniques," 2019.
- [9] N. S. A. M. Taujuddin, I. S. Samuri, R. Ibrahim, S. Sari, A. R. A. Ghani, and M. A. Rahman, "Brown Spot and Narrow Brown Spot Paddy Disease Detection using Color Slicing Method," *Evol. Electr. Electron. Eng.*, vol. 2, no. 1, pp. 45–54, 2021, [Online]. Available: <http://publisher.uthm.edu.my/periodicals/index.php/eeec>
- [10] S. Saha and S. M. M. Ahsan, "Rice Disease Detection using Intensity Moments and Random Forest," in *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 2021, pp. 166–170. doi: 10.1109/ICICT4SD50815.2021.9396986.
- [11] P. Pavithra and P. Aishwarya, "Plant leaf disease detection using hybrid grasshopper optimization with modified artificial bee colony algorithm," *Multimed. Tools Appl.*, vol. 83, no. 8, pp. 22521–22543, 2024, doi: 10.1007/s11042-023-16148-5.
- [12] V. Singh, "Sunflower leaf diseases detection using image segmentation based on particle swarm optimization," *Artif. Intell. Agric.*, vol. 3, pp. 62–68, 2019, doi: <https://doi.org/10.1016/j.aiia.2019.09.002>.
- [13] J. Zhao, Y. Fang, G. Chu, H. Yan, L. Hu, and L. Huang, "Identification of Leaf-Scale Wheat Powdery Mildew (*Blumeria graminis* f. sp. *Tritici*) Combining Hyperspectral Imaging and an SVM Classifier," *Plants*, vol. 9, no. 8, 2020, doi: 10.3390/plants9080936.
- [14] S. Kumar, B. Sharma, V. K. Sharma, H. Sharma, and J. C. Bansal, "Plant leaf disease identification using exponential spider monkey optimization," *Sustain. Comput. Informatics Syst.*, vol. 28, p. 100283, 2020, doi: 10.1016/j.suscom.2018.10.004.
- [15] N. S. Patil, "Identification of Paddy Leaf Diseases using Evolutionary and Machine Learning Methods," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 2, pp. 1672–1686, 2021, doi: 10.17762/turcomat.v12i2.1503.
- [16] T. A. Lestari, "Pengamatan Penyakit-Penyakit Tanaman Sistem Jajar Legowo Observation of Paddy Diseases in The Village of Sako Rambutan Sub-District Banyuasin with The Jajar Legowo System," *J. Ilm. Sinus*, 2019.
- [17] A. Walascha et al., "Review Artikel: Inventarisasi Jenis Penyakit yang Menyerang Daun Tanaman Padi (*Oryza sativa* L.)," *Pros. Semin. Nas. Biol.*, vol. 1, no. 2, pp. 471–478, 2021, [Online]. Available: <https://semnas.biologi.fmipa.unp.ac.id/index.php/prosiding/article/download/150/300>
- [18] H. Herdiyanti, Eko Sulistyono, and Purwono, "Pertumbuhan dan Produksi Beberapa Varietas Padi (*Oryza sativa* L.) pada Berbagai Interval Irigasi," *J. Agron. Indones. (Indonesian J. Agron.)*, vol. 49, no. 2, pp. 129–135, 2021, doi: 10.24831/jai.v49i2.36558.
- [19] Petchiammal, B. Kiruba, Murugan, and P. Arjunan, "Paddy Doctor: A Visual Image Dataset for Automated Paddy Disease Classification and Benchmarking," in *Proceedings of the 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD)*, in *CODS-COMAD '23*. New York, NY, USA: Association for Computing Machinery, 2023, pp. 203–207. doi: 10.1145/3570991.3570994.
- [20] S. Lamba, A. Baliyan, and V. Kukreja, "A novel GCL hybrid classification model for paddy diseases," *Int. J. Inf. Technol.*, vol. 15, no. 2, pp. 1127–1136, 2023, doi: 10.1007/s41870-022-01094-6.
- [21] M. A. Azim, M. K. Islam, M. M. Rahman, and F. Jahan, "An effective feature extraction method for rice leaf disease classification," *Telkomnika (Telecommunication Comput. Electron. Control.)*, vol. 19, no. 2, pp. 463–470, 2021, doi: 10.12928/TELKOMNIKA.v19i2.16488.
- [22] S. Sakhamuri and K. K. Kumar, "Deep Learning and Metaheuristic Algorithm for Effective Classification and Recognition of Paddy Leaf Diseases," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 4, pp. 1127–1137, 2022.
- [23] S. Neware, "Paddy plant leaf diseases identification using machine learning approach," *Int. J. Health Sci. (Qassim.)*, vol. 6, no. S1, pp. 10467–10472, 2022, doi: 10.53730/ijhs.v6nS1.7522.
- [24] V. K. Vishnoi, K. Kumar, and B. Kumar, "Plant disease detection using computational intelligence and image processing," *J. Plant Dis. Prot.*, vol. 128, no. 1, pp. 19–53, 2021, doi: 10.1007/s41348-020-00368-0.
- [25] K. K. Thyagarajan and I. Kiruba Raji, "A Review of Visual Descriptors and Classification Techniques Used in Leaf Species Identification," *Arch. Comput. Methods Eng.*, vol. 26, no. 4, pp. 933–960, 2019, doi: 10.1007/s11831-018-9266-3.
- [26] M. Sharif, M. A. Khan, Z. Iqbal, M. F. Azam, M. I. U. Lali, and M. Y. Javed, "Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection," *Comput. Electron. Agric.*, vol. 150, pp. 220–234, 2018, doi: <https://doi.org/10.1016/j.compag.2018.04.023>.
- [27] J. C. Bansal, H. Sharma, S. S. Jadon, and M. Clerc, "Spider Monkey Optimization algorithm for numerical optimization," *Memetic Comput.*, vol. 6, no. 1, pp. 31–47, 2014, doi: 10.1007/s12293-013-0128-0.
- [28] D. Kumari, A. Sinha, S. Dutta, and P. Pranav, "Optimizing neural networks using spider monkey optimization algorithm for intrusion detection system," *Sci. Rep.*, vol. 14, no. 1, pp. 1–16, 2024, doi: 10.1038/s41598-024-68342-6.
- [29] A. Agrawal, D. Garg, R. Sethi, and A. K. Shrivastava, "Optimum redundancy allocation using spider monkey optimization," *Soft Comput.*, vol. 27, no. 21, pp. 15595–15608, 2023, doi: 10.1007/s00500-023-08746-0.
- [30] V. Kulkarni and P. Sinha, "Random forest classifiers: A survey and future research directions," *Int. J. Adv. Comput.*, vol. 36, pp. 1144–1153, Jan. 2013.
- [31] A. F. Agarap, "An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification," pp. 5–8, 2017, [Online]. Available: <http://arxiv.org/abs/1712.03541>.