The Academic
International Journal of Multidisciplinary Research

An Online Peer Reviewed / Refereed Journal
Volume 2 | Issue 10 | October 2024
ISSN: 2583-973X (Online)
Website: www.theacademic.in

Towards Enhanced Rice Foliar Disease Detection: A Methodological Approach of Preprocessing, Segmentation, and Feature Extraction

Govindarajan S
Research Scholar, PG & Research Department of Computer Science,
Presidency College (Autonomous), Chennai.

Dr. Mary Vennila S
Associate Professor & Head, PG & Research Department of Computer Science,
Presidency College (Autonomous), Chennai.

ARTICLE DETAILS

Research Paper

Keywords:

Rice leaf disease,

Preprocessing ,

Segmentation,

feature extraction,

Machine Learning ,

Deep Learning.

DOI:

10.5281/zenodo.14106337

ABSTRACT

Rice (Oryza Sativa) is a staple food for almost half of the global population, particularly in Africa and Asia, where it serves as a crucial dietary component. However, rice crops are highly susceptible to both abiotic stresses, such as drought and salinity, and biotic stresses, including pests and a variety of bacterial, viral, and fungal diseases. Identifying these diseases promptly is vital to maintain crop yield and quality. This paper, presents a comprehensive approach to rice leaf disease detection using advanced machine learning and deep learning techniques. By employing meticulous preprocessing, segmentation, and feature extraction methods, the study seeks to improve disease classification accuracy. The paper also compares the effectiveness of various techniques, demonstrating the promise of advanced deep learning algorithms in accurately identifying specific rice foliar diseases based on image data.

Introduction

India cultivates a wide variety of crops, many of which are susceptible to various insect pests and pathogens. The predominantly subtropical to tropical climate creates favorable conditions for pest proliferation, often more so than for disease-causing pathogens. Preventive measures and early diagnosis are essential to minimize the impact of these pathogens on crops. Effective disease detection enables producers to monitor crops closely, detect initial symptoms early, and take measures to control disease spread at a low cost, thereby preserving the bulk of the yield. Manual identification of diseases, however, is often prone to errors and misclassification. This paper presents a method that addresses these challenges by employing advanced image processing and neural network techniques to accurately identify and classify leaf diseases [1]. A disease in plants refers to any abnormal condition that damages the plant or disrupts its normal functions. Diseases are often identifiable through symptoms, which are visible changes observed in the plant. While numerous types of diseases affect rice crops, this study concentrates on three specific types that share similar symptoms: Bacterial Blight (BB), Brown Spot (BS), Leaf Blast (LB), and Healthy (H) as the control [2]. In this study, image processing techniques and a multiclass Support Vector Machine are applied to classify rice leaf diseases. Bacterial Blight (BB) affects the vascular system of the rice plant, causing water-soaked lesions that expand and turn yellow, leading to leaf drying and potential yield reduction. Brown Spot (BS), a fungal disease, targets leaves, leaf sheaths, and panicles, producing large brown spots that can damage leaf tissue and reduce grain quality if the infection spreads to seeds. Leaf Blast (LB), caused by the fungus *Magnaporthe oryzae*, is one of the most destructive rice diseases, forming spindle-shaped lesions on leaves and panicles that can lead to entire plant collapse under severe conditions. Proper detection and classification of these diseases are crucial for applying appropriate treatments to safeguard crop yield and quality [3]. Manual diagnosis is challenging and time-consuming, as it requires assessing numerous parameters. Therefore, implementing automated systems is essential to assist farmers in early and more accurate disease detection. In this process, advanced machine learning techniques play a crucial role in enhancing the accuracy of disease classification [4] [5].

Related Works

Phadikar et al. [6] developed an automated system for classifying brown spot and leaf diseases in rice plants by analyzing morphological changes. The study employed Otsu's segmentation algorithm to segment images, and radial hue distribution from the center to the edge of the spots was used as a feature for disease classification.

Kholis Majid et al. [7] contributed to the development of a mobile application for identifying paddy plant diseases using fuzzy entropy and a probabilistic neural network classifier, designed to run on the Android operating system. The application targets four specific diseases: brown spot, leaf blast, tungro, and bacterial leaf blight, achieving a diagnostic accuracy of 91.46%.

Rothe et al. [8] proposed a work in light of example acknowledgment framework for distinguishing proof and order of Versatile neuro-fuzzy derivation framework utilized Hu's moments as components for the preparation technique. The arrangement precision is 85 percent.

Ferentinos et al. [9] proposed a CNN model was trained on an extensive dataset of 87,848 images covering 25 plant varieties across 58 classes, including healthy plants. Among several models tested, the highest-performing model achieved an accuracy of 99.53% in correctly identifying the categories.

Similarly, Mohanty et al. [11] explored a CNN was trained on a dataset of 54,306 images spanning 14 crop types, comprising 26 diseases along with healthy leaves. Although this model reached an accuracy of 99.35%, its performance sharply dropped to 31.4% when applied to a real-world dataset, highlighting the challenges of generalizability in varied conditions.

Additionally, Xie et al. [12] explores the complexity of disease severity classification, emphasizing its greater difficulty compared to disease identification. This challenge is amplified by high intra-class similarities within the same disease category, which complicates accurate classification due to subtle variations among the images.

Gayathri Devi et al. [13] developed an automated approach for recognizing diseases in paddy leaves using image processing techniques. Their method employed a combination of grayscale co-occurrence matrix, discrete wavelet transform (DWT), and scale-invariant feature transform (SIFT) to extract relevant features. These extracted features were then fed into various classifiers—including multiclass SVM, Naïve Bayes, backpropagation neural network, and K-nearest neighbors (KNN)—to classify the plants as diseased or healthy.

Kaya et al. [14] examined the impact of four distinct transfer learning models on deep neural networks for plant classification, utilizing four public datasets. Their findings showed that transfer learning significantly enhances automated plant identification and can notably improve models that previously had lower classification performance.

Dos et al. [15] applied Convolutional Neural Networks (CNNs) for detecting weeds in soybean crop images, classifying them into grass and broadleaf categories. They developed an image database with over fifteen thousand images, capturing soil, soybean plants, and both types of weeds. The CNNs in their study, representing a deep learning architecture, demonstrated notable effectiveness in image recognition tasks.

Methodology

The rice leaf disease detection system follows key stages, including image acquisition, preprocessing, segmentation, feature extraction, and classification. This process is illustrated in figure 1.

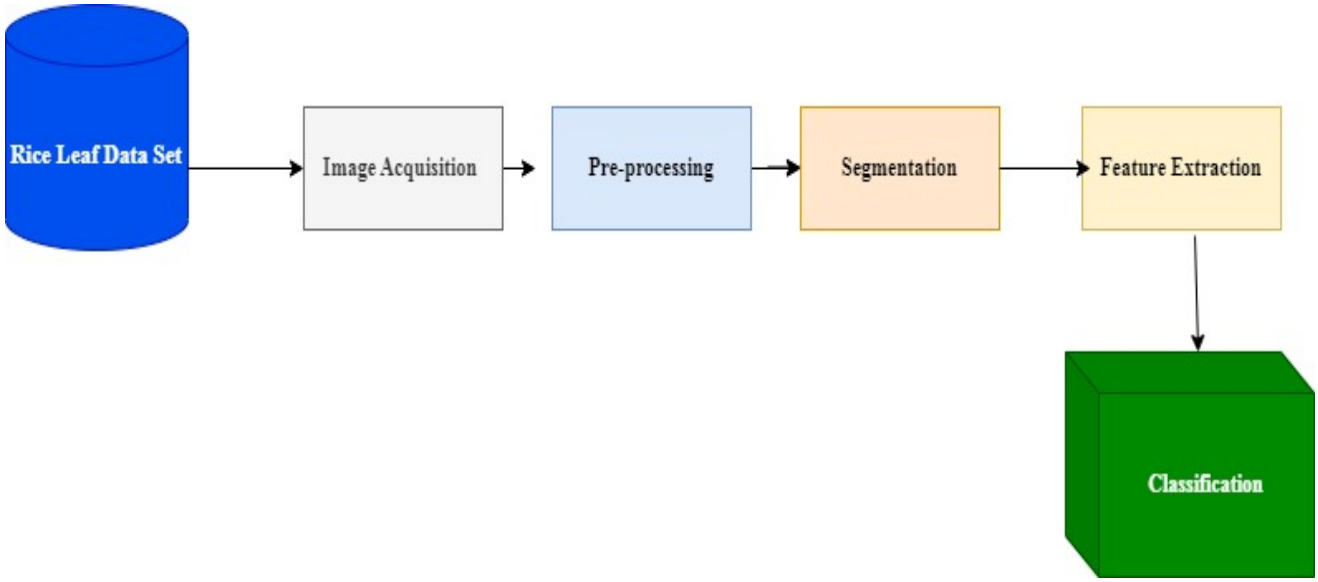

Figure 1: Approach for detecting diseases in rice leaf.

Results and Discussion

All results and discussions were processed and run in the Google Colab environment for efficient execution and analysis.

Rice Leaf Data Set

The rice leaf dataset is split into training and testing sets, with the training set organized into four subfolders (Bacterial blight, Brown spot, Leaf smut, and Healthy), each containing 35 images. Similarly, the test set has four subfolders with 15 images each, designated for evaluating the model, as illustrated figure 2. in the bar graph below.

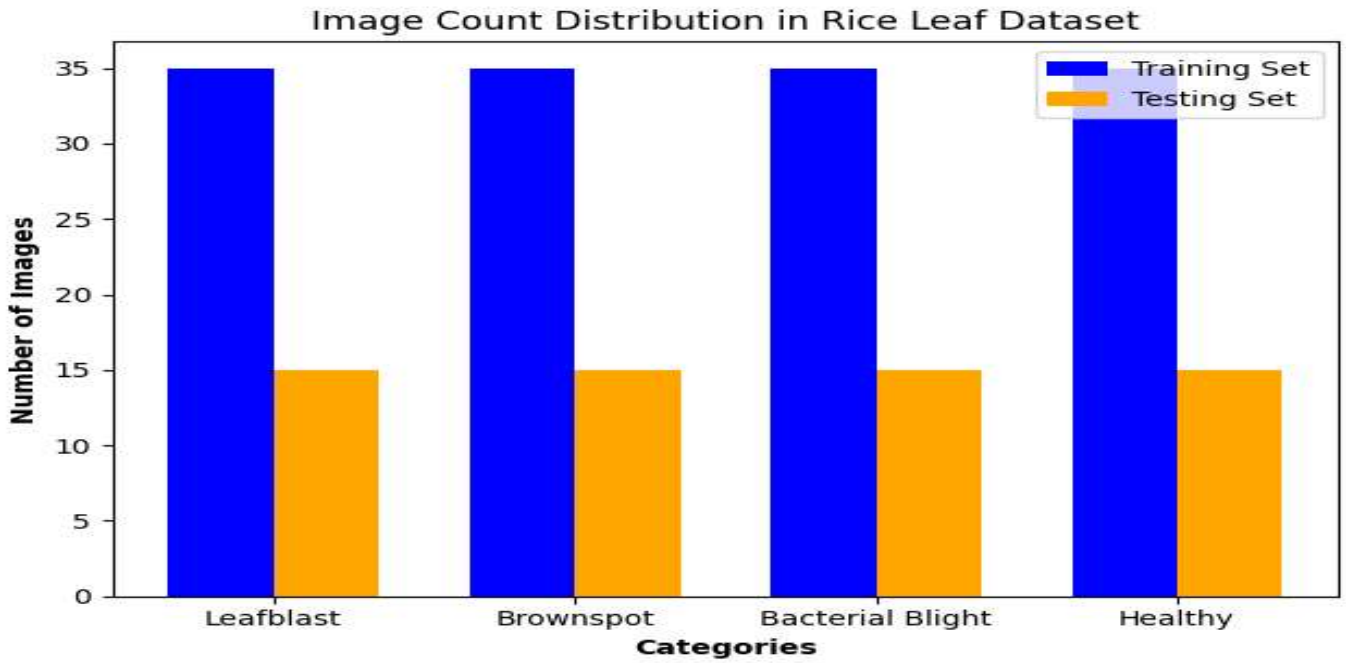

Figure 2: Distribution of Image Counts in Training and Testing Datasets for Rice Leaf Diseases.

Image Acquisition

The rice leaf dataset utilized in this research was obtained from a public repository on Kaggle.

Preprocessing

In this study, we implemented a comprehensive image preprocessing pipeline to enhance the quality of the rice leaf dataset obtained from Kaggle. The preprocessing process began with reading images from both the training and testing datasets which is shown in figure 3. Each image was resized to a uniform dimension of 256x256 pixels to ensure consistency across the dataset. Following resizing, the images were converted to grayscale, which simplifies the data and reduces computational complexity without losing critical information necessary for subsequent analysis.

Image Resizing:

$$I_{Resized}(x\ and\ ) = I_{Original}\ \frac{x}{IN_{original}}\ .IN_{new},\frac{and}{H_{origianl}}\ .H_{new}$$

$$I_{gray}(x\ and\ ) = 0.299.R(x,and) + 0.587.G(x,and) + 0.114.B(x,and)$$

Gray Scale Conversion

Computational Complexity Reduction:

$$Complexity_{original} = O(N.C)$$

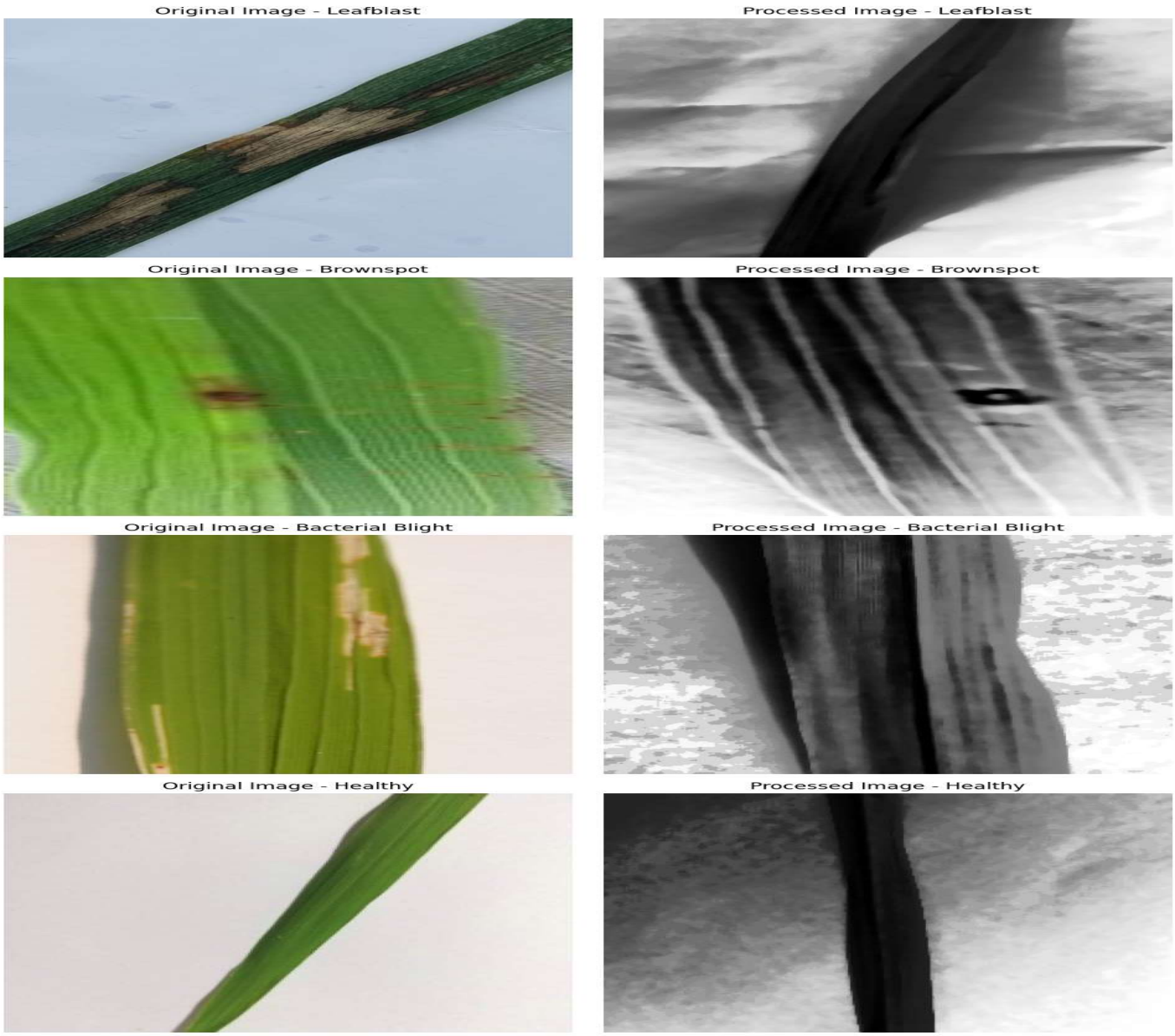

Figure 3: Comparison of Original and Processed Rice Leaf Images Across Different Disease Categories.

To further improve the image quality, a median filter was applied for denoising, effectively reducing the impact of noise that could interfere with feature extraction. Finally, histogram equalization was performed to enhance contrast and improve the visibility of details within the images. The processed images were then saved into corresponding output directories, preserving the structure of the original dataset. This preprocessing step is crucial for ensuring the robustness and accuracy of the machine learning models used in the classification of rice leaf diseases.

Segmentation

To prepare the rice leaf images for classification, segmentation was performed to isolate and highlight key areas indicative of disease. Using Otsu's thresholding and Canny edge detection methods, each image was segmented to clearly differentiate leaf regions from the background. The segmentation process was applied to both the preprocessed training and testing datasets, organized into four categories: Bacterial blight, Brown spot, Leaf smut, and Healthy which is shown in figure 4.

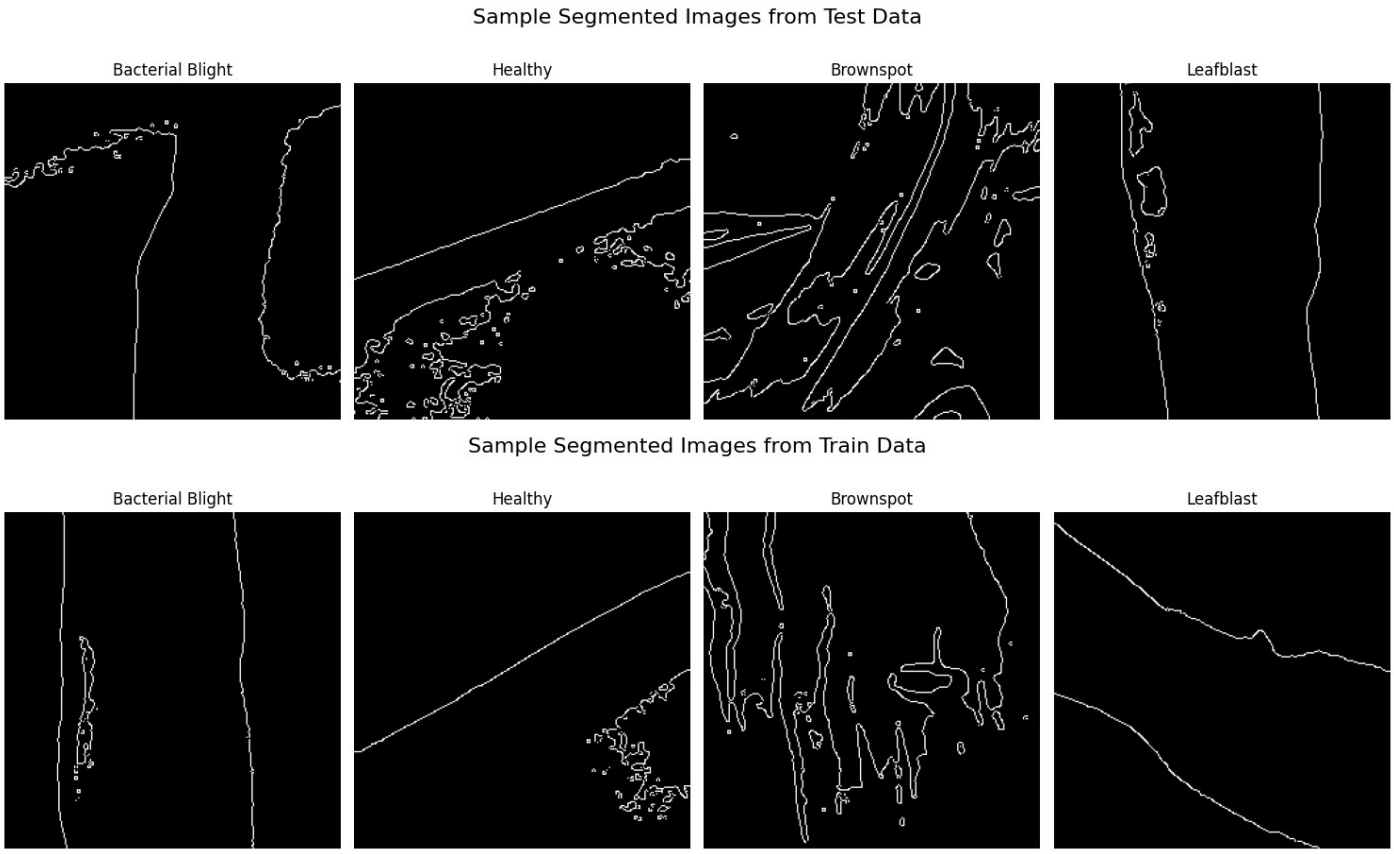

Figure 4: Sample Segmented Images of Rice Leaf Disease Categories in Training and Testing Datasets

Initially, Otsu's thresholding was used, a method that calculates an optimal threshold value automatically by minimizing intra-class variance in grayscale images. This technique produced binary images that effectively isolated leaf areas, helping to eliminate irrelevant background details. Following this, Canny edge detection was applied to emphasize the boundaries of the segmented regions, allowing for clear demarcation of features that may contain disease indicators.

Segmented images were saved in dedicated directories for each category, with each image processed and stored in its respective folder. This segmentation step provided a clean, focused dataset where only the leaf regions were retained, facilitating accurate feature extraction and improving the performance of subsequent classification algorithms.

Feature Extraction

In this study, feature extraction was applied to segmented rice leaf images using two key techniques: Histogram of Oriented Gradients (HOG) and Gray-Level Co-occurrence Matrix (GLCM), aiming to capture crucial shape and texture information for classification of rice leaf diseases. Paths to segmented training and testing datasets were set, with each dataset containing subfolders categorized by disease type.

For shape-based features, the HOG method was implemented through gradient orientation and magnitude calculations, where each image was divided into cells, and histograms of gradient orientations were generated to produce a vector representing the image's overall shape. Additionally, GLCM features were extracted to provide statistical texture analysis using properties such as contrast, dissimilarity, homogeneity, energy, and correlation by examining pixel intensity pairs at specific distances and angles.

The feature extraction process was automated using the extract_features() function, which iterated through each category folder to process images in grayscale format. HOG and GLCM features were extracted for each image, combined into a single feature vector, and stored with the category label which is shown in figure 5. The final output was organized into a DataFrame with columns for each category and the respective extracted features.

Train Features:

	Category	HOG_0	HOG_1	HOG_2	HOG_3	HOG_4	HOG_5	HOG_6	\
0	Bacterial Blight	0.0	0.0	0.0	0.0	0.002787	0.0	0.0	
1	Bacterial Blight	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	
2	Bacterial Blight	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	
3	Bacterial Blight	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	
4	Bacterial Blight	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	

	HOG_7	HOG_8	...	HOG_34591	HOG_34592	HOG_34593	HOG_34594	HOG_34595	\
0	0.0	0.0	...	0.363569	0.000805	0.363569	0.002908	0.001139	
1	0.0	0.0	...	0.000000	0.000000	0.000000	0.000000	0.000000	
2	0.0	0.0	...	0.000000	0.000000	0.000000	0.000000	0.000000	
3	0.0	0.0	...	0.322116	0.000000	0.281524	0.001320	0.250173	
4	0.0	0.0	...	0.000000	0.000000	0.000000	0.000000	0.000000	

	GLCM_Contrast	GLCM_Dissimilarity	GLCM_Homogeneity	GLCM_Energy	\
0	2460.208517	9.899694	0.886786	0.844178	
1	1694.051593	6.817341	0.922106	0.892365	
2	1126.773591	4.526379	0.951390	0.934815	
3	1217.191498	4.867142	0.953191	0.936375	
4	2110.273775	8.386887	0.935552	0.917456	

	GLCM_Correlation
0	0.330542
1	0.330801
2	0.167078
3	0.215321
4	0.037479

[5 rows x 34602 columns]

Test Features:

	Category	HOG_0	HOG_1	HOG_2	HOG_3	HOG_4	HOG_5	HOG_6	\
0	Bacterial Blight	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	
1	Bacterial Blight	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	
2	Bacterial Blight	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	
3	Bacterial Blight	0.075658	0.0	0.106579	0.0	0.002364	0.0	0.0	
4	Bacterial Blight	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.0	

	HOG_7	HOG_8	...	HOG_34591	HOG_34592	HOG_34593	HOG_34594	HOG_34595	\
0	0.0	0.0	...	0.000000	0.000000	0.000000	0.000000	0.000000	
1	0.0	0.0	...	0.235894	0.002905	0.247291	0.001024	0.116604	
2	0.0	0.0	...	0.000000	0.000000	0.000000	0.000000	0.000000	
3	0.0	0.0	...	0.316355	0.000790	0.190772	0.000395	0.133839	
4	0.0	0.0	...	0.000000	0.000000	0.000000	0.000000	0.000000	

	GLCM_Contrast	GLCM_Dissimilarity	GLCM_Homogeneity	GLCM_Energy	\
0	1189.829994	4.762469	0.953324	0.938169	
1	1011.027665	4.041360	0.961397	0.947808	
2	2533.200092	10.182200	0.889864	0.849075	
3	1949.911734	7.876440	0.899686	0.859754	
4	1492.386872	5.996094	0.934445	0.909448	

	GLCM_Correlation
0	0.159656
1	0.204476
2	0.349100
3	0.354354
4	0.311165

[5 rows x 34602 columns]

Figure 5: Sample Extracted Features from Training and Testing Datasets Using HOG and GLCM Methods.

The extracted features for both training and testing datasets were then saved to CSV files (train_features.csv and test_features.csv) for further analysis. This structured approach to feature extraction efficiently handles large datasets, providing a comprehensive feature set ready for use in classification tasks.

To view the extracted features, the CSV files for the training and testing datasets (train_features.csv and test_features.csv) were loaded into DataFrames. This step allows inspection of the feature data to ensure that feature extraction was successful and organized as expected.

Using train_features.head() and test_features.head(), the first five rows of each dataset were displayed. Each row represents an image from the dataset, with columns containing the category label and feature values obtained from HOG and GLCM methods. This display enables a quick visual verification of the extracted features and confirms that both datasets are structured and ready for subsequent classification analysis.

Algorithm for Preprocessing, Segmentation, and Feature Extraction

Step 1: Preprocessing

1. **Input**: Raw images from the dataset containing subfolders for each disease category.
2. **For each image in the dataset**:
    - Read the image using OpenCV.
    - Convert the image to grayscale.
    - Resize the image to a standard dimension (e.g., 256x256 pixels).
    - Apply denoising techniques (e.g., median filter) to reduce noise.
    - Perform histogram equalization to improve contrast.
3. **Output**: Preprocessed images saved in a designated segmented dataset folder structure, maintaining the category organization.

Step 2: Segmentation

1. **Input**: Preprocessed images.
2. **For each preprocessed image**:
    - Apply segmentation techniques (e.g., color-based segmentation, edge detection) to isolate relevant areas of the image (i.e., leaf regions).
    - Store the segmented images in a specified folder for further analysis.
3. **Output**: Segmented images saved in a separate folder structure, maintaining the category organization.

Step 3: Feature Extraction

1. **Input**: Segmented images from the previous step.
2. **Initialize**: Prepare to collect features in a data structure (e.g., a list or DataFrame).
3. **For each category in the dataset**:
    - Access the folder containing segmented images of that category.
    - **For each image in the category**:
        - Read the image in grayscale format.
        - Extract HOG features:
            - Divide the image into cells and calculate the gradient orientations and magnitudes.
            - Generate a feature vector representing the shape of the image.
        - Extract GLCM features:
            - Compute the Gray-Level Co-occurrence Matrix for the image.
            - Calculate texture properties: contrast, dissimilarity, homogeneity, energy, and correlation.
        - Combine the HOG and GLCM feature vectors into a single feature vector.
        - Append the category label and feature vector to the data structure

Conclusion

The primary focus is on the essential tasks of preprocessing, segmentation, and feature extraction of the rice leaf dataset. These foundational steps are crucial for preparing the dataset for future classification tasks. The preprocessing stage involves cleaning and enhancing the images to improve the quality of the input data. Techniques such as grayscale conversion, resizing, denoising, and histogram equalization are employed to ensure that the images are standardized and free from noise. This step is vital, as the performance of any subsequent classification model heavily relies on the quality of the input data.

Following preprocessing, the segmentation process isolates the relevant portions of the images, specifically the areas corresponding to the rice leaves. By employing various segmentation techniques, the methodology aims to accurately delineate the leaf regions from the background, ensuring that the feature extraction process is conducted on the most relevant parts of the images. This step is instrumental in minimizing background noise and enhancing the features that are critical for identifying disease patterns.

Feature extraction is the next crucial step, where key characteristics of the segmented images are quantified. By implementing methods such as Histogram of Oriented Gradients (HOG) and Gray-Level Co-occurrence Matrix (GLCM), significant shape and texture features are derived from the images. These features serve as valuable inputs for machine learning and deep learning models in future work. The combination of these feature extraction techniques provides a robust representation of the images, capturing essential information that is pivotal for effective disease classification.

References:

Mangla, N., Raj, P. B., Hegde, S. G., & Pooja, R. (2019). Paddy leaf disease detection using image processing and machine learning. Int J Innov Res Elec Electron Instrument Control Eng, 7(2), 97-99.

Kurniawati, N. N., Abdullah, S. N. H. S., Abdullah, S., & Abdullah, S. (2009, December). Investigation on image processing techniques for diagnosing paddy diseases. In 2009 international conference of soft computing and pattern recognition (pp. 272-277). IEEE.

Satgunalingam, V., & Thaneeshan, R. (2020). Automatic Paddy Leaf Disease Detection Based on GLCM Using Multiclass Support Vector Machine. Int. J. Comput, 39(1), 97-106.

S. Ghosal and K. Sarkar, "Rice Leaf Diseases Classification Using CNN With Transfer Learning," 2020 IEEE Calcutta Conference (CALCON), 2020, pp. 230-236, doi: 10.1109/CALCON49167.2020.9106423.

Santanu Phadikar and Jaya Sil, Rice Disease Identification using Pattern Recognition Techniques, Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008) 25-27 December, 2008. DOI:10.1109/ICCITECHN.2008.4803079.

S. Phadikar, J. Sil, and A. K. Das, "Classification of rice leaf diseases based on morphological changes," International Journal of Information and Electronics Engineering, vol. 2, no. 3, p. 460, 2012.

Kholis Majid, Yeni Herdiyeni, Annu Rauf, "I-Pedia: Mobile Application For Paddy Disease Identification Using Fuzzy Entropy And Probabilistic Neural Network", ICACSIS, 2013.

P. R. Rothe, "Cotton Leaf Disease Identification Using Pattern Recognition Techniques", International Conference On Pervasive Computing, 2015.

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and electronics in agriculture, 145, 311-318.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, 1419.

Xie, S., Yang, T., Wang, X., & Lin, Y. (2015). Hyper-class augmented and regularized deep learning for fine-grained image classification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2645-2654).

Gayathri Devi, T., & Neelamegam, P. J. C. C. (2019). Image processing based rice plant leaves diseases in Thanjavur, Tamilnadu. Cluster Computing, 22(Suppl 6), 13415-13428.

Kaya, A., Keceli, A. S., Catal, C., Yalic, H. Y., Temucin, H., & Tekinerdogan, B. (2019). Analysis of

transfer learning for deep neural network based plant classification models. Computers and electronics in agriculture, 158, 20-29.