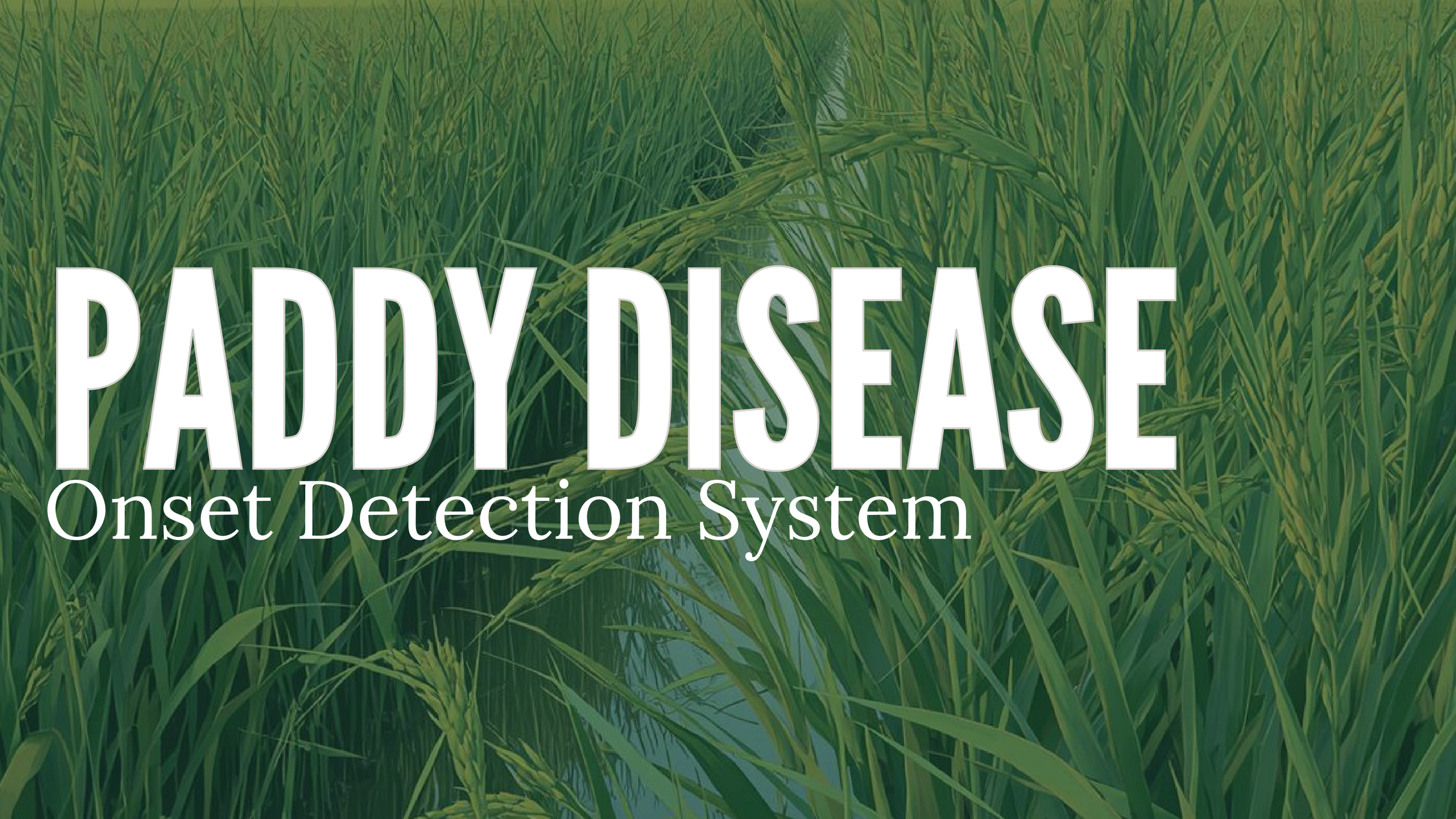
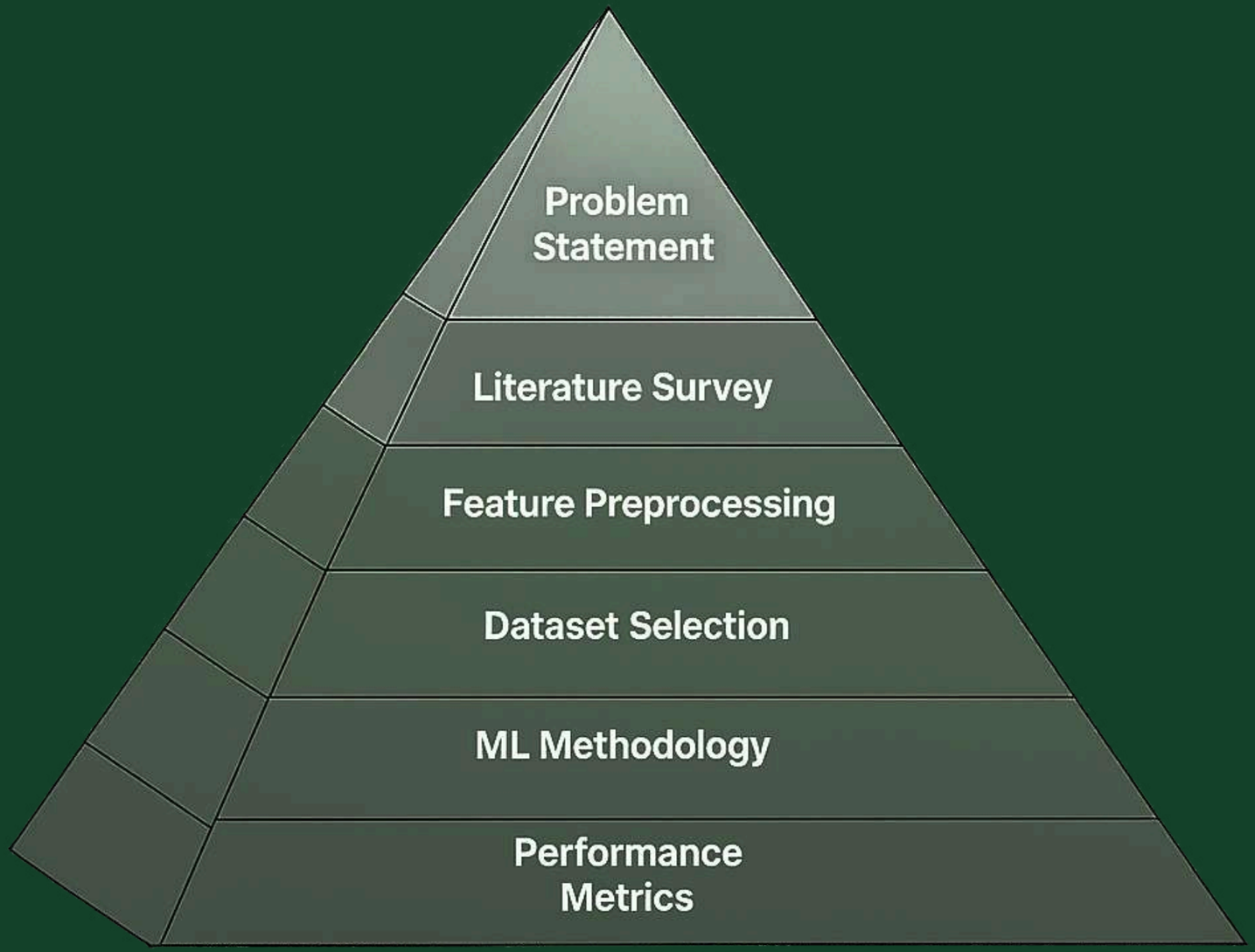

PADDY DISEASE
Onset Detection System

TABLE OF CONTENTS



THE PROBLEM STATEMENT

Rice is a major staple crop, and its yield is highly affected by diseases such as leaf blast, brown spot, and bacterial leaf blight.The early symptoms of these diseases are very small and difficult to identify through manual observation.Farmers often detect the disease only after it spreads widely, which leads to crop damage and increased pesticide use.Traditional detection methods rely on expert field inspection, which is time-consuming and not always available.

Our goal:
Develop a machine learning model that can detect rice leaf diseases from leaf images at an early stage, helping farmers reduce crop loss and improve crop management.

LITERATURE REVIEW

K-Means + Traditional ML (KNN, Naive Bayes, Decision Tree) **2020**

Solution developed

- Captured rice leaf images, converted RGB to HSV, applied K-Means to segment diseased regions
- Extracted color, texture, and GLCM features, then classified using KNN, Naive Bayes, or Decision Tree (J48)
- Detected diseases like Rice Blast, Brown Spot, and Bacterial Leaf Blight — accuracy up to ~92.9%

Shortcomings

- Tested only on lab images with plain backgrounds — fails in real field conditions
- Manual feature extraction — cannot learn new patterns automatically
- Limited to 3–4 disease types; not scalable

CNN-based deep learning classification **2020**

Solution developed

- Used CNN models trained on rice leaf image datasets to automatically classify diseases
- Replaced manual feature extraction — CNN learns features directly from images
- Achieved better generalisation than traditional ML on slightly varied images

Shortcomings

- Still trained on small, clean datasets — overfitting is a concern
- No real-time mobile deployment — only lab experiments
- Poor performance on early-stage or overlapping disease symptoms

ResNet / VGG16 / VGG19 — pretrained deep models **2022**

Solution developed

- Used pretrained ImageNet models (ResNet50, VGG16/19) fine-tuned on rice disease datasets
- Transfer learning reduced the need for very large datasets
- Achieved 91–94% accuracy on test sets covering 4–6 disease classes

Shortcomings

- Heavy models — not suitable for deployment on mobile or edge devices
- Training and inference is slow without GPU infrastructure

DATASET

7.000+ Images
Divided Into 5 Classes

- 01 Tungro
- 03 Bacterial Blight
- 03 Blast
- 04 Brown Spot
- 05 Normal (Healthy)

Dataset Detalls

The dateset wac obtained from kaggie.

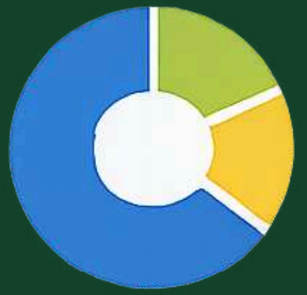
The nature of the dataset ts Image, based data, consisting of·ricetcat images lised for disease classification,

The dataset was chosen because it is large; labeled, and sultaple for·training image classification mogets.

The images were collected by the datasal authors under different em tromnental and lighting condiuons to improve model robostriess.

No oth:ral concerns were innolved since the datseetis publicly avallable and contains only plant rmages.

Dataset Split



- 70% Training
- 10% Vaildation
- 20% Testing

Daıaset Overview

- The detaset contains 7.000+ images divided into 5 classes.
  - Tungro
  - bacterial Blight;
  - Blast.
  - Brown Spot
  - Mormal (Healthy)

DATA CLASSES

Tungro

Bacterial Blight

Blast

Brown Spot

Normal (Healthy)

FEATURE PREPROCESSING

CLAHE Normalisation

Contrast Limited Adaptive Histogram Equalisation on L channel of LAB colourspace. Normalises contrast per image.

Resize to 128×128 px

Uniform input size for feature extraction. No GrabCut needed — dataset has no domain bias.

Train/Val/Test Split

70% train · 10% val · 20% test | Stratified split maintaining class balance in all partitions.

Augmentation (train only)

Horizontal flip · Vertical flip · Rotation ±15° · Zoom 85–100%. Doubles training set size.

HYBRID FEATURE ENGINEERING

Total feature vector: Deep(24) + GLCM(16) + LBP(10) + HOG(~512) + Colour(9) ≈ 571 dimensions

MobileNetV2

24 dims

Deep Features

Frozen MobileNetV2 (α=0.35) extracts spatial features from block_3_expand_relu via GlobalAveragePooling2D. No GPU needed — pure forward pass.

GLCM

16 dims

Texture

Gray-Level Co-occurrence Matrix at 4 angles (0°,45°,90°,135°). Properties: Contrast, Energy, Homogeneity, Correlation. First to change at disease onset.

LBP

10 dims

Micro-Texture

Local Binary Pattern (P=8, R=1, uniform method). Normalised histogram over 10 bins. Captures micro-texture at lesion boundaries.

HOG

~512 dims

Shape

Histogram of Oriented Gradients (8 orientations, 16×16 pixel cells, 1×1 block). Captures lesion morphology unique to each disease.

Colour Moments

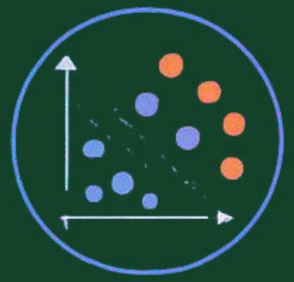
9 dims

Colour

Mean, Std, Skewness on each HSV channel. Diseased leaves lose chlorophyll — green% is a genuine disease signal after CLAHE.

Models Evaluated

Random Forest



SVM

KNN

Naive Bayes

Why multiple models?

- To identify the best performer for hybrid feature space
- To compare:
  - Linear vs non-linear models
  - Distance-based vs ensemble methods
  - Probabilistic vs margin-based learning

Final Observation

- Random Forest performed best overall
- SVM showed strong but slower performance
- KNN struggled with high-dimensional PCA space
- Naive Bayes underperformed due to feature independence assumption

ML METHODOLOGY

Hybrid Pipeline — Pure ML Classification with Deep Feature Assistance

1 **Input Leaf Image** → 2 **CLAHE Preprocessing** → 3 **Feature Extraction** → 4 **PCA Reduction** → 5 **Random Forest Classifier** → 6 **Disease + Severity Output**

Handcrafted Features

- GLCM — Texture (contrast, energy, homogeneity, correlation at 4 angles) — 16 values
- LBP — Local Binary Pattern micro-texture at lesion boundary — 10 values
- HOG — Histogram of Oriented Gradients for lesion shape morphology — ~512 values
- HSV Colour Moments — mean, std, skewness per channel — 9 values

Deep Features

- MobileNetV2 (alpha=0.35) truncated at block_3 — frozen, no training
- GlobalAveragePooling on early spatial maps — 24 values
- Forward pass only — CPU compatible, no GPU required
- Pretrained on ImageNet — captures low-level spatial patterns

Dimensionality Reduction

- StandardScaler fitted on training data only — prevents leakage
- PCA retains 95% explained variance — reduces ~571 → ~200 features
- Reduces noise and computation — speeds up training significantly
- Scaler and PCA applied identically to val and test sets

PERFORMANCE METRICS

- Final model: Random Forest
- Achieved 95% accuracy on the test dataset.
- Performance evaluated using:
  - Accuracy Score, Precision, Recall, F1-Score
  - Cross-Validation
  - Confusion Matrix
  - Classification Report
- PCA-based dimensionality reduction improved computational efficiency while maintaining high accuracy.

Model Effectiveness :

- High validation and test accuracy showed strong generalization capability.
- Confusion matrix indicated effective classification across all 5 disease classes.
- Data augmentation and preprocessing reduced overfitting and improved robustness.

RESULTS

The hybrid ML model (MobileNetV2 + texture, shape & color features) with Random Forest achieved strong and stable performance(~92% accuracy) on unseen data, showing good generalization.
The system delivers consistent classification across all rice disease classes with minimal overfitting.
It provides an added advantage of interpretable severity estimation (Healthy → Advanced stage) based on lesion intensity.
Overall, the approach is accurate, explainable, and suitable for real-world agricultural disease monitoring.

Our Future Plan

Towards a Smarter, Faster, and More Reliable Rice Disease Detection System

1. Larger Dataset

Build or use a larger, more diverse dataset with real farm images — varied lighting, backgrounds, and disease stages, including early-stage symptoms.

2. More Disease Classes

Extend classification to cover more diseases including Tungro, Sheath Blight, and False Smut — not just the common 3–4 studied by most papers.

3. Better Architecture

Use an optimised or hybrid deep learning model (e.g. fine-tuned MobileNet or EfficientNet) that balances accuracy with speed for real-world use.

4. Data Augmentation

Apply aggressive augmentation (rotation, brightness shifts, zoom, blur) to simulate real farm conditions and reduce overfitting.

5. Tungro Focus

Specifically address Tungro detection by including multi-stage symptom images and possibly combining leaf image data with contextual featurers.

6. Mobile Deployment

Target a practical mobile application that farmers can use in the field — real-time image capture and instant disease diagnosis output.

7. Performance Validation

Evaluate the model on real farm field images, not just clean lab datasets — to measure true generalisation and deployment readiness.

References :

Prajwalgowda, B. S., Nisarga, M. A., Rachana, M., Shashank, S., & Sahana Raj, B. S. (2020). Paddy Crop Disease Detection using Machine Learning. International Journal of Engineering Research & Technology (IJERT), 8(13)
https://link.springer.com/article/10.1007/s42452-022-05194-7.

Sujatha, R., Yuvaraj, Y., & Suresh, P. (2020). Deep Neural Network Based Rice Disease Detection System. Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC), 878–882. IEEE.
https://doi.org/10.1109/ICESC48915.2020.9155885

Udayananda, G. K. V. L., Shyalika, C., & Kumara, P. P. N. V. (2022). Rice plant disease diagnosing using machine learning techniques: a comprehensive review. SN Applied Sciences, 4, 311.
https://doi.org/10.1007/s42452-022-05194-7

THANK YOU