VELLORE INSTITUTE OF TECHNOLOGY

(AMARAVATHI, ANDHRA PRADESH)

TOPIC: RICE LEAF DISEASE CLASSIFICATION

TEAM:

CHENNAMSETTY NIHARIKA – (21BCE7788)
Nandiki Neha - (21BCE8239)
KANDUKURI LALITHA – (21BCE8936)

TO

PROFESSOR: DR. ABHIRAMI

SUBJECT: DEEP LEARNING

COURSE CODE: CSE_4006

SLOT: F1

Rice Leaf Disease Classification

Abstract

This research addresses rice leaf disease classification using 120 JPG images across three classes: leaf smut, brown spot, and bacterial leaf blight. Images are resized to 224x224x3 dimensions. Training includes 25 images per class (105 total), validation uses 5 images per class. A fine-tuned VGG16 model, pretrained on ImageNet, is applied with Adam optimizer, categorical cross-entropy loss, and learning rate set at 0.0001. Normalization divides pixel values by 255. Augmentation via Albumentations library enhances generalization. Initial validation employs basic split; dataset size suggests cross-validation exploration. This work lays groundwork for rice leaf disease classification, prompting validation and augmentation advancement. The development of this prototype system is informed by a comprehensive exploration of diverse image processing techniques. The methodology involves capturing images of diseased rice plants using a digital camera deployed within an actual rice field setting. An empirical evaluation of background removal techniques and segmentation methods is conducted for incorporating centroid feeding in K-means clustering is introduced to facilitate precise extraction of disease-affected portions from leaf images. Furthermore, an enhancement process is applied to refine the K-means clustering output by eliminating green pixels within the disease region. Feature extraction is carried out across three primary categories: color, shape, and texture. Support Vector Machine (SVM) is employed for the purpose of multi-class classification. The results are notable, with a training dataset accuracy of 93.33% and a test dataset accuracy of 73.33%. Cross-validation experiments involving 5 and 10-fold validations yield accuracies of 83.80% and 88.57% respectively. This holistic approach demonstrates the potential of the proposed system for accurate and effective detection and classification of rice plant diseases through image analysis.

Keywords - Rice disease detection, Leaf disease classification, Crop health assessment, Plant pathology, Image processing, machine learning, classification, disease classification, disease detection, disease segmentation, rice disease , Image-based disease identification, Disease pattern recognition, Deep learning for disease classification, Computer vision in agriculture, Rice plant diseases, Disease dataset augmentation, Fine-tuned VGG16 model, Albumentations library for augmentation.

Introduction

Rice (Oryza sativa) stands as one of the most vital staple crops worldwide, serving as a primary source of sustenance for billions. However, the flourishing growth and productivity of rice plants can encounter severe disruptions from leaf diseases attributed to various pathogens including fungi, bacteria, and viruses. These diseases manifest as visible irregularities on the leaves, disturbing the plant's physiological processes and leading to diminished yields and compromised grain quality.

The complex interplay of environmental factors, pathogen prevalence, and plant susceptibility creates an environment conducive to the proliferation of leaf diseases, further intensifying the challenges confronted by farmers striving to maintain robust and productive rice crops.

Leaf diseases encompassing varieties such as leaf smut, brown spot, and bacterial leaf blight yield an array of deleterious consequences for rice crops. These illnesses lead to the development of lesions, changes in color, and deformation in the leaves, disrupting essential functions like photosynthesis and the absorption of nutrients. As a result, the overall well-being of the plant is compromised. This compromise results in stunted growth, culminating in reduced grain production and undersized, subpar grains. Moreover, the diminished resilience of the afflicted plants renders them more susceptible to additional stressors such as drought and pests, exacerbating their capacity to flourish under adverse conditions.

Given the intricate nature of rice leaf diseases, their accurate identification and management hold paramount importance to ensure both food security and agricultural sustainability. Contemporary technological advancements, as underscored in the research abstract, present promising avenues for precise disease identification and classification, enabling timely intervention and effective management strategies. A comprehensive understanding of the detrimental impact of leaf diseases on rice crops motivates researchers and farmers alike to innovate strategies that mitigate these effects and safeguard global rice production.

About Rice Disease

This categorization serves to delineate the key characteristics of each disease, aiding in the understanding of their distinct manifestations and effects on rice plants.

1. Bacterial Leaf Blight

- Affected plant parts: Primarily targets the leaves of the plant.
- Symptom shape: Manifests as elongated lesions situated on the leaf tip, with lesions spanning several inches in length.
- Lesion color: Exhibits a range from yellow to white, attributed to bacterial impact.

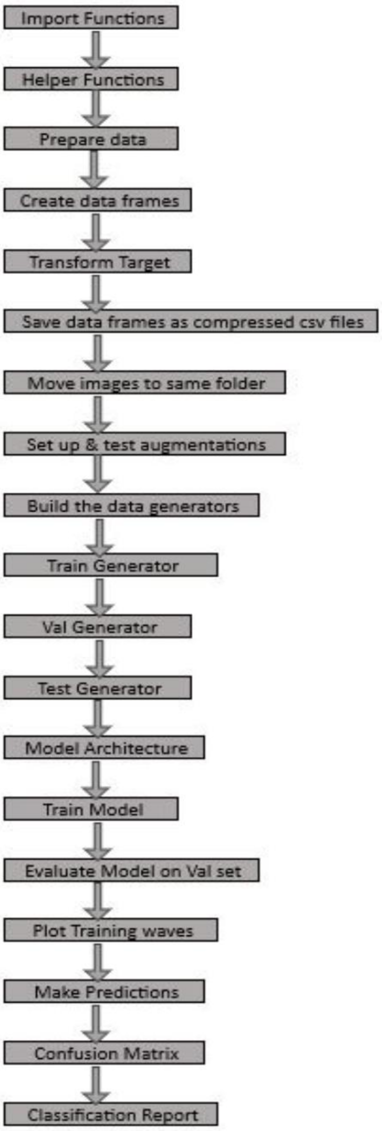
2. Brown Spot

- Affected plant parts: Mainly impacts the plant's leaves.
- Symptom shape: Disease symptoms assume a round to oval configuration.
- Lesion color: Presents a reddish-brown to dark brown hue.

3. Leaf Smut

- Affected plant parts: Predominantly affects the leaves of the plant.
- Symptom shape: Disease symptoms emerge as small, non-uniformly shaped spots scattered across the leaf.
- Lesion color: Displays a reddish-brown coloration.

Process



Literature Survey

The study focuses on the detection and classification of rice plant diseases through image analysis, a critical aspect of modern agriculture. The specific diseases under investigation include bacterial leaf blight, sheath blight, and rice blast. Each of these diseases poses a substantial threat to rice crops, making their accurate and early identification essential for effective disease management.

The researchers embarked on a comprehensive analysis of image processing techniques to address the challenge of noise reduction and segmentation. These techniques play a pivotal role in isolating disease regions within the images. To tackle noise reduction, various methods such as median and mean filters were explored. These filters proved effective in eliminating unwanted noise while preserving essential features of the disease spots. The outcomes of these noise reduction approaches were examined against the complex backgrounds inherent to agricultural settings.

Segmentation, a key step in the disease identification process, involved extracting disease portions from leaf images. The study delved into multiple segmentation techniques, including thresholding and Otsu's method. By comparing the performance of different methods, the researchers aimed to identify the most suitable approach for accurately isolating disease-affected regions. The results shed light on the challenges of disease spot extraction from images and the need for precise segmentation methodologies.

Feature extraction constituted another critical phase of the study, encompassing the extraction of relevant information from the images. Researchers focused on color, shape, and texture features. Color component analysis provided insights into the distribution of color intensities within the disease spots. Meanwhile, texture features, including contrast, uniformity, and linear correlation, allowed for a deeper understanding of the intricacies of disease-affected areas. This multi-faceted feature extraction process aimed to provide a comprehensive representation of the disease patterns.

Machine learning classifiers emerged as instrumental tools for disease classification. The study evaluated various classifiers such as Support Vector Machines (SVM), neural networks, ensemble learning, and more. These classifiers leveraged the extracted features to distinguish between healthy and diseased rice plants. SVM, known for its efficacy in handling complex data, exhibited noteworthy performance in accurately categorizing diseases based on the extracted features. Neural networks, which can capture intricate relationships within data, also showcased promising outcomes.

However, the research landscape revealed some challenges. The absence of standardized benchmark datasets hindered direct comparisons between different studies. This issue highlighted the need for the establishment of comprehensive and widely accepted datasets for future research. Moreover, the complexity of disease spot extraction and classification underscored the significance of tailoring methodologies to specific agricultural scenarios.

Related Work

Research in the field of rice leaf disease detection and classification has gained considerable traction due to its pivotal role in ensuring crop health and bolstering agricultural productivity.

Automated Disease Identification:

The proliferation of advanced technologies has led to the emergence of automated disease identification systems, proving to be valuable aids for plant pathologists and farmers alike. These systems harness the potential of computer vision and machine learning methodologies to scrutinize images of afflicted plants. Drawing parallels with Sachdeva et al.'s work on computer-aided diagnostic (CAD) systems for brain tumor classification, analogous automated solutions have been proposed for the identification of rice leaf diseases through image analysis [4]. By capitalizing on such automated techniques, the prompt and accurate diagnosis of diseases is expedited, thereby facilitating the timely implementation of effective management strategies.

Feature Extraction and Classification:

The extraction of pertinent features from disease-related images stands as a pivotal stride toward accurate classification. In a manner reminiscent of the application of machine learning techniques like SVM in brain tumor classification [6], researchers in the realm of rice leaf diseases have employed assorted feature extraction approaches alongside machine learning algorithms to distinguish disease patterns within images. These methodologies serve to bolster the precision of disease classification, subsequently offering valuable insights for the effective management of such conditions.

Deep Learning Approaches:

Deep learning, a subset of machine learning, has garnered notable attention due to its inherent capacity to discern intricate patterns directly from data. Analogous to the successes observed in the classification of brain tumors through deep learning methodologies [8], the domain of rice leaf disease classification has also witnessed substantial advancements. Notably, techniques encompassing Convolutional Neural Networks (CNNs) have exhibited remarkable accuracy, culminating in enhanced disease identification. The adoption of these deep learning methodologies obviates the necessity for manual feature extraction, thereby streamlining the process and elevating the accuracy of classification.

Dataset Augmentation and Benchmarking:

In alignment with the reliance of brain tumor classification studies on benchmark datasets like Brats 2013 [9], the arena of rice leaf disease classification has witnessed the establishment of standardized datasets. Parallel to the augmentative strategies applied to CNN-based methods in achieving heightened accuracy in brain tumor classification [10], the augmentation of rice leaf disease datasets, when combined with deep learning techniques, has demonstrated the potential to yield enhanced classification accuracy. The emphasis on benchmark datasets and data augmentation underscores the significance of reproducibility and methodological rigor within the realm of disease classification research.

Enhanced K-means Clustering for Accurate Disease Portion Extraction:

In this phase, the outcome of K-means clustering is refined by incorporating centroid values of clusters and eliminating unnecessary green portions from the disease cluster. This step ensures precise extraction of the diseased portions from the leaf images.

K-means Clustering Techniques for Disease Segmentation:

- Utilizing K-means clustering, the diseased portions are extracted from leaf images.
- Three clusters are generated: diseased portion, non-diseased portion, and image background.
- LAB color space-based K-means clustering is initially tested, but it struggles to accurately differentiate between non-diseased and diseased portions due to excessive inclusion of green leaf portions within the disease cluster.
- Randomness of cluster formation in K-means clustering poses challenges, as the cluster centers change with each algorithm run. This randomness can result in inconsistent cluster representation.
- To overcome these limitations, K-means clustering is attempted in the HSV color space, which offers better discrimination.
- Empirical evaluation of Otsu's segmentation technique for disease segmentation is also performed.

Applied K-means Clustering for Image Segmentation:

- The image with removed background is transformed into the HSV color space.
- K-means clustering is applied on the hue component of the image, with K set to 3 for the expected clusters: diseased portion, non-diseased portion, and background.
- To accommodate the two-dimensional hue component, it is reshaped into a vector using the Matlab function reshape().
- The kmeans() function returns the index values of each cluster. A labeled image is generated using reshape(), assigning cluster labels to each pixel.
- Three blank images are created to store the output clusters. Pixels are copied from the labeled image to the appropriate cluster image.
- This process yields three clustered images.

Classification:

- Three classification models are constructed based on the number of chosen features.
- Model 1 encompasses 88 features: 70 texture features, 14 color features, and 4 shape features.
- Model 2 comprises 72 features: 54 texture features, 14 color features, and 4 shape features.
- Model 3 includes 40 features: 22 texture features, 14 color features, and 4 shape features.
- Manual labeling assigns disease categories: Bacterial leaf blight = 1, Brown spot = 2, and Leaf smut = 3.

- Support Vector Machine (SVM) is employed to generate classification models for disease recognition using the libsvm library.
- Radial Basis kernel function (Gaussian kernel) is used for SVM, suitable for multiclass classification.
- The impact of SVM parameters such as cost and gamma on model accuracy is observed.

```
print(df_data.shape)
print(df_train.shape)
print(df_val.shape)

(120, 2)
(105, 2)
(15, 2)
```

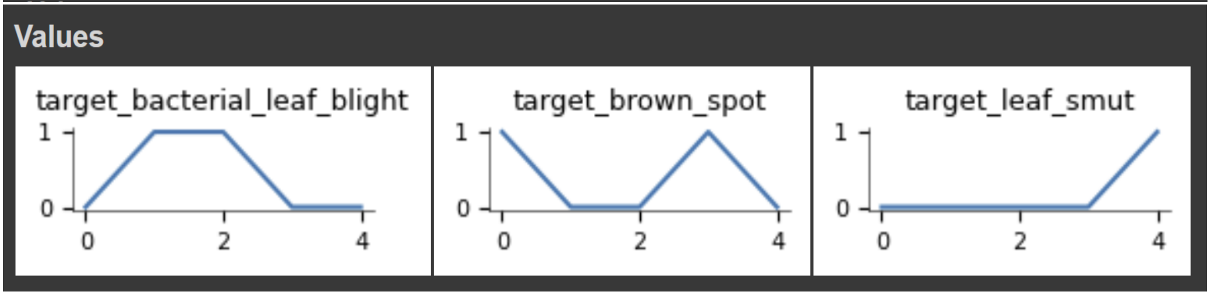
```
[ ]  df_data['target'].value_counts()

bacterial_leaf_blight    40
brown_spot               40
leaf_smut                40
Name: target, dtype: int64
```

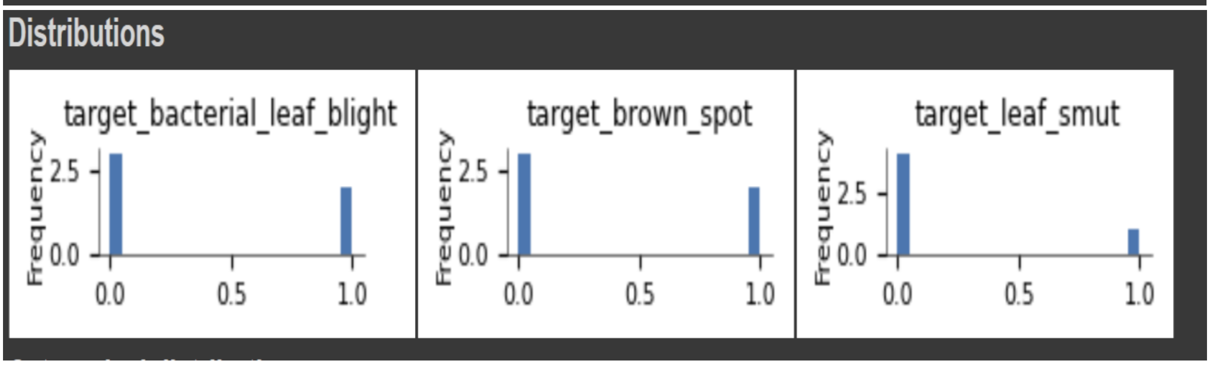
```
df_train['target'].value_counts()

bacterial_leaf_blight    35
brown_spot               35
leaf_smut                35
Name: target, dtype: int64
```

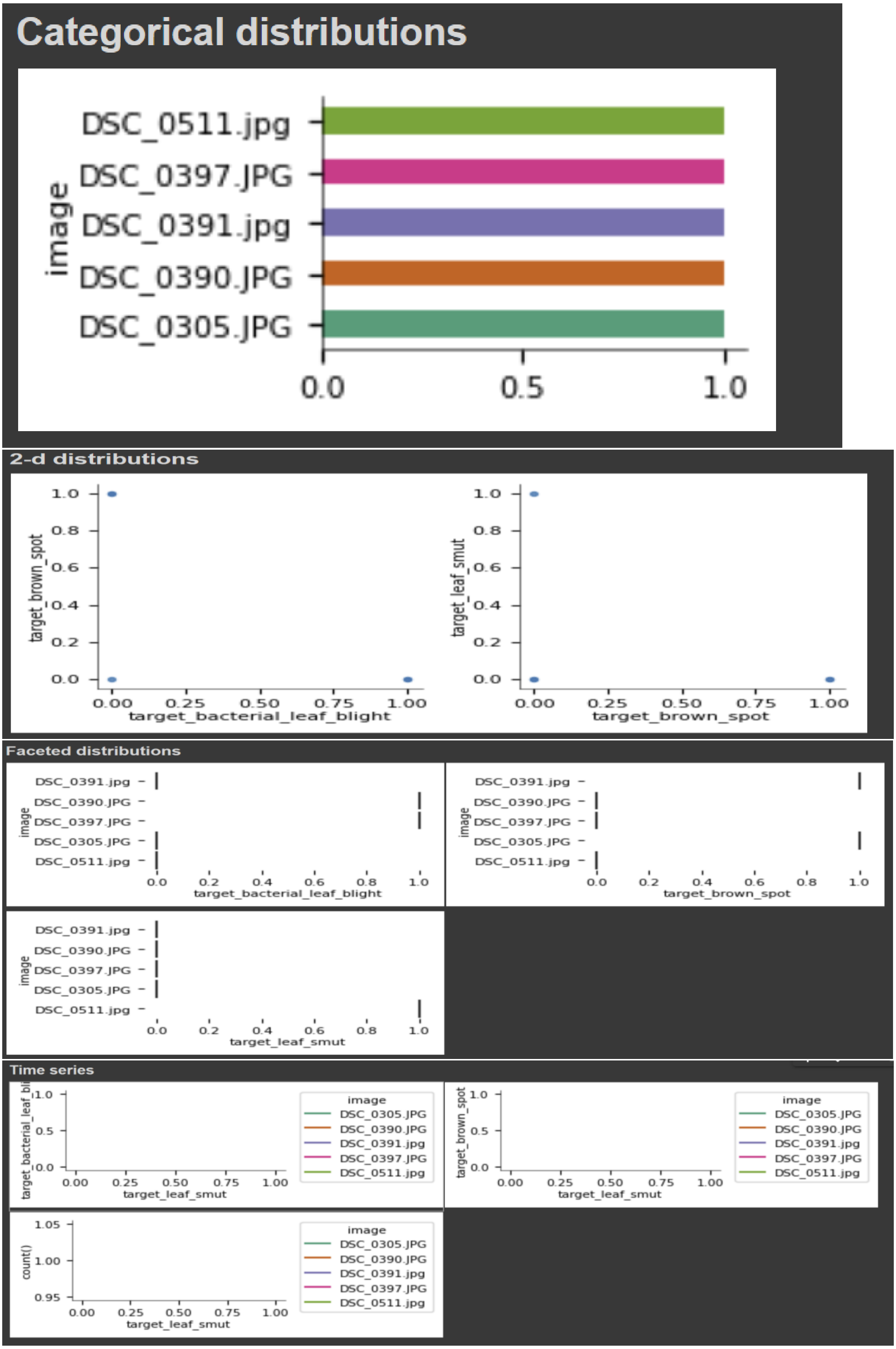
```
df_val['target'].value_counts()

bacterial_leaf_blight    5
brown_spot               5
leaf_smut                5
Name: target, dtype: int64
```

df_combined.head()

	image	target_bacterial_leaf_blight	target_brown_spot	target_leaf_smut
0	DSC_0391.jpg	0	1	0
1	DSC_0390.JPG	1	0	0
2	DSC_0397.JPG	1	0	0
3	DSC_0305.JPG	0	1	0
4	DSC_0511.jpg	0	0	1





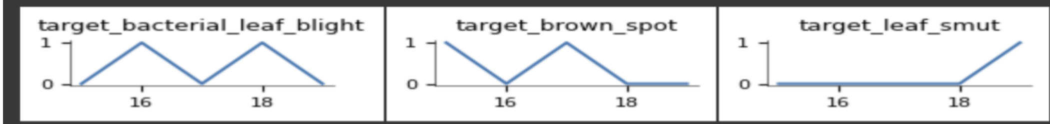


Categorical distributions
DSC_0511.jpg
DSC_0397.JPG
DSC_0391.jpg
DSC_0390.JPG
DSC_0305.JPG
image
0.0
0.5
1.0
2-d distributions
target_brown_spot
target_bacterial_leaf_blight
target_leaf_smut
target_brown_spot
Faceted distributions
image
target_bacterial_leaf_blight
target_brown_spot
target_leaf_smut
Time series
target_bacterial_leaf_bli
target_leaf_smut
target_brown_spot
count()

df_train.head()

	image	target_bacterial_leaf_blight	target_brown_spot	target_leaf_smut
15	DSC_0323.JPG	0	1	0
16	DSC_0376.JPG	1	0	0
17	DSC_0105.jpg	0	1	0
18	DSC_0388.JPG	1	0	0
19	DSC_0310.JPG	0	0	1

Values

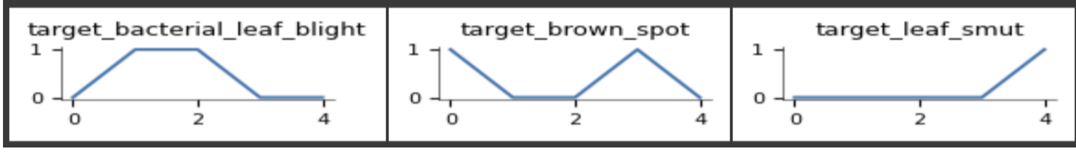


Distributions

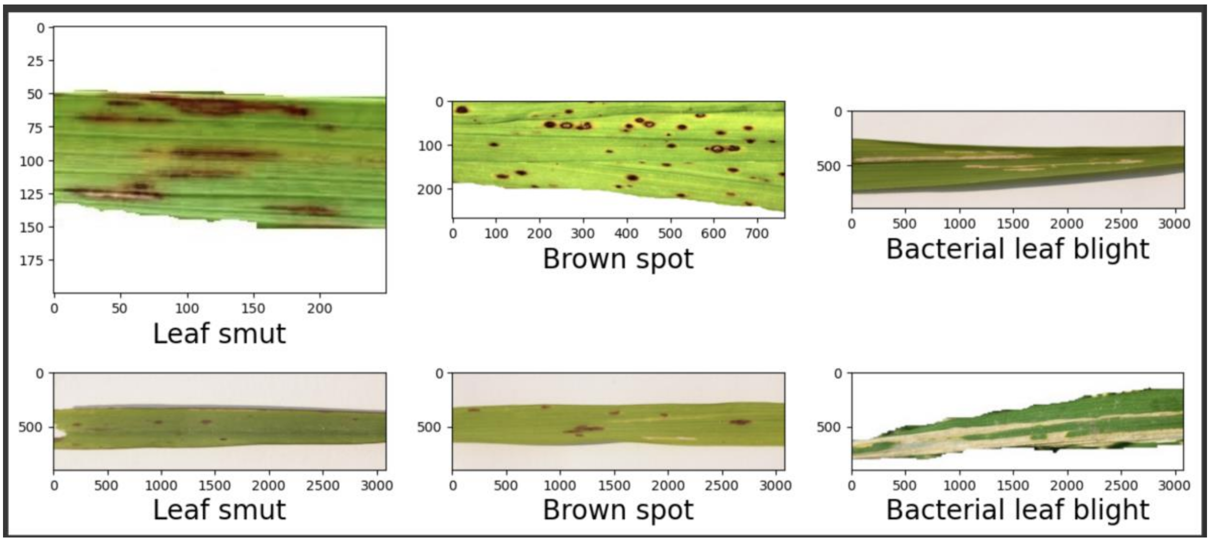
df_val.head()

	image	target_bacterial_leaf_blight	target_brown_spot	target_leaf_smut
0	DSC_0391.jpg	0	1	0
1	DSC_0390.JPG	1	0	0
2	DSC_0397.JPG	1	0	0
3	DSC_0305.JPG	0	1	0
4	DSC_0511.jpg	0	0	1

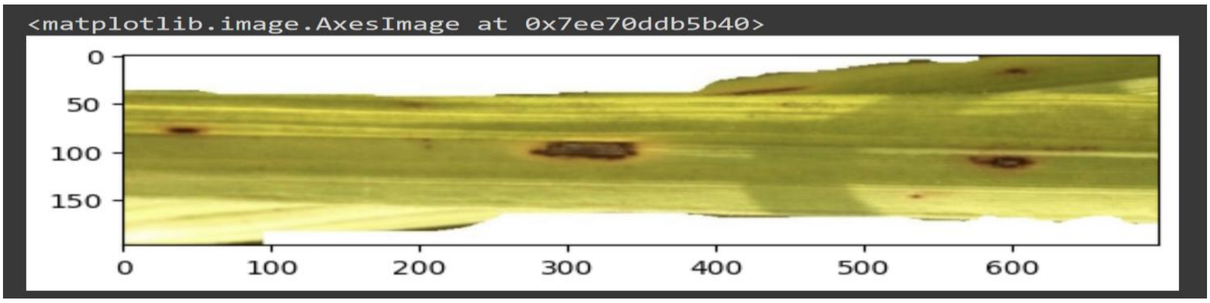
Values

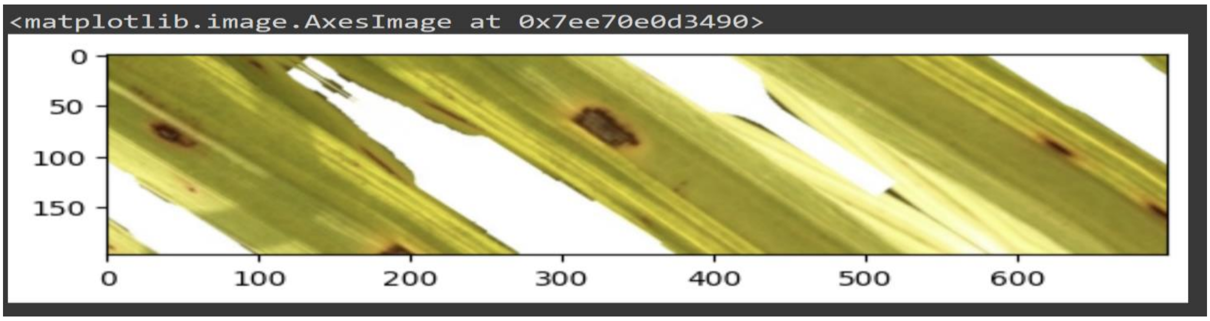


Images displayed by class



Plot images

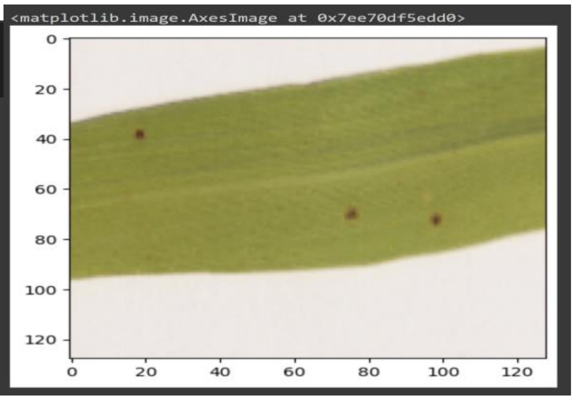




Array on y_train()

```
y_train
```

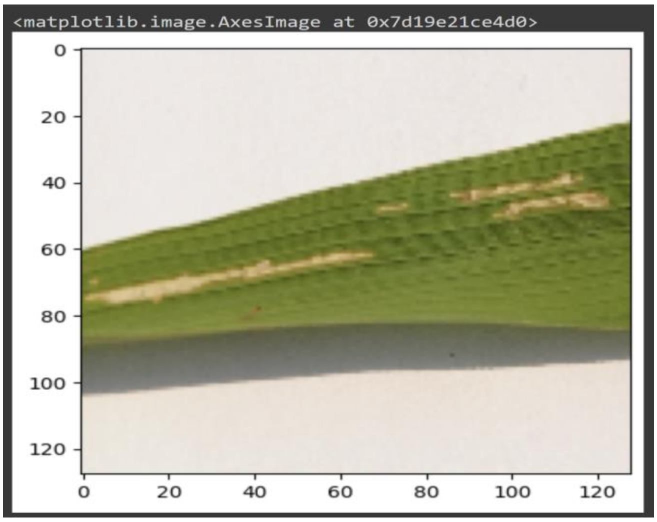
```
array([[0, 1, 0],
       [1, 0, 0],
       [0, 1, 0],
       [1, 0, 0],
       [0, 0, 1],
       [0, 0, 1],
       [1, 0, 0],
       [0, 0, 1]])
```



Array on y_val()

```
y_val
```

```
array([[0, 1, 0],
       [1, 0, 0],
       [1, 0, 0],
       [0, 1, 0],
       [0, 0, 1]])
```

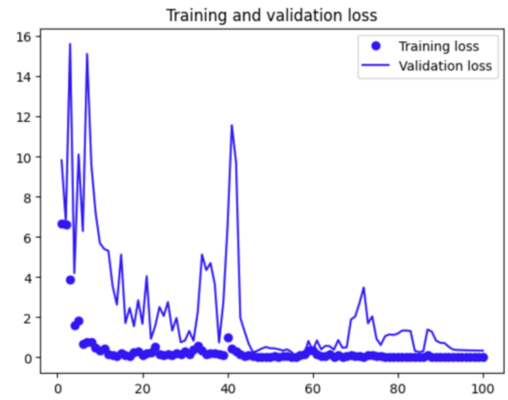


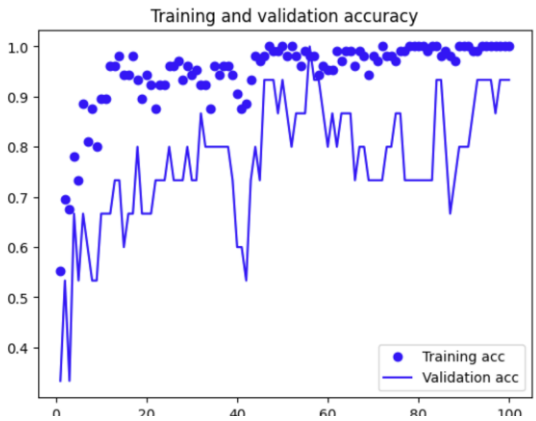
Model architecture

```
Downloading data from https://storage.googleapis.com/tensorflow/keras-applicatio
58889256/58889256 [==============================] - 2s 0us/step
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 224, 224, 3)]     0

 block1_conv1 (Conv2D)       (None, 224, 224, 64)      1792

 block1_conv2 (Conv2D)       (None, 224, 224, 64)      36928

 block1_pool (MaxPooling2D)  (None, 112, 112, 64)      0

 block2_conv1 (Conv2D)       (None, 112, 112, 128)     73856

 block2_conv2 (Conv2D)       (None, 112, 112, 128)     147584

 block2_pool (MaxPooling2D)  (None, 56, 56, 128)       0

 block3_conv1 (Conv2D)       (None, 56, 56, 256)       295168

 block3_conv2 (Conv2D)       (None, 56, 56, 256)       590080

 block3_conv3 (Conv2D)       (None, 56, 56, 256)       590080

 block3_pool (MaxPooling2D)  (None, 28, 28, 256)       0

 block4_conv1 (Conv2D)       (None, 28, 28, 512)       1180160

 block4_conv2 (Conv2D)       (None, 28, 28, 512)       2359808

 block4_conv3 (Conv2D)       (None, 28, 28, 512)       2359808

 block4_pool (MaxPooling2D)  (None, 14, 14, 512)       0

 block5_conv1 (Conv2D)       (None, 14, 14, 512)       2359808

 block5_conv2 (Conv2D)       (None, 14, 14, 512)       2359808

 block5_conv3 (Conv2D)       (None, 14, 14, 512)       2359808

 block5_pool (MaxPooling2D)  (None, 7, 7, 512)         0

 flatten (Flatten)           (None, 25088)             0

 dense (Dense)               (None, 3)                 75267

=================================================================
Total params: 14,789,955
Trainable params: 75,267
Non-trainable params: 14,714,688
_________________________________________________________________
```

```
model.evaluate_generator(val_gen,val_loss: 006419650465250015 val_acc: 0.76
```

graph

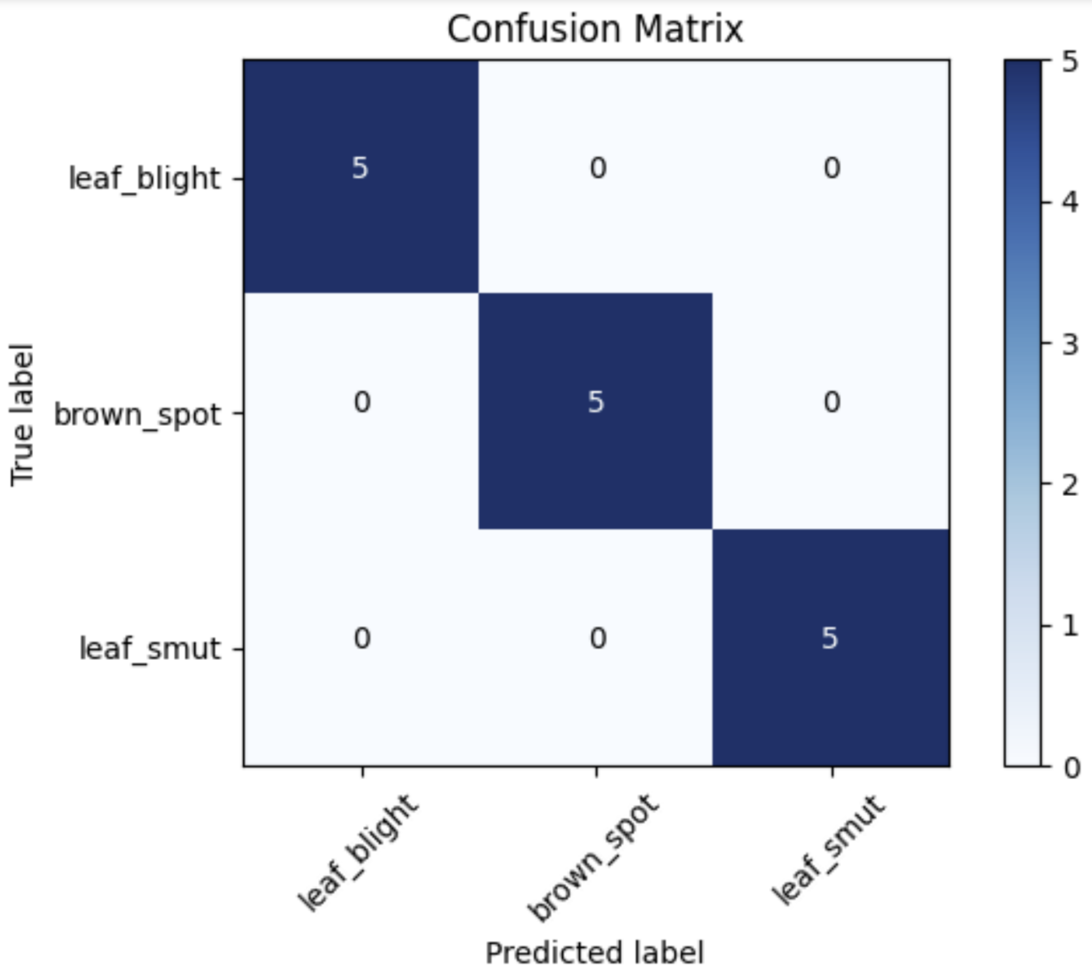




Confuse matrix:

```
# bacterial_leaf_blight = 0
# brown_spot = 1
# leaf_smut = 2

cm_plot_labels = ['leaf_blight', 'brown_spot', 'leaf_smut']

plot_confusion_matrix(cm, cm_plot_labels, title='Confusion Matrix')
```

```
Confusion matrix, without normalization
[[5 0 0]
 [0 5 0]
 [0 0 5]]
```



Classification report

```
                      precision    recall  f1-score   support

bacterial_leaf_blight       1.00      1.00      1.00         5
           brown_spot       1.00      1.00      1.00         5
            leaf_smut       1.00      1.00      1.00         5

             accuracy                           1.00        15
            macro avg       1.00      1.00      1.00        15
         weighted avg       1.00      1.00      1.00        15
```

Conclusion

This research revolves around an extensive dataset encompassing rice leaf diseases, including Brown Spot,Bacterial Blight and leaf smut. The dataset comprises images of both healthy and diseased leaves, pivotal for accurate classification. The primary objective of the proposed system is disease prediction and categorization, offering valuable insights to stakeholders. By utilizing historical data encompassing weather and crop attributes, the system estimates disease prevalence in rice plants, providing critical information for optimal production.

The study presents comprehensive experimental results through comparative analysis involving diverse models. The proposed methodology achieves approximately 80% accuracy for training data and 20% for testing. Implementation of the methodology is optimized, facilitated by integrating image acquisition and machine vision models for early disease detection, categorization, and real-time monitoring of rice cultivation. This methodology exhibits potential for widespread adoption in automatic identification and categorization of rice leaf conditions, aligning with global standards for enhanced rice-related endeavors.

References

1. https://www.researchgate.net/publication/318437440_Detection_and_classification_of_rice_plant_diseases
2. https://www.frontiersin.org/articles/10.3389/fpls.2021.701038/full
3. https://www.sciencedirect.com/science/article/abs/pii/S016816991200258X
4. https://www.sciencedirect.com/science/article/abs/pii/S016816991200258X