Wheat Disease Classification Using Machine Learning Techniques

Fahim Arsad Nafis, Mirza Kamrul Bashar Shuhan, Nafisa Mehjabin, Shahana Shultana
Department of Computer Science
George Mason University
Fairfax, Virginia, USA
{fnafis2, mshuhan, nmehjabi, sshultan}@gmu.edu

***Abstract*—Wheat diseases present a significant challenge to global food security, highlighting the necessity of early detection and effective classification systems. This study introduces a machine-learning approach to classify wheat diseases using diverse image datasets covering various breeds and regions. To address the challenges of imbalanced datasets, we apply preprocessing techniques such as colour standardization, data augmentation and weight balancing, which significantly enhance model performance. We utilize five models for classification: a customized Convolutional Neural Network (CNN), MobileNetV2, ResNet50, Vision Transformer (ViT), and Convolutional Vision Transformer (CvT). Recognizing that accuracy alone is insufficient for imbalanced datasets, we use the F1-score, normalized confusion matrices, ROC curves, and Grad-CAM visualizations for detailed analysis. Notably, simpler models like the Vision Transformer outperformed more complex architectures.**

***Index Terms*—Wheat, Disease, Image Classification, Machine Learning**

I. Introduction

Wheat is a vital global crop that provides a fifth of the world's food calories, making its production critical for food security and increasing demand for wheat-based foods [1]. Despite the large-scale production of wheat, there is a pressing need to increase yield to meet the demands of a growing population. Recent trends indicate that global wheat yield losses an average of 21.5% which is the most significant loss occurring in food-deficit regions [2]. This presents a significant challenge for agricultural researchers. Moreover, factors such as climate conditions, fertilizer usage, unpredictable abiotic and biotic stresses, and disease attacks can all negatively impact wheat yield. Among these, wheat stripe rust disease poses a significant threat, causing a 30-40% reduction in production, as highlighted in [3]. Wheat is susceptible to several other types of fungal diseases such as Stripe rust, which is caused by the airborne fungus *Puccinia striiformis*. It spreads rapidly within the field and primarily affects the leaves, but it can also impact glumes and awns, leading to infection hotspots [3]. There are three main types of wheat rust diseases: (a) yellow/stripe rust, (b) leaf/brown rust, and (c) stem/black rust, all of which can affect the entire wheat plant. In some regions of South Asia, shortages of fungicides and protective equipment have hindered timely control which further complicated the management of these diseases [4].

Crop disease diagnosis is traditionally done manually through field visits, microscopes, or labour-intensive methods, which are time-consuming and prone to errors. While techniques like spectroscopy can detect diseases, they are costly and require specialized equipment and expertise. As there are over 25,000 varieties of wheat adapted to various temperate climates, creating a generalized solution for disease control becomes a more difficult task. Additionally, many wheat diseases are interrelated, making identification and classification based solely on symptoms particularly challenging. Early and accurate diagnosis is essential for minimizing yield losses, but since manual inspection is labour-intensive, there is a need in the agricultural research domain for automated disease detection to improve efficiency and reduce costs [5].

Given the large scale of global wheat production, advanced image-based techniques for automatic disease classification offer an efficient solution. Machine learning models can analyze and classify wheat disease symptoms from images, providing real-time insights for farmers. In this study, we classify wheat diseases using labelled images of infected plants, leveraging state-of-the-art advanced models to detect various wheat diseases based on visual symptoms. Our study aims to improve disease detection and help reduce crop losses while answering these questions-

- **RQ1:** How can we develop a machine learning-based system to classify wheat plant diseases using real-world image data considering variations across different wheat breeds and regions?
- **RQ2:** How do various transfer learning pre-trained models perform in the context of wheat disease classification?

We employ a range of machine learning approaches, including Convolutional Neural Networks (CNN), pre-trained models like MobileNetV2, ResNet50, and ViT, as well as custom CvT models for enhanced performance in wheat disease classification. Additionally, we apply data mining techniques to analyze the results and ensure the accuracy of disease classification. Our study demonstrates the importance of balanced class weighting to handle imbalanced datasets and effective preprocessing enhances the model's performance.

II. Background

In this section, we briefly discuss the machine learning models (Sect. II-A) and the evaluation metrics (Sect. II-B) we used for this study.

A. Machine Learning Models

Convolutional Neural Networks [6] are particularly well-suited for image classification tasks, making them an ideal choice for detecting and classifying wheat diseases. Their ability to automatically learn spatial hierarchies of features through convolutional layers enables accurate detection of subtle patterns in disease symptoms, even amidst variations in lighting, angles, or background noise.

MobileNetV2 [7] is an efficient deep-learning model designed for resource-constrained environments, making it ideal for tasks requiring a balance between accuracy and computational efficiency. Its use of depthwise separable convolutions and inverted residual blocks enables faster processing with reduced memory requirements, while still capturing critical features for image classification.

ResNet50 [8] is a powerful deep-learning model designed to address the vanishing gradient problem through its use of residual connections, enabling the training of very deep networks. Its ability to extract complex hierarchical features makes it highly effective for image classification tasks, such as wheat disease detection.

Vision Transformers (ViTs) [9] is a cutting-edge deep learning model that leverages the transformer architecture for image classification tasks. Unlike CNNs, ViTs divide images into patches and process them as sequences, enabling the model to capture long-range dependencies and global context effectively. This makes them particularly suitable for complex datasets with diverse patterns, such as wheat disease classification, where understanding spatial relationships is crucial.

Convolutional Vision Transformers (CvTs) [10] combine the strengths of CNNs and ViTs to achieve high performance in image classification tasks. By integrating convolutional layers into the transformer architecture, CvTs effectively capture both local features (via convolutions) and global relationships (via self-attention mechanisms). This hybrid approach makes them well-suited for tasks like wheat disease classification, where both fine-grained details and broader spatial dependencies are crucial.

B. Evaluation Metrics and Tools

Learning curves plot the model's performance (e.g., accuracy or loss) over training epochs for both the training and validation datasets. These curves help identify issues such as overfitting or underfitting.

Confusion matrix is a tabular representation of the model's performance, summarizing the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) predictions. It is structured as:

$$\begin{bmatrix} \text{TP} & \text{FP} \\ \text{FN} & \text{TN} \end{bmatrix}$$

This matrix provides insights into specific types of errors made by the model.

F1 score is the harmonic mean of precision and recall, offering a balanced evaluation of a model's performance, especially under class imbalance. It is defined as:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1)$$

Where:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

Receiver Operating Characteristic (ROC) curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various thresholds. The Area Under the Curve (AUC) quantifies the model's ability to distinguish between classes. TPR and FPR are given by:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (3)$$

Gradient-weighted Class Activation Mapping (Grad-CAM) [11] is a visualization technique that highlights regions of an input image that strongly influence the model's prediction. It uses the gradients of the target class with respect to feature maps to generate a heatmap, aiding in interpretability.

Loss functions are used to measure the difference between the model's predictions and the actual target values. They provide a feedback signal to optimize the model parameters during training. The Cross Entropy Loss is widely used for multi-class classification. It computes the negative log probability of the true class, encouraging the model to assign higher probabilities to the correct class.

For a batch of predictions p and true labels y:

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^{N} \text{class\_weights}[y_i] \cdot \log(p_i[y_i]) \quad (4)$$

where:

- N: Number of samples in the batch.
- $p_i[y_i]$: Predicted probability for the true class y_i.
- `class_weights`: Weight for each class to handle imbalance.

III. Dataset

The details of the datasets used for this project are summarized in Table I. Each dataset has been assigned a unique ID for better understanding and reference.

The first dataset [12] comprises high-resolution images of wheat plants affected by various diseases, along with healthy plants. The dataset includes 14 classes, although the Stem Fly and Tan Spot classes are excluded due to inaccurate

TABLE I: Dataset Details

ID	Dataset Name	Origin	Dataset Size	No of Images
1	Wheat Disease Data Set	Worldwide, mainly South Asia	17.47 GB	21200
2	Wheat Leaf dataset	Ethiopia	1.52 GB	407
3	CGIAR Computer Vision for Crop Disease	Ethiopia and Tanzania	3 GB	1486
4	DAE-Mask	China	1.21 GB	2236

Fig. 1: Sample Images of Dataset Classes - Aphid (1), Black Rust (2), Brown Rust (3), Head Blight (4), Leaf Blight (5), Mildew (6), Mite (7), Septoria (8), Smut (9), Stripe Rust (10), Yellow Dwarf (11), and Healthy wheat (12)

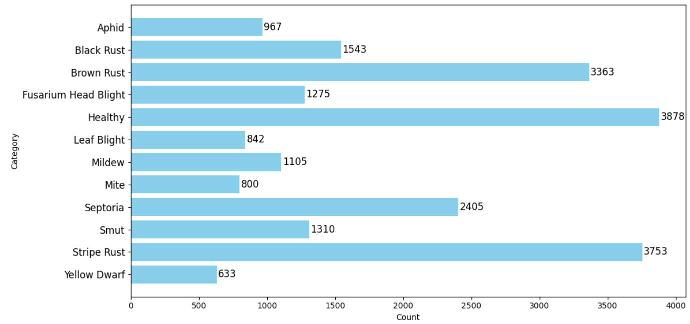

Fig. 2: Class Distribution of Difference Disease Classes

representation in the majority of their images. The images originate from diverse regions worldwide, covering a variety of wheat breeds, with a significant focus on South Asia. This is critical for addressing localized disease patterns, making the dataset particularly valuable for early intervention.

The second dataset [13] contains images of wheat leaves affected by two significant diseases, Stripe Rust and Septoria, along with healthy leaves. The dataset was collected from the Holeta wheat farm in Ethiopia under real, uncontrolled farm conditions, capturing the natural heterogeneity of the environment. Plant pathologists have validated the classification of the data set to ensure accuracy.

The third dataset [14] focuses on identifying wheat rust in images collected from Ethiopia and Tanzania. It includes 876 training images and 610 testing images, categorized into three classes: 'Healthy Wheat,' 'Leaf Rust,' and 'Stem Rust.' For images showing both types of rust, the classification is based on the rust that appears most prominently. The data was gathered mainly in-field by CIMMYT and its partners, with additional images sourced from Google Image.

The fourth dataset [15] focuses on detecting four major wheat diseases: Wheat Stripe Rust, Wheat Powdery Mildew, Wheat Yellow Dwarf, and Wheat Scab. These diseases pose a significant threat to wheat production worldwide. The images are collected from wheat fields in China, reflecting real-world farming conditions to support practical and reliable disease detection.

These four datasets are merged into a single comprehensive dataset through a Python script. The script is designed to merge similar dataset classes into unified classes, ensuring consistency across the combined dataset. The final dataset comprises a total of 12 classes, with representative images from each class displayed in Figure 1. The distribution of the number of images in each class is detailed in Figure 2. Following this, the data was split into three subsets under each class: 60% for testing, 20% for training, and 20% for validation. By combining the dataset, we ensure that the final dataset is well-structured for a reliable analysis.

IV. Methodology

The methodology for our study involves a systematic workflow in collecting and preprocessing data, training multiple models, and evaluating their performance. The workflow, illustrated in Figure 3, consists of six key stages.

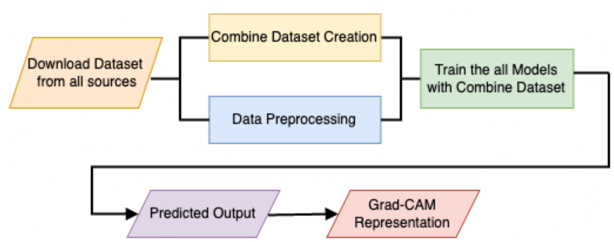

Fig. 3: Workflow

A. Data Preprocessing

The study utilizes a structured dataset organized into training, validation, and testing subsets, each within designated 'train', 'val', and 'test' subdirectories under the primary directory 'all'. This setup facilitates efficient data access and preprocessing.

1) Image Loading and Transformation: Images are processed using the `datasets.ImageFolder` class from the `torchvision` package, assigning labels according to their directory names. The transformations applied are:

- **Color Standardization:** Images are converted to RGB format to ensure uniformity across all channels.
- **Resizing:** Images are resized to 224×224 pixels to meet the input requirements of the model.
- **Array Conversion:** Images are converted into NumPy arrays to enhance data manipulation capabilities.

2) Data Stacking and Archival: The processed images and labels are arranged in arrays:

- **Image Stacking:** Image arrays are stacked into a single three-dimensional array.
- **Label Organization:** Labels are aligned into a one-dimensional array.

3) Integration with PyTorch: A custom `NumpyDataset` class is developed to work with PyTorch's data utilities, facilitating:

- Efficient data loading for both training and validation phases.
- Normalization and dimension adjustment of data through the `__getitem__` method, scaling pixel values to [0, 1], and conforming to PyTorch's (C, H, W) format.

B. Model Implementation

Several models are explored to determine which would perform best on our wheat disease classification task. The selection of each model was based on its architectural benefits and ability to effectively handle image data.

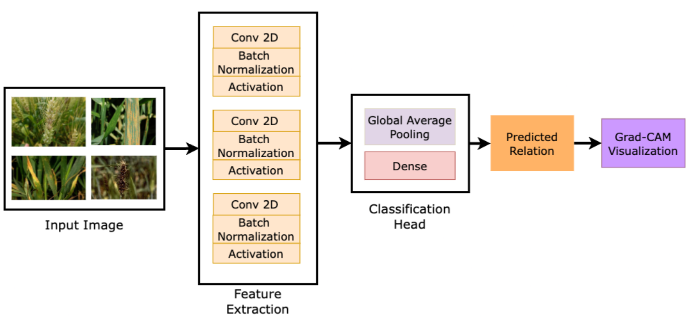

Fig. 4: Architecure of Custom Convolutional Neural Network

*1) **Custom CNN**:* To introduce non-linearity and stabilize training, the CNN model, shown in Figure 4, consists of a sequence of convolutional layers, each followed by batch normalization and ReLU activation functions. The architecture integrates an adaptive average pooling layer preceding the fully connected output layer, which effectively reduces feature dimensions while preserving essential information for classification.

*2) **MobileNetV2**:* This model is renowned for its efficient architecture, making it particularly appropriate for settings with constrained computational power, such as mobile devices.

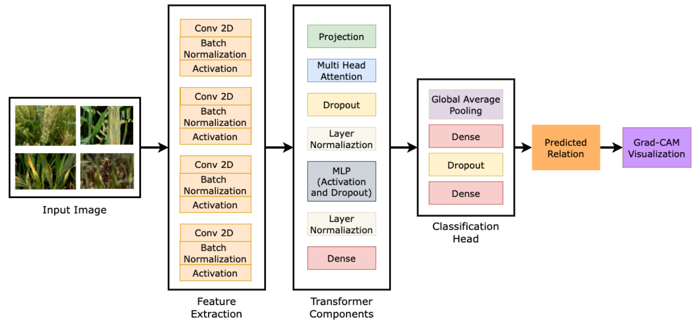

Fig. 5: Architecure of Custom Convolutional Vision Transformer

The depthwise separable convolutions are the main innovation of MobileNetV2. This method divides convolutions into depthwise and pointwise processes, where the former applies a single filter per input channel, and the latter combines the outputs using a 1x1 convolution. This architecture was incorporated into our research, utilizing its effective processing powers to minimize computational demands and sustain high performance, which makes it a perfect fit for real-time applications.

*3) **ResNet50**:* We employed ResNet50 due to its remarkable capacity to manage extremely deep networks by utilizing residual connections, which prevent the vanishing gradient issue and guarantee reliable feature learning. This architecture's proven effectiveness across diverse image recognition tasks offers significant reliability and accuracy for our complex datasets. Furthermore, ResNet50's residual blocks and deep layers are adept at extracting intricate characteristics from high-dimensional data, making it an optimal choice for achieving high performance in our wheat disease image classification objectives.

*4) **Vision Transformer**:* Leveraging the `timm` library, a pre-trained ViT model was adapted by setting the number of output classes to match our dataset. The ViT applies self-attention mechanisms that consider global dependencies within the image, potentially capturing more complex patterns than traditional CNNs.

*5) **Convolutional Vision Transformer**:* By combining the advantages of transformers and CNNs into a single architecture, the Convolutional Vision Transformer (CvT) model, illustrated in Figure 5, presents a novel strategy. In order to properly preprocess the input into detailed feature maps, which are subsequently fed into later transformer blocks, this integration begins with a sequence of convolutional layers. Through the use of self-attention mechanisms, these transformer blocks enable the model to comprehend global dependencies within the data by focusing on distinct portions of the input data in different layers. The CvT model is designed to work exceptionally well in tasks requiring nuanced comprehension of complex visual inputs by fusing localized feature extraction via convolutions with the global contextual capabilities of transformers.

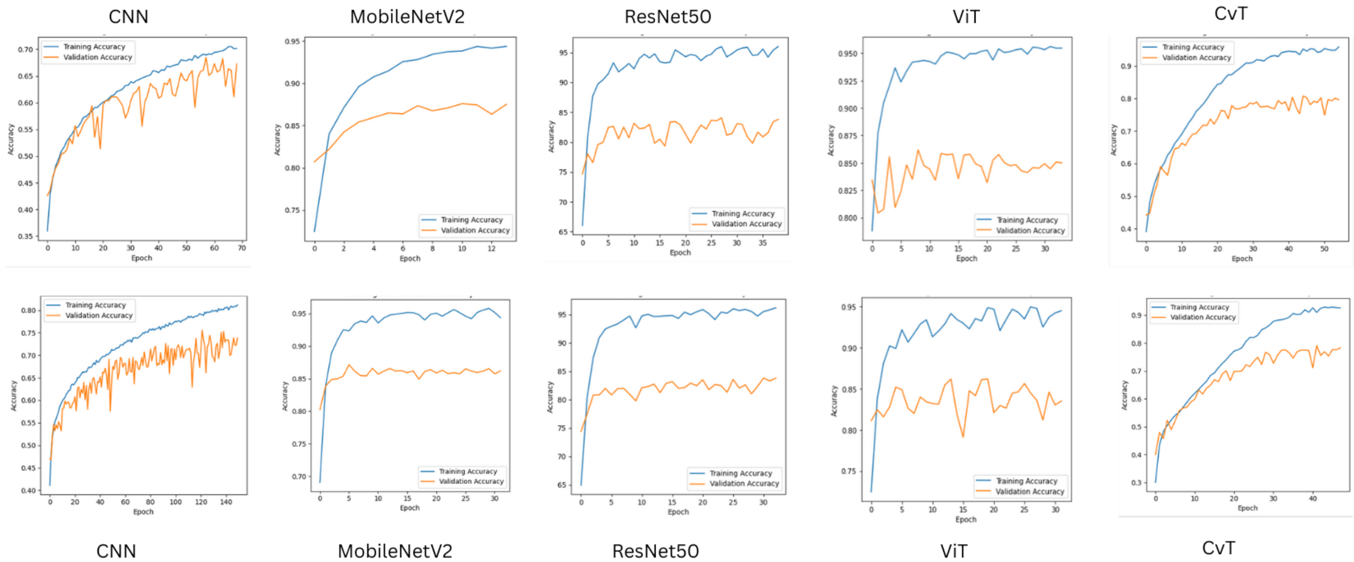

Fig. 6: Training and Validation Accuracy for All Applied Models: Without Class Balancing (Top), With Class Balancing(Bottom)

V. Experiments

A. Training Procedures

All models in our study, including the CNN, MobileNetV2, ResNet-50, CvT, and ViT, were trained with a constant learning rate of 0.0001 using the Adam optimizer. To address and compare the effects of class imbalance, experiments were conducted with and without class weights. The cross-entropy loss function was used as the main criterion for all models. By including class weights, the model becomes more sensitive to under-represented classes, resulting in a more fair learning environment across a variety of class distributions. All models had an early stopping mechanism in their training routine to prevent overfitting, and the patience parameter was selected based on non-improvement in validation loss performance. Each model underwent training for up to 150 epochs with a batch size of 32. This batch size was chosen to balance the computational load and model performance effectively. To guarantee robust generalization, validation was systematically performed for each training epoch for all models. A dedicated validation dataset was used to continuously track and assess the model's performance.

In our Wheat Disease Classification Using Machine Learning Techniques project, the training and validation loops were carefully designed to record accuracy metrics and losses, offering vital information on how well each model was learning. To secure the best state for further evaluations and deployment, models were saved and updated only when validation loss improved. This ensured that only the most effective model configurations were retained. Plots of accuracy and loss for training and validation were also used to visualize the training history, as depicted in Figure 6. Along with monitoring the models' performance over time, these visualizations helped detect any patterns that would point to possible problems like overfitting or underfitting.

This comprehensive methodology underscores our project's dedication to robustness and precision in applying diverse machine-learning architectures for the automated classification of wheat diseases. We made sure our models were accurate and generalizable by incorporating thorough validation procedures, which makes them ideal for real-world uses in agricultural monitoring and management.

B. Class Weights

Resolving class imbalance was an essential stage in our Wheat Disease Classification work to make sure that our models correctly identified and categorized diseases in all categories, even less common ones. To tackle this issue, we employed the `compute_class_weight` function from the `sklearn.utils.class_weight` library. The class weights determined by this function are inversely proportional to the input data's class frequencies. Here is the formula that is used:

$$\text{class\_weight}_i = \frac{\text{total\_samples}}{\text{number\_of\_classes} \times \text{frequency}_i} \tag{5}$$

where:

- class_weight_i is the weight for class i.
- total_samples is the total number of samples in the dataset.
- number_of_classes is the total number of unique classes in the dataset.
- frequency_i is the number of samples for class i.

In order to correct for imbalances, these weights are then applied to the loss function during model training. This modifies the weights assigned to each class throughout the optimization process, penalizing misclassifications of under-represented classes more severely than those of over-represented ones. Table II lists the precise values determined for the class weights in our study. This table offers a clear reference for comprehending the relative importance of each class, assisting in demonstrating how our models were directed to focus more on less common diseases, improving the efficacy and fairness of our classification approach.

TABLE II: Class Weights Used to Address Class Imbalance in Wheat Disease Classification

Class	Class Weight
Aphid	1.97299
Black Rust	1.23712
Brown Rust	**0.43199**
Fusarium Head Blight	1.49586
Healthy	0.49197
Leaf Blight	2.26601
Mildew	1.72600
Mite	2.38403
Septoria	0.79302
Smut	1.45589
Stripe Rust	0.50837
Yellow Dwarf	**3.18756**

VI. Result

This section highlights the results obtained from our experiments.

A. Normalized Confusion Matrix

We have normalized our confusion matrices as our dataset is imbalanced. Figure 7 depicts the confusion matrices of the different models we experimented with. All the matrices show high values along the diagonal, indicating that most predictions align with the true labels. We can see that a common misclassification occurred between Aphids and Mites because both are insect-based classes having similar visual features. Many models struggled with classifying Leaf Blight as it has a low presence in our dataset.

B. Test Accuracy

Table III presents the test accuracy of different models under two scenarios: (a) With Class Balancing and (b) Without Class Balancing. Among the models, MobileNetV2 achieves the highest accuracy with class balancing (0.8844), demonstrating its efficiency in handling imbalanced datasets. ViT also performs exceptionally well, with accuracies of 0.8781 (with balancing) and 0.8284 (without balancing), showcasing its robustness. ResNet50 follows, maintaining consistent performance across both conditions, while CvT and CNN exhibit moderate to lower accuracy, especially without class balancing (0.6766 and 0.6306, respectively). The results emphasize that advanced models like MobileNetV2 and ViT, combined with class balancing, significantly improve classification performance compared to simpler architectures like CNN.

TABLE III: Test Accuracy of Different Models

	Test Accuracy	
	With Class Weights	Without Class Weights
CNN	0.7606	0.6306
MobileNetV2	0.8844	0.8567
ResNet50	0.8306	0.8135
ViT	0.8781	0.8284
CvT	0.8024	0.6766

C. F-1 Score

In section VI-B, we have demonstrated the test accuracy of our experiments with different models in two scenarios. However, the F1 score is considered a better measure for imbalanced datasets because it provides a balanced evaluation of a model's performance by considering both precision and recall.

Table IV summarizes the F1 scores of different models for wheat disease classification under two conditions: With Class Balancing and Without Class Balancing, highlighting the higher score in bold for each comparison. Across all models, class balancing significantly improves performance for most disease classes, particularly for challenging ones like Aphid, Leaf Blight, and Mite. For example, CNN improves from 0.48 to 0.55 for "Aphid" and from 0.31 to 0.39 for "Leaf Blight" with class balancing. Advanced models like MobileNetV2, ResNet50, and Vision Transformers (ViT) exhibit consistently strong performance, with MobileNetV2 achieving an F1 score of 0.87 (weighted) with class balancing compared to 0.84 without. Similarly, ViT achieves the highest weighted F1 score of 0.88 with class balancing. In contrast, simpler models like CNN and Convolutional Vision Transformers (CvT) benefit more significantly from balancing, indicating that handling class imbalance is crucial for their optimal performance. These results underscore the effectiveness of class balancing in improving model performance, especially for underrepresented classes.

D. ROC Curve

The ROC curve in figure 8 illustrates the classification performance of the models for various classes. These models demonstrate excellent separability for classes such as Fusarium Head Blight, Septoria, and Yellow Dwarf, which achieve near-perfect AUC scores across different models. Similarly, high AUC values for Healthy, Mildew, and Smut reflect strong performance for these classes. However, slightly lower AUCs for classes like Black Rust and Leaf Blight suggest some difficulty in distinguishing these diseases, possibly due to overlapping features. These results confirm the model's ability to perform well across most classes.

E. Grad-CAM

In figure 9, Grad-CAM visualization highlights the areas of an image that contributed most to the model's predictions. For example, the Grad-CAM for a correctly predicted image for a "Healthy" class shows that the model focuses on the wheat's healthy regions. The heatmap emphasizes these

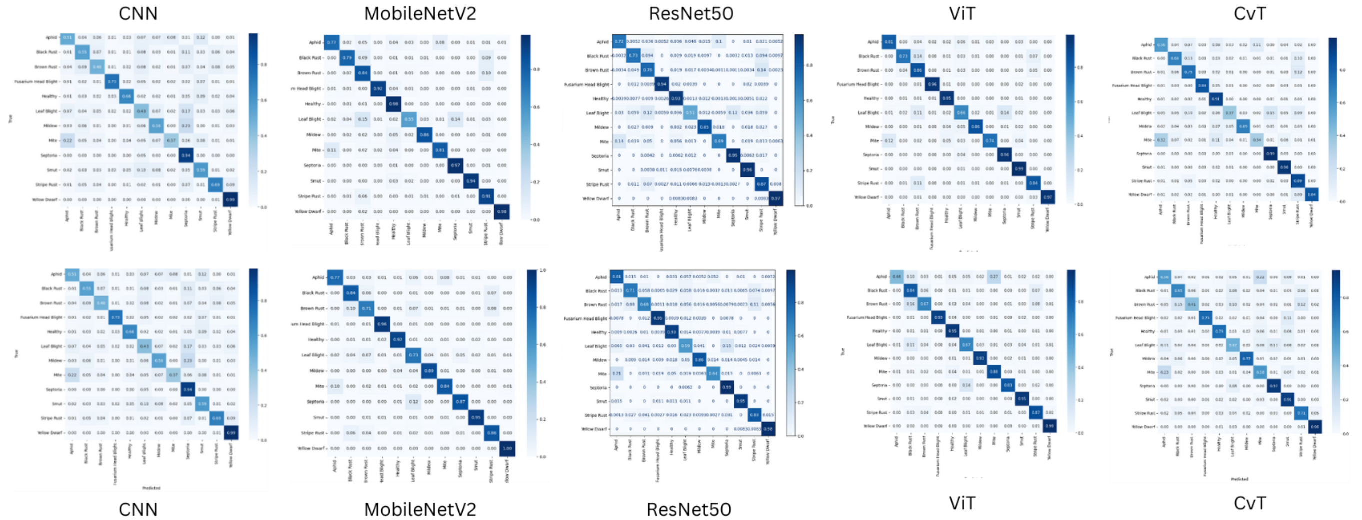

Fig. 7: Normalized Confusion Matrices for All Applied Models: Without Class Balancing (Top), With Class Balancing(Bottom)

TABLE IV: F1 Score of Different Models

	CNN		MobileNetV2		ResNet-50		ViT		CvT	
	With	Without	With	Without	With	Without	With	Without	With	Without
Aphid	**0.55**	0.48	**0.81**	0.77	**0.76**	0.73	**0.84**	0.6	**0.56**	0.49
Black Rust	**0.61**	0.51	**0.8**	0.75	**0.73**	0.7	**0.77**	0.66	**0.66**	0.54
Brown Rust	**0.75**	0.6	**0.85**	0.81	**0.8**	0.77	**0.84**	0.76	**0.79**	0.55
Fusarium Head Blight	**0.8**	0.77	**0.92**	**0.92**	**0.95**	0.94	**0.96**	0.95	**0.89**	0.78
Healthy	**0.85**	0.76	**0.94**	0.93	**0.92**	**0.92**	**0.95**	0.94	**0.88**	0.8
Leaf Blight	**0.39**	0.31	**0.61**	0.55	**0.56**	0.5	**0.6**	0.54	**0.45**	0.34
Mildew	**0.74**	0.53	**0.88**	0.84	**0.84**	**0.84**	0.89	**0.9**	**0.78**	0.74
Mite	**0.51**	0.44	**0.82**	0.8	**0.74**	0.72	**0.83**	0.79	0.41	**0.44**
Septoria	**0.89**	0.73	0.95	**0.98**	**0.95**	0.93	**0.95**	0.88	**0.93**	0.88
Smut	**0.62**	0.52	**0.95**	0.93	**0.92**	**0.94**	**0.93**	**0.93**	**0.9**	0.73
Stripe Rust	**0.84**	0.74	**0.9**	0.86	0.81	**0.83**	**0.87**	0.86	**0.83**	0.75
Yellow Dwarf	**0.91**	0.57	**0.98**	0.96	**0.94**	0.91	**0.98**	0.89	**0.9**	0.74
Weighted F1 Score	**0.76**	0.64	**0.87**	0.84	**0.84**	0.83	**0.88**	0.83	**0.8**	0.68

regions, reflecting the model's ability to identify relevant features for accurate classification. Another example of an incorrectly predicted image demonstrates a misclassification where the model predicted "Stripe Rust" instead of the actual class "Brown Rust." The Grad-CAM heatmap reveals that the model focused on parts of the image that may not contain distinctive features for "Brown Rust," explaining the error.

VII. Related Work

Researchers worldwide are working to provide guidance that helps farmers make informed decisions. During the past two decades, various approaches including conventional statistical methods, image processing, segmentation techniques, and machine learning have been applied to accurately recognize and classify wheat diseases.

Machine Learning-based Approach. Early machine-learning approaches for wheat disease classification relied solely on manual feature extraction techniques. These included RGB feature-based processing [16], Histogram of Oriented Gradient (HOG) with random forest classification [17], and Gray-level co-occurrence matrices (GLCM) with k-means clustering [18]. Despite achieving promising results, these approaches were limited by time-consuming feature extraction processes and offered poor scalability. H. Khan et al. [19] attempted to address these limitations by combining multiple feature extractors (HOG, LBP, HM, CH) with random forest classification, but the approach remained computationally intensive and struggled with environmental variations.

Deep Learning-based Approach. Following the emergence of Convolutional Neural Networks (CNNs), various deep-learning approaches have been successfully applied to wheat disease classification. R. Pryzant et al. [20] compared RNN and DNN models using hyperspectral images for wheat disease prediction, finding RNN achieved 6.7% higher accuracy. Sood and Singh [21] evaluated ResNet50 and VGG16 models to classify wheat rust from healthy samples using preprocessed image data. Similarly, Kumar and Kukreja [22] utilized Mask-RCNN with ResNet50 backbone to detect wheat mosaic virus, achieving 97% detection accuracy on 15,536 manually labelled images. While these models eliminated the need for manual feature extraction and achieved high

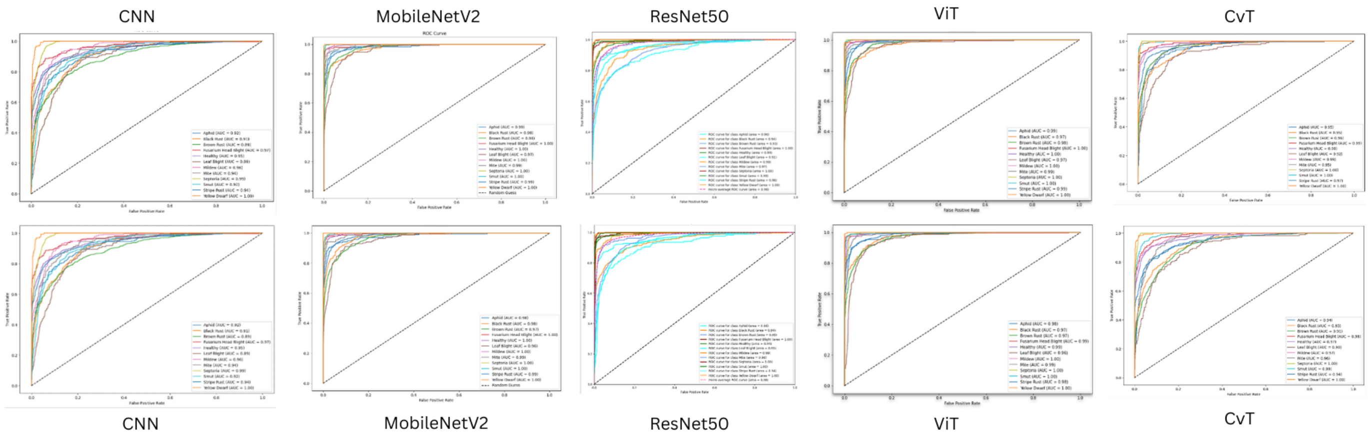

Fig. 8: ROC Curve for All Applied Models: Without Class Balancing (Top), With Class Balancing(Bottom)

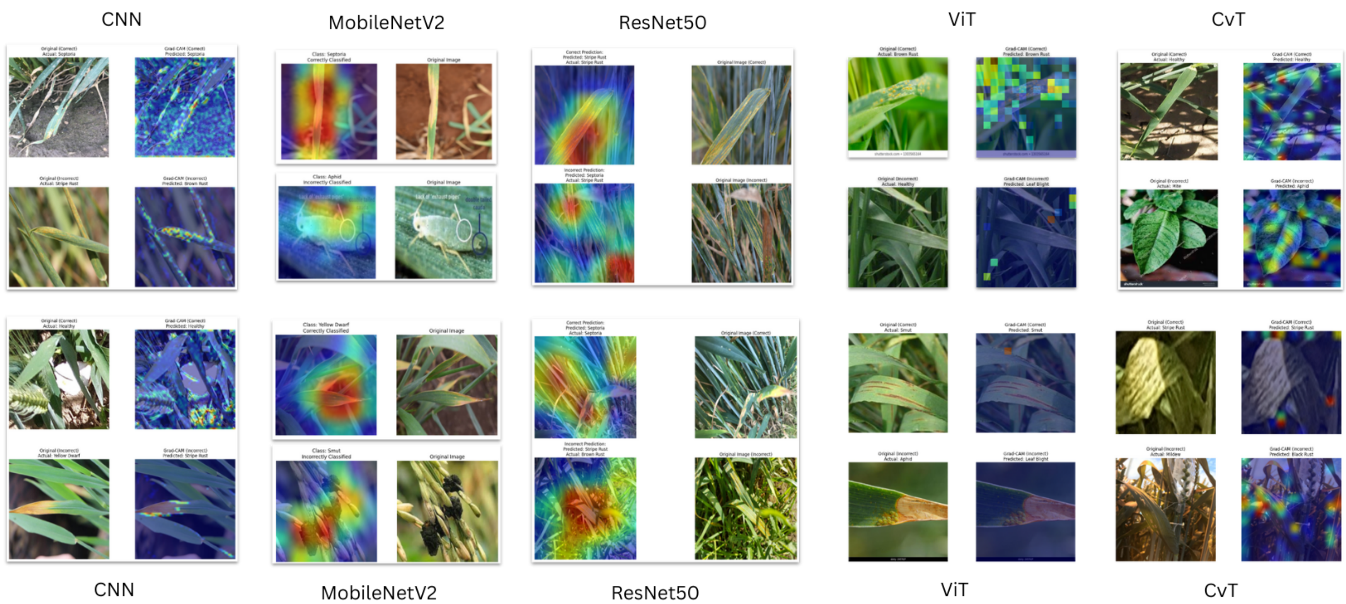

Fig. 9: GradCAM visualization for All Applied Models: Without Class Balancing (Top), With Class Balancing(Bottom)

accuracies, they face some critical limitations. The requirement for extensive training data and computational resources makes them impractical for rapid deployment in resource-constrained environments. Moreover, these models lack the flexibility to adapt to new disease outbreaks without complete retraining on the entire dataset when incorporating new disease classes.

Transfer Learning and Attention-based Approach. Recent transformer-based approaches have shown promising advancements in agricultural disease detection. Reedha et al. [23] demonstrated the effectiveness of Vision Transformer (ViT) in processing image patches using self-attention mechanisms. Leveraging augmentation and transfer learning, they achieved high performance despite limited data. Wu et al. [24] introduced a dual-scale approach combining different patch sizes, while Thai et al. [25] achieved 90% F1-score with an optimized ViT model for edge deployment. Yasamin et al. [26] explored CNN-ViT hybrid architectures to balance accuracy and speed. While ViTs excel at capturing global feature relationships compared to CNN's template matching approach, they require huge computational resources. However, combining pre-trained models with ViT architectures offers a promising direction for extracting comprehensive features while managing computational constraints.

VIII. Findings

In summary, our findings highlight the following key insights and outcomes of the research.

- We have seen high diagonal values in the confusion matrix indicate that most predictions align with the true labels.

- Common misclassification occurs between Aphids and Mites, likely due to their similar visual features as insect-based classes.
- Leaf Blight is often misclassified due to its low representation in the dataset.
- F1 scores improve significantly with class balancing across all models, particularly for challenging classes like Aphid, Leaf Blight, and Mite.
- MobileNetV2 achieves the highest accuracy with class balancing (88.44%) but has some inconsistency and unevenly distributed F1 scores.
- Vision Transformers (ViT) perform robustly, with accuracies of 87.81% (with balancing) and 82.84% (without balancing). Additionally, it showed consistency in the F1 scores, making it the best candidate in our study.
- ResNet50 shows consistent performance across both conditions, while simpler models like CNN and CvT perform better with class balancing but exhibit lower accuracies without it.
- Grad-CAM successfully identified areas of the image most influential for predictions, providing insights into the model's decision-making process.

IX. Conclusion

Our research addresses the challenges of wheat disease classification across multiple breeds and regions. This study proves that complex models don't always yield better results. Careful data preprocessing and smart model choices are key to creating accurate tools for agricultural diagnostics. These insights improve agricultural technology and offer a better framework for effective disease identification.

X. Division of Work

All group members have equally contributed to every phase of this project, including the proposal, dataset preparation, implementation, presentation, and report writing.

XI. Acknowledgment

We acknowledge the use of ChatGPT [27] and Grammarly [28] for proofreading, grammar checks, and editorial refinements in the preparation of this report.

References

[1] O. Erenstein, M. Jaleta, K. A. Mottaleb, K. Sonder, J. Donovan, and H.-J. Braun, "Global trends in wheat production, consumption and trade," in *Wheat improvement: food security in a changing climate*. Springer International Publishing Cham, 2022, pp. 47–66.

[2] S. Savary, L. Willocquet, S. J. Pethybridge, P. Esker, N. McRoberts, and A. Nelson, "The global burden of pathogens and pests on major food crops," *Nature Ecology & Evolution*, vol. 3, no. 3, pp. 430–439, 2019. [Online]. Available: https://doi.org/10.1038/s41559-018-0793-y

[3] H. R. Bukhari, R. Mumtaz, S. Inayat, U. Shafi, I. U. Haq, S. M. H. Zaidi, and M. Hafeez, "Assessing the impact of segmentation on wheat stripe rust disease classification using computer vision and deep learning," *IEEE Access*, vol. 9, pp. 164 986–165 004, 2021.

[4] V. Kukreja and D. Kumar, "Automatic classification of wheat rust diseases using deep convolutional neural networks," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2021, pp. 1–6.

[5] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation—a review," *Information Processing in Agriculture*, vol. 7, no. 1, pp. 1–19, 2020.

[6] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2022.

[7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[9] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[10] H. Wu, B. Xiao, N. Codella, M. Liu, X. Dai, L. Yuan, and L. Zhang, "Cvt: Introducing convolutions to vision transformers," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 22–31.

[11] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[12] Kaggle, "Wheat Disease Data Set," https://www.kaggle.com/datasets/freedomfighter1290/wheat-disease, [accessed on 12/03/2024].

[13] Wheat Leaf Dataset. https://www.kaggle.com/datasets/olyadgetch/wheat-leaf-dataset. [accessed on 12/03/2024].

[14] "ICLR Workshop Challenge: CGIAR Computer Vision for Crop Disease," https://zindi.africa/competitions/iclr-workshop-challenge-1-cgiar-computer-vision-for-crop-disease/data, [accessed on 12/03/2024].

[15] R. Mao, Y. Zhang, Z. Wang, X. Hao, T. Zhu, S. Gao, and X. Hu, "DAE-Mask: a novel deep-learning-based automatic detection model for in-field wheat diseases," *Precision Agriculture*, Dec. 2023. [Online]. Available: https://doi.org/10.1007/s11119-023-10093-x

[16] A. Badage, "Crop disease detection using machine learning: Indian agriculture," *Int. Res. J. Eng. Technol*, vol. 5, no. 9, pp. 866–869, 2018.

[17] S. Ramesh, R. Hebbar, M. Niveditha, R. Pooja, N. Shashank, P. Vinod *et al.*, "Plant disease detection using machine learning," in *2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C)*. IEEE, 2018, pp. 41–45.

[18] A. S. Tulshan and N. Raul, "Plant leaf disease detection using machine learning," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2019, pp. 1–6.

[19] H. Khan, I. U. Haq, M. Munsif, Mustaqeem, S. U. Khan, and M. Y. Lee, "Automated wheat diseases classification framework using advanced machine learning technique," *Agriculture*, vol. 12, no. 8, p. 1226, 2022.

[20] R. Pryzant, S. Ermon, and D. Lobell, "Monitoring ethiopian wheat fungus with satellite imagery and deep feature learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 39–47.

[21] S. Sood and H. Singh, "An implementation and analysis of deep learning models for the detection of wheat rust disease," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE, 2020, pp. 341–347.

[22] D. Kumar and V. Kukreja, "Image-based wheat mosaic virus detection with mask-rcnn model," in *2022 International Conference on Decision Aid Sciences and Applications (DASA)*. IEEE, 2022, pp. 178–182.

[23] R. Reedha, E. Dericquebourg, R. Canals, and A. Hafiane, "Transformer neural network for weed and crop classification of high resolution uav images," *Remote Sensing*, vol. 14, no. 3, p. 592, 2022.

[24] S. Wu, Y. Sun, and H. Huang, "Multi-granularity feature extraction based on vision transformer for tomato leaf disease recognition," in *2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST)*. IEEE, 2021, pp. 387–390.

[25] H.-T. Thai, N.-Y. Tran-Van, and K.-H. Le, "Artificial cognition for early leaf disease detection using vision transformers," in *2021 International Conference on Advanced Technologies for Communications (ATC)*. IEEE, 2021, pp. 33–38.

[26] Y. Borhani, J. Khoramdel, and E. Najafi, "A deep learning based approach for automated plant disease classification using vision transformer," *Scientific Reports*, vol. 12, no. 1, p. 11554, 2022.

[27] OpenAI, "ChatGPT [Large language model]," https://chatgpt.com/, [accessed on 12/03/2024].
[28] Grammarly, "Grammarly: Free AI Writing Assistance," https://app.grammarly.com/, [accessed on 12/03/2024].