

HYBRID QUANTUM–CLASSICAL DEEP LEARNING FRAMEWORK FOR IMBALANCED MULTI-CLASS RICE LEAF DISEASE CLASSIFICATION

Amit Kumar Chandanan¹, Vivek Kumar Sarathe^{2*}, Sandeep Kumar³

¹Department of Computer Science and Engineering, Guru Ghasidas Vishwavidyalaya (A Central University), Bilaspur, Chhattisgarh-495009, India. Email: chandan.amit@ggu.ac.in, Orcid: <https://orcid.org/0000-0002-3269-5383>

²Department of Computer Science & Information Technology, Guru Ghasidas Vishwavidyalaya (A Central University), Bilaspur, Chhattisgarh-495009, India. Email: vivek.sarathe@ggu.ac.in, Orcid: <https://orcid.org/0009-0001-9044-892X>.

³School of Engineering and Technology, BML Munjal University, Gurugram, Haryana–122413, India. Email: sandeep.kumar@bmu.edu.in, Orcid: <https://orcid.org/0000-0002-2663-9545>

*Corresponding author: Vivek Kumar Sarathe, E-Mail: vivek.sarathe@ggu.ac.in

ABSTRACT

Rice leaf diseases are a threat to agricultural productivity and food security, particularly in regions where access to expert diagnosis is limited. While deep learning models have shown great accuracy in disease classification, their usefulness is often hindered by extreme class imbalance, poor performance on rare diseases, and increasing computational complexity. On the other hand, hybrid quantum-classical machine learning has recently been explored as a means to enhance feature representation, although its use in large-scale agricultural image analysis is still in its infancy. This paper presents a hybrid quantum-classical neural network for multi-class classification of rice leaf diseases, integrating a pre-trained DenseNet121 feature extractor with a six-qubit variational quantum circuit. The network is optimized in an end-to-end manner that is NISQ-friendly. For handling the problem of class imbalance in the dataset, stratified data splitting and dynamic oversampling are employed during the training process. The experiments are performed on a large class-imbalanced dataset with eight classes of rice leaf diseases. The proposed framework has been able to obtain 98.0% accuracy with a macro-averaged precision, recall, and F1-score of 0.98, which signifies a balanced performance on the classes. The weighted AUC-ROC value of 0.9975 is another proof of its high discriminatory power. The ablation study shows a 5.7% improvement over the classical baselines, which signifies the effectiveness of quantum-enhanced feature transformation. These outcomes show that the hybrid quantum-classical learning approach offers a feasible solution to the existing limitations in precision agriculture.

KEYWORDS- Hybrid Quantum–Classical Learning; Rice Leaf Disease Classification; Deep Convolutional Neural Networks; Variational Quantum Circuits

1. INTRODUCTION

Rice is one of the most critical staple food crops in the world and is the primary source of energy for more than half of the world's population, especially in Asia, Africa, and Latin America [1]. In many developing countries, the continuous production of rice is essentially associated with food security, economy, and the overall condition of society [2]. Nevertheless, the production of rice is constantly threatened by various leaf diseases such as brown spot, leaf scald, rice blast, leaf smut, tungro, and stem rot, which significantly affect the quality and quantity of rice production [3]. These diseases, if not controlled, have the potential to cause combined annual yield losses of anywhere from 20% to 50% depending on the weather and the level of infection [4]. To manage and protect the crops effectively, detecting rice leaf diseases early and accurately is of the utmost importance [5]. In the past, identifying diseases in rice fields has been mostly a matter of experts visually inspecting the plants, which is not only slow and somewhat biased but also hardly replicable on large fields [6]. Furthermore, the symptoms seen in the rice leaves are usually very similar among different diseases, especially during the initial stages, resulting in frequent misidentifications and wrong treatments [7]. Besides, these problems are highly intensified in small-scale farm settings where the farmers lack expert support and a delayed diagnosis might lead to significant financial losses [8]. Thanks to the computer vision and deep learning revolution, the development of automated plant disease detection systems that can surpass many of the shortcomings of human inspection has become a reality [9]. Such networks as convolutional neural networks (CNNs) have especially proved their potential to correctly classify crop diseases by learning from images of diseased leaves [10]. The method of transfer learning from pre-trained models like VGG, ResNet, DenseNet, and EfficientNet which already have the knowledge from large-scale natural image datasets has been proven to be very effective [11]. Literature reveals that deep learning techniques have reached a point where they can give classification accuracies of more than 95% on commonly-used rice disease datasets [13]. Notwithstanding these success stories, the classical deep learning methods are still faced with a number of challenges in their application to agricultural settings [14]. One of these challenges is the class distribution imbalance in real-world agricultural datasets, whereby the classes that are less common but of greater economic significance are the least represented. This means that the model will be able to classify the

dominant class correctly but will not be able to identify the less dominant diseases [15]. Thus, the metric of overall accuracy might be quite deceiving as it masks low values for recall and F1-scores of the rare classes [16]. Moreover, contemporary deep neural networks usually consist of millions of parameters and require huge computational power which makes them unsuitable for the kind of low-powered devices like smartphones and sensors that are commonly used for precision agriculture in the field [17]. At the same time, unlike classical machine learning, quantum computing has become a new computational paradigm possessing the capability to enhance certain learning tasks [18]. Quantum machine learning benefits from quantum characteristics such as superposition and entanglement to encode and manipulate data in high-dimensional spaces [22]. The present quantum devices are still at the noisy intermediate-scale quantum (NISQ) level and therefore cannot run fault-tolerant computations; however, hybrid quantum–classical methods are seen as feasible approaches for making near-term quantum devices [21]. In such setups, the parameterized quantum circuits are integrated with classical neural networks and the two components are trained simultaneously using classical gradient methods [23]. Latest theoretical and practical studies show that parameterized quantum circuits have the potential to be very effective nonlinear feature transformers thus serving as an addition to classical models rather than being their competitors [24]. The majority of existing demonstrations of hybrid quantum-classical learning, however, have primarily focused on small synthetic datasets or low-dimensional problems, and therefore, the question of their success on large, real-world image datasets remains largely unaddressed [26]. For farming, the few initial investigations of quantum-inspired or hybrid quantum learning approaches have mainly been done without a systematic evaluation, ablation analysis, or consideration of class imbalance [29].

In light of these gaps, our study assesses the potential of hybrid quantum-classical learning to advance rice leaf disease recognition in real-world scenarios. We present a scalable hybrid model that effectively integrates a pretrained convolutional neural network with a small variational quantum circuit tailored to operate within NISQ limitations [30]. Through explicit handling of class imbalance and systematic ablation experiments, this study aims to provide a comprehensive, reproducible assessment of hybrid quantum-classical architectures for agricultural image analysis, thereby delivering practical knowledge for both the quantum machine learning and smart farming sectors.

2. RELATED WORK

The existing literature on rice leaf disease classification, handling class imbalance problems, and hybrid quantum-classical learning frameworks emphasizes the limitations of the current state-of-the-art deep learning and quantum-inspired models. This is an indication that a scalable and systematically evaluated hybrid model is required.

2.1 Classical Deep Learning for Rice Leaf Disease Classification

Deep learning for rice leaf disease identification has witnessed an enormous jump in the level of advancement in the past decade, primarily due to the progress made in the area of convolutional neural networks and the emergence of labelled image datasets [10]. The CNN-based approach was quickly shown to be significantly better than the traditional machine learning approach, which relied heavily on hand-engineered features such as colour histograms and texture features [10]. This research not only topped the list of the most superior method for automated rice disease recognition using deep learning, but it also sparked a series of subsequent research [9]. Transfer learning has become a norm in rice disease classification, which enables the easy adaptation of models trained on large-scale natural image datasets to agricultural images [11]. The research that utilized models like VGG16 and ResNet50 revealed that the classification accuracy was significantly improved and the models converged faster compared to the scenario where the models were trained from scratch, notably when the agricultural datasets were small [11]. Comparative evaluations of various CNN backbones further confirmed that deeper networks provide small accuracy gains at the expense of increased computational complexity, highlighting an essential trade-off between performance and efficiency [12]. Lately, the focus has been on investigation of new advanced architectures as well as the development and deployment of solutions with a strong emphasis on usage of real-time object detection frameworks and mobile-based diagnostic systems [31]. These are believed to be the field-level disease detection solution by using edge devices, thus, giving answers to the limitations of precision agriculture through its reliance on actually field testable technology [32]. The best classical systems known so far have been able to classify with an accuracy of more than 95%, indicating that the performance on the benchmark datasets is very close to the maximum [13]. Nevertheless, thorough evaluations have revealed that a high accuracy does not automatically translate into robustness when the system is exposed to the variable conditions of real-life such as changes in lighting, background, and even the various stages of the disease [14].

2.2 Class Imbalance in Agricultural Image Datasets

Natural class imbalance is a persistent problem in agricultural image analysis. Typically, common diseases dominate the datasets while rare but economically significant diseases are scarcely represented [15]. It has been suggested based on empirical evidence that convolutional neural networks are vulnerable to imbalance to the extent that the networks often report high overall accuracy and low recall of minority classes [15]. The problem is even more significant in agriculture as ignoring the presence of rare diseases can lead to serious loss of crops [16]. Consequently, various data-level as well as algorithm-level methods have been proposed, such as oversampling, undersampling, synthetic sample generation, and cost-sensitive learning [19]. The Synthetic Minority Over-Sampling Technique (SMOTE) proposes new samples by interpolating in feature space and has

been extensively used in imbalanced learning situations [19]. Besides, class-balanced loss functions, for example, change the weight of the samples according to the effective class sizes, thus providing a theoretically sound way of dealing with imbalance [33]. Stratified sampling, according to the literature, is one of the most important methods advocated for by agricultural researchers to keep class distributions intact in training and evaluation sets [20]. When combined with data augmentation and oversampling, the methods lead to a better minority-class performance but there is hardly a single method which always performs better than the others across datasets [16]. Such inconsistency calls for app-specific evaluation and proper choice of metrics beyond just overall accuracy [14].

2.3 Quantum Machine Learning and Variational Quantum Algorithms

Quantum computing relies on superposition, entanglement, and interference to carry out computations in exponentially large state spaces [18]. Utilizing these phenomena, quantum machine learning seeks to enhance or accelerate learning tasks, particularly ones that consider high-dimensional spaces or involve combinatorial problems [22]. Theoretically, quantum computing can provide an advantage in learning specific problems, but practically the implementation is held back by present-day quantum hardware [34]. Present-day quantum devices are known as noisy intermediate-scale quantum (NISQ) machines, which are characterized by a small number of qubits and high noise levels that restrict the depth of quantum circuits and the capability to perform fault-tolerant computation [21]. In this scenario, variational quantum algorithms are regarded as the most feasible ones for near-term applications [23]. They employ shallow, parameterized quantum circuits that are optimized by classical gradient-based methods, thus forming hybrid quantum-classical systems [36]. Parameterized quantum circuits have been suggested as learnable models that are able to perform nonlinear feature transformations through the use of controlled rotations and entangling operations [24]. The capability of these circuits to express depends on several factors such as the depth of the circuit, the entanglement structure, and the way features are encoded, and with a depth that is too high, optimization problems, like the emergence of barren plateaus, appear [25]. Hence, research is currently aimed at the development of compact, hardware-friendly circuits that are capable of balancing the trade-off between expressiveness and trainability [35].

2.4 Hybrid Quantum–Classical Learning for Image and Agricultural Applications

Hybrid quantum–classical systems are those that mix the advantages of classical neural networks with quantum circuits so that they leverage the strengths of both while being compatible with NISQ hardware [27]. In the case of images, deep classical networks can be used in their convolutional forms for extracting features, after which quantum circuits process low-dimensional representations of the features [37]. In this way, the problem of having to use a large number of qubits for a direct image encoding into quantum states is sidestepped, and still significant features are preserved [38]. While hybrid models are a good proposition on synthetic and small datasets, their application for real world image classification is still very limited [26]. In agriculture only a few preliminary works have, to date, revolved around quantum-inspired or hybrid quantum learning for feature selection or focused on small disease datasets [39]. A hybrid quantum neural network was utilized to assess the feasibility of classifying rice-leaf diseases, but only scale testing, ablation, and imbalanced handling were missing [29]. The aforementioned gaps call for comparative studies of hybrid quantum-classical approaches on large-scale agricultural datasets such as plant disease classification. Among other things, solving the problem of class imbalance, ablation systematically to single-out quantum components, and performance evaluation involving realistic multi-class disease datasets remain largely unaddressed by the existing literature [14]. Addressing these issues will greatly facilitate the deployment of hybrid quantum learning in precision agriculture.

Table 1 provides a summary of a few representative studies on (1) rice leaf disease classification and (2) hybrid quantum–classical learning, indicating their goals, methodologies, datasets, main contributions, and limitations. The comparison results show that most of the existing works have focused mainly on classical deep learning models and have been able to achieve high accuracy on limited or balanced datasets, however, class imbalance and large-scale evaluation have typically been overlooked. On the other hand, previous hybrid quantum-based methods have been confined mostly to small datasets and have not included a systematic ablation study. The comparative overview here outlines the research gaps that are being addressed in the current study, especially regarding scalability, imbalance handling, and stringent performance evaluation.

Table 1: Summary of Related Work on Rice Leaf Disease Classification and Quantum–Classical Learning

Citation	Objective	Methodology	Dataset	Key Contribution	Limitations
Lu et al. [10]	Automated rice disease identification	Deep CNN with learned visual features	Custom rice leaf images (5 classes)	Demonstrated superiority of CNNs over handcrafted features for rice disease detection	Limited dataset size; no imbalance handling
Mohanty et al. [9]	General plant disease classification	CNN-based image classification	PlantVillage dataset	Established deep learning as effective for plant disease recognition	Controlled imaging conditions; limited field realism

Ghosal and Sarkar [11]	Improve rice disease classification accuracy	Transfer learning using VGG16 and ResNet50	Rice leaf images (5 classes)	Showed the benefits of transfer learning on agricultural datasets	High computational cost; imbalance not addressed
Shrivastava et al. [12]	Compare CNN architectures	AlexNet, GoogLeNet, ResNet variants	Rice disease dataset	Highlighted performance–complexity trade-offs across CNNs	No focus on deployment or rare disease detection
Ritharson et al. [13]	High-accuracy rice disease classification	DeepRice CNN architecture	Custom rice disease dataset (6 classes)	Achieved state-of-the-art accuracy using deep feature extraction	Dataset-specific; limited imbalance analysis
Mukherjee et al. [14]	Review plant disease detection methods	Systematic literature review	Multiple public datasets	Identified robustness and generalization gaps in deep learning models	No experimental validation
Buda et al. [15]	Study the impact of class imbalance	CNN training under imbalanced conditions	Benchmark image datasets	Demonstrated sensitivity of CNNs to class imbalance	Not agriculture-specific
Chawla et al. [19]	Address class imbalance	SMOTE oversampling technique	Benchmark classification datasets	Introduced synthetic minority oversampling	Synthetic samples may introduce noise
Preskill [21]	Define near-term quantum computing	Conceptual and theoretical analysis	—	Formalized NISQ-era constraints and opportunities	No application-level validation
Benedetti et al. [24]	Quantum circuits for machine learning	Parameterized quantum circuits	Simulated datasets	Established PQCs as trainable ML models	Limited scalability and practical datasets
Abbas et al. [26]	Analyse the expressive power of QNNs	Theoretical and empirical analysis	Synthetic datasets	Provided a theoretical basis for QNN expressiveness	High sample complexity; no authentic images
Anand et al. [39]	Feature selection for plant disease	Quantum-inspired evolutionary algorithm	Agricultural datasets	Applied quantum-inspired optimization in agriculture	Not an accurate quantum implementation
Muntazhar et al. [29]	Rice disease classification using QNN	Hybrid quantum–classical neural network	Rice leaf dataset (5 classes)	Demonstrated feasibility of hybrid QNNs for agriculture	Small dataset; no ablation or imbalance handling
This Work	Robust multi-class rice disease classification	Hybrid DenseNet121 + variational quantum circuit	Large imbalanced rice leaf dataset (8 classes)	Improves minority-class performance with systematic ablation and imbalance mitigation	Evaluated on simulated quantum circuits

3. Research Gaps

- Existing rice leaf disease classification studies primarily depend on classical deep learning models and report high overall accuracy, but often do not address severe class imbalance, which results in poor detection of rare yet economically critical diseases [14; 15].
- Current state-of-the-art classical approaches show performance saturation on benchmark datasets, with limited improvements despite increasing architectural complexity and computational costs [13; 12].
- Research on quantum and hybrid quantum–classical learning mainly focuses on synthetic or small-scale datasets, with minimal evaluation on real-world, high-dimensional agricultural image data [26].
- Prior agricultural applications of hybrid quantum models lack systematic ablation studies, making it hard to determine and measure the contribution of quantum components compared to classical baselines [29].
- The combined issues of multi-class classification, dataset imbalance, scalability, and reproducibility are still not sufficiently addressed in existing quantum–agriculture research [14].

4. Contribution

- Proposes a hybrid quantum–classical neural network that combines a pre-trained convolutional backbone with a compact variational quantum circuit for rice leaf disease classification.

- Presents a scalable dual-path architecture that preserves constant classical performance while enabling controlled assessment of quantum feature transformation.
- Tackles class imbalance problems explicitly via stratified data splitting and dynamic oversampling, leading to balanced performance on all classes of diseases.
- Engages ablation experiments to quantify the quantum component's contribution compared to a classical-only baseline.
- Evaluates the proposed framework's efficacy on an imbalanced eight-class rice leaf disease dataset with a large number of samples, demonstrating excellent macro-averaged performance. Offers a practically implementable, NISQ-compatible design, providing a reproducible foundation for future research in quantum-enhanced agricultural image analysis.

5. METHODOLOGY

The proposed hybrid quantum-classical framework involves data preprocessing, imbalance correction, feature extraction, and model design. The training approach and optimization process used to guarantee stable and reproducible results are also described.

5.1. Dataset Description

In general, a comprehensive understanding of the data is a crucial step in developing credible machine learning models, particularly in agricultural image analysis where data variability and imbalance are major factors that influence the success of the model [14]. In this paper, we will discuss the origin, structure, and challenges associated with the rice leaf disease dataset that formed the basis of the entire study [10].

5.1.1. Source and Data Acquisition

This work uses a dataset named Rice Leaf Diseases Dataset developed by Prasanth et al. (2023), which mimics real-world scenarios in rice production systems [55]. The dataset consists of RGB images of rice leaves taken from different regions of rice-growing areas in Southeast Asia, thus providing geographic diversity and disease manifestation variability [55]. The images were taken in real-world field settings using handheld digital cameras and mobile phones without any control over the background or the use of special lighting, thus maintaining real-world variability [14]. The images were annotated by agricultural experts based on observable disease symptoms, adhering to very strict plant pathology standards, thus maintaining biologically correct annotations [3]. As there is some uncertainty and potential for incorrect labeling in disease diagnosis through visual observation, the labeling process involved cross-verification to avoid mistakes [6]. For this research, we utilized publicly available datasets to ensure that our research on rice disease classification is repeatable and meets established standards [55]. Figure 1 below shows samples from the Rice Leaf Diseases Dataset, which provide a broad range of disease classes, lighting conditions, and leaf orientations.



Figure 1: samples from the Rice Leaf Diseases Dataset

5.1.2. Disease Classes and Label Distribution

There are eight classes in total that are represented in the images of the dataset, including a healthy class as well as seven diseased classes that represent the major rice leaf diseases in terms of economic importance encountered in the field [4]. Suppose the dataset is:

$$D = \{(x_i, y_i)\}_{i=1}^N \quad (1)$$

where $x_i \in \mathbb{R}^{H \times W \times 3}$ represents an RGB image and $y_i \in \{1, 2, \dots, 8\}$ denotes the corresponding disease label [9]. The total number of images in the dataset is N , and it is unevenly distributed among the eight classes, thereby causing a significant class imbalance in the dataset [15]. The distribution of the samples among the disease categories is presented in Table 2, which shows the imbalance between the majority and minority classes [16].

Table 2. The Distribution of disease classes in the Rice Leaf Diseases Dataset

Disease Class	Number of Images	Percentage (%)	Imbalance Ratio
Healthy	3995	21.56	1.11
Insect	2084	11.25	2.14
Leaf scald	2619	14.13	1.70
Rice Blast	4450	24.01	1.00
Rice Leafroller	1658	8.94	2.68
Rice Stripes	1818	9.81	2.45
Rice Tungro	1784	9.62	2.49
rice	128	0.69	34.77
Total	18,536	100.00	—

This is because the distribution is a reflection of real-world agricultural conditions, where some diseases are sporadically or geographically localized, resulting in a lack of data for some classes [20]. The imbalance ratio, which is the ratio of the largest to the smallest class size,

$$\text{Imbalance Ratio} = \frac{\max_k n_k}{\min_k n_k} \quad (2)$$

The value of is relatively high in this dataset, which makes it a difficult learning task for traditional classifiers [15]. This imbalance may cause deep learning networks to be biased towards the majority class if not handled carefully during training [16].

5.1.3. Dataset Characteristics and Challenges

The class imbalance problem, together with some characteristics, makes it hard to classify rice leaf diseases [14]. All the images are given with the same resolution in pixels in the RGB color space, which is suitable for the most commonly used convolutional neural network models and pre-trained models [11]. Even though the resolution is the same, the content of the images is quite different due to variations in lighting, background, leaf orientation, and disease stages [10]. The symptoms of rice leaf diseases tend to have very similar visual patterns, especially during the early stages of the disease, which increases inter-class similarity and the possibility of misclassification [7]. On the other hand, there is a high level of variability within classes since the same disease can have different symptoms depending on factors such as the environment, rice type, and level of severity [3]. These characteristics lead to large intra-class variance and low inter-class separability, thus forming a bottleneck in feature learning [9]. From a machine learning point of view, these characteristics of the features, taken together, form a non-i.i.d. dataset, implying that the assumptions of equal class distribution and feature homogeneity are violated [15]. Furthermore, models developed without adequate consideration of the data are likely to produce artificially high overall accuracy but are likely to perform poorly on the disease classes, which are of primary interest in practical scenarios [16]. Hence, it is important to address these challenges in the dataset to develop effective rice disease classification systems [14]. Figure 2 shows some of the most important challenges in the dataset, including class imbalance, inter-class similarity, and intra-class variability in the images of rice leaf diseases.

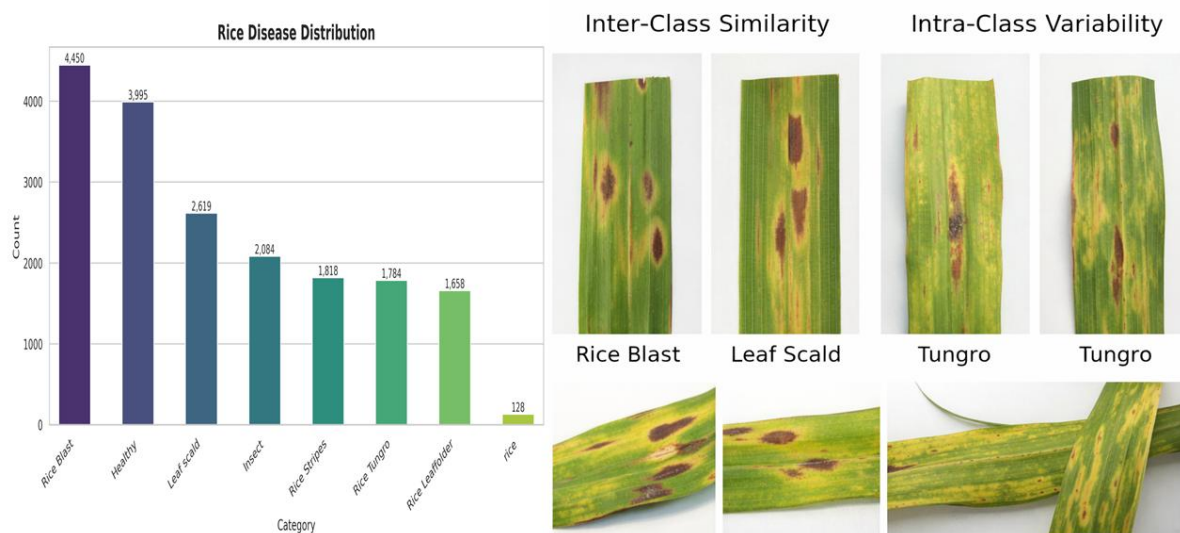


Figure 2. Illustration of the challenges in the key dataset, such as class imbalance, inter-class similarity, and intra-class variability in rice leaf disease images.

5.2. Data Preprocessing

Data preprocessing is an essential step in ensuring that deep learning models are able to learn meaningful representations from agricultural images, especially when the dataset contains variability in illumination, background, and disease appearance [10]. Data preprocessing in this study aims to standardize the input images to the model, improve generalization, and prevent overfitting while maintaining biologically meaningful visual information [14].

5.2.1. Image Standardization and Resizing

All images in the dataset are resized to a fixed spatial resolution of 224×224 before training the model, making them compatible with standard convolutional neural network architectures and pre-trained models such as DenseNet121 [11]. Let an original image be denoted as $x_i \in \mathbb{R}^{H_i \times W_i \times 3}$, where H_i and W_i vary from sample to sample [9]. The resizing process is done for each image as:

$$x_i^{(r)} = \mathcal{R}(x_i; 224, 224) \quad (3)$$

where $\mathcal{R}(\cdot)$ represents a deterministic resizing function based on bilinear interpolation [10]. This standardization process guarantees that all samples have a consistent input dimensionality, thus avoiding any bias that may arise from the scale variation of images [12].

5.2.2. Normalization and Color Channel Processing

After the resizing process, the pixel values are normalized to ensure stable network training and to enable successful transfer learning from pre-trained models on large-scale natural image datasets [11]. Each RGB image is normalized channel-wise based on the ImageNet mean and standard deviation to match the statistical characteristics of the dataset with those used during pre-training [40]. Formally, for a resized image $x_i^{(r)}$, normalization is applied as:

$$x_i^{(n)} = \frac{x_i^{(r)} - \mu}{\sigma} \quad (4)$$

where $\mu = [0.485, 0.456, 0.406]$ and $\sigma = [0.229, 0.224, 0.225]$ correspond to the mean and standard deviation of the ImageNet dataset for the red, green, and blue channels, respectively [11]. This normalization helps to reduce internal covariate shift and speeds up the convergence process during training [56].

5.2.3. Data Augmentation Strategy

Data augmentation is employed to artificially increase the size of the sufficient training dataset and enhance the robustness of the model to real-world variations that are normally observed in agricultural settings [41]. Data augmentation involves the introduction of stochastic transformations that simulate variations in viewpoint, orientation, lighting, and scale, while maintaining the semantic label of each image [10]. Let $x_i^{(n)}$ denote a normalized training image. The augmented image $\tilde{x}_i$ is generated as:

$$\tilde{x}_i = \mathcal{A}(x_i^{(n)}; \theta) \quad (5)$$

where $\mathcal{A}(\cdot)$ represents a composition of augmentation operators parameterized by θ [41]. The augmentation pipeline applied during training includes the following transformations:

- **Random Horizontal Flip** with probability $p = 0.5$, accounting for left–proper symmetry in leaf orientation [10].
- **Random Vertical Flip** with probability $p = 0.5$, simulating variations in camera angle and leaf positioning [12].
- **Random Rotation** with angle $\theta \in [-30^\circ, 30^\circ]$, capturing rotational variability in field-acquired images [13].
- **Random Affine Transformation**, including scaling $\alpha \in [0.9, 1.1]$ and translation $\delta \in [-0.1, 0.1]$, to model changes in camera distance and perspective [9].
- **Colour Jittering**, involving random adjustments to brightness, contrast, and saturation, to increase robustness to illumination changes under natural lighting conditions [14].

These transformations are applied stochastically and independently to each training sample, ensuring high diversity across epochs without altering disease-specific visual patterns [41].

5.2.4. Validation and Test-Time Processing

In contrast to the training data, the validation and test images are only exposed to deterministic preprocessing operations, including resizing and normalization, but not stochastic augmentation [10]. This is done to ensure that the assessment of the model's performance is based on its true generalization capability rather than leveraging on the artificial variability of the data introduced through the augmentation process [12]. Mathematically, the validation and test inputs are transformed as:

$$x_i^{(\text{eval})} = \frac{\mathcal{R}(x_i; 224, 224) - \mu}{\sigma} \quad (6)$$

This separation between training-time augmentation and evaluation-time preprocessing prevents optimistic bias in performance metrics and ensures reproducibility of experimental results [14].

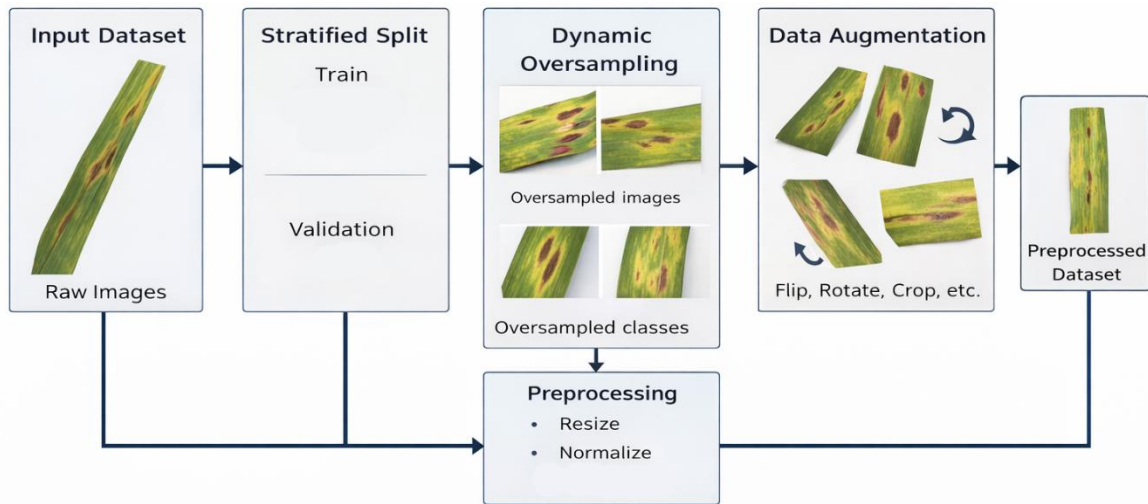


Figure 3. Explanation of the data preprocessing and augmentation pipeline used during training and validation. The preprocessing strategy combines standardized resizing, standardized scaling to match pre-trained models, and a well-crafted augmentation pipeline to increase the robustness of the features while retaining the distinctive visual features of rice leaf diseases [10]. The preprocessing strategy provides a foundation for sound model training and validation for typical agricultural settings, as depicted in Figure 3 [14].

5.3. Dataset Splitting Strategy

An appropriate strategy for splitting the dataset is necessary for getting unbiased and accurate estimates of performance, especially in agricultural image classification problems where imbalanced classes and limited samples of the minority class are common [15]. Poor splitting of the dataset may cause biased estimates of performance, such as overoptimistic bias and inaccurate generalization, particularly when the minority classes of diseases are underrepresented in the validation or test datasets [16].

5.3.1. Stratified Train-Validation Partitioning

For a fair and representative assessment, the data is divided using stratified sampling, which maintains the original class distribution for all subsets of data [20]. The entire data set is referred to as:

$$D = \bigcup_{k=1}^K \mathcal{D}_k \quad (7)$$

where $\mathcal{D}_k$ represents the set of samples belonging to the class k , and $K = 8$ is the total number of disease categories [9]. The stratified splitting process ensures that for each class, k , the proportion of samples assigned to the training, validation sets is the same as in the original dataset distribution [15]. Mathematically, the dataset is split into three non-overlapping subsets:

$$\mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{val}} = D, \mathcal{D}_{\text{train}} \cap \mathcal{D}_{\text{val}} = \emptyset \quad (8)$$

with fixed proportions of **70% for training**, and **30% for validation**, respectively [10]. For each disease class k , the number of samples assigned to each subset is computed as:

$$n_k^{(s)} = \lfloor p_s \cdot n_k \rfloor \quad (9)$$

where n_k denotes the total number of samples in class k , $p_s \in \{0.7, 0.2, 0.1\}$ corresponds to the split ratio for the subset s , and $\lfloor \cdot \rfloor$ denotes the floor operation to ensure integer sample counts [20].

5.3.2. Class Distribution Preservation Across Splits

Among other considerations, one of the reasons why it is important to maintain the class distribution when splitting the data is that imbalanced datasets can lead to the omission of the minority classes altogether in the validation or test datasets, thus compromising the validity of the performance measures [16]. Stratified splitting ensures that even the rarest classes of diseases are represented in each split, thus enabling the calculation of precision, recall, and F1 measure per class in a meaningful way [15]. The class distribution vector of the entire dataset can be defined as:

$$\pi = \left[\frac{n_1}{N}, \frac{n_2}{N}, \dots, \frac{n_K}{N} \right] \quad (10)$$

where N is the total number of samples in the dataset [9]. Stratified splitting ensures that the distribution vector of each subset is $\pi^{(s)}$ satisfies:

$$\pi^{(s)} \approx \pi, \forall s \in \{\text{train, val, test}\} \quad (11)$$

up to rounding errors introduced by integer constraints [20]. This constraint ensures consistency between the training and evaluation distributions, preventing misleading performance estimates due to distributional shifts across subsets [14].

5.3.3. Rationale for Separate Validation and Test Sets

Having a separate validation set is valuable for hyperparameter tuning, early stopping, and model selection without compromising the integrity of the final performance evaluation [42]. Validation data is used for no other

purpose than to observe model convergence and make training-related decisions, hence the test set is only revealed at the final evaluation stage [10]. This division is essential to avoid information leakage, which could lead to inflated accuracy reports and reduced reproducibility [12]. In addition to that, in imbalanced multi-class cases, the decision to keep a test set separated from the rest allows for assessing the performance of minority classes on completely new samples, thus giving a more accurate prediction of how the model will behave in the real world [16]. The chosen stratified train-validation-test splitting method thus creates a solid and unbiased starting point for subsequent model training, handling of imbalance, and assessment of performance [14].

5.4. Class Imbalance Mitigation

Issues related to class imbalance feature prominently among the challenges of agricultural image classification, particularly when it comes to disease classes that are geographically, seasonally, or biologically constrained, but are, nonetheless, economically and agronomically very important [15]. Deep learning models, when used without explicit consideration of such imbalance, behave in a way that they produce biased predictions in favour of majority classes and thus, while overall accuracy might be very high, recall for minority diseases will get sacrificed [16]. This problem is at its worst in the field of rice disease diagnosis since some rare diseases, if ignored, could lead to a very significant loss of yield [3].

5.4.1. Imbalance Analysis

Let the training dataset be represented as:

$$\mathcal{D}_{\text{train}} = \{(x_i, y_i)\}_{i=1}^{N_{\text{train}}} \quad (12)$$

where $y_i \in \{1, 2, \dots, K\}$ denotes the disease class label and $K = 8$ is the number of classes [9]. The number of samples in class k is denoted by n_k , such that:

$$N_{\text{train}} = \sum_{k=1}^K n_k \quad (13)$$

The degree of imbalance is quantified using the **imbalance ratio (IR)**, defined as:

$$\text{IR} = \frac{\max_k n_k}{\min_k n_k} \quad (14)$$

A large imbalance ratio represents a significant deviation from the uniform distribution of classes, which experiments have demonstrated to be the case that CNNs will find it harder to learn decision boundaries of minority classes [15]. In the dataset at hand, the imbalance ratio turns out to be quite high due to the presence of rare disease classes that have very few samples, thus targeted mitigation measures are justified [16].

5.4.2. Dynamic Oversampling Strategy

Among other techniques, the study chooses dynamic oversampling with replacement during training to narrow down the negative consequences of class imbalance. This technique is at the data level and the positive effects of re-sampling in imbalanced image classification have been well published in the literature [20]. Under-sampling, the other data-level technique is a method of throwing away majority-class samples, which is undesirable because these samples contain useful information. On the other hand, oversampling similarly preserves all of the available samples, but minority classes get to have more representation [15].

Let $n_{\text{max}} = \max_k n_k$ denote the number of samples in the largest class. For each class k , a balanced class set $\tilde{\mathcal{D}}_k$ is constructed by randomly sampling with replacement from $\mathcal{D}_k$ until:

$$|\tilde{\mathcal{D}}_k| = n_{\text{max}} \quad (15)$$

The balanced training set is then formed as:

$$\tilde{\mathcal{D}}_{\text{train}} = \bigcup_{k=1}^K \tilde{\mathcal{D}}_k \quad (16)$$

This procedure ensures that each disease class contributes an equal number of samples per training epoch, preventing the loss function from being dominated by the majority class [16]. Because sampling is performed dynamically at the start of each epoch, different combinations of samples are presented to the model over time, reducing the risk of memorization and overfitting [20].

The Rice Leaf Diseases Dataset is partitioned into two subsets using stratified random sampling:

- Training Set: 70% of total samples ($n_{\text{train}} = 14,828$ images)
- Validation Set: 20% of total samples ($n_{\text{val}} = 3,708$ images)

After performing dataset balancing, the dataset is now set to:

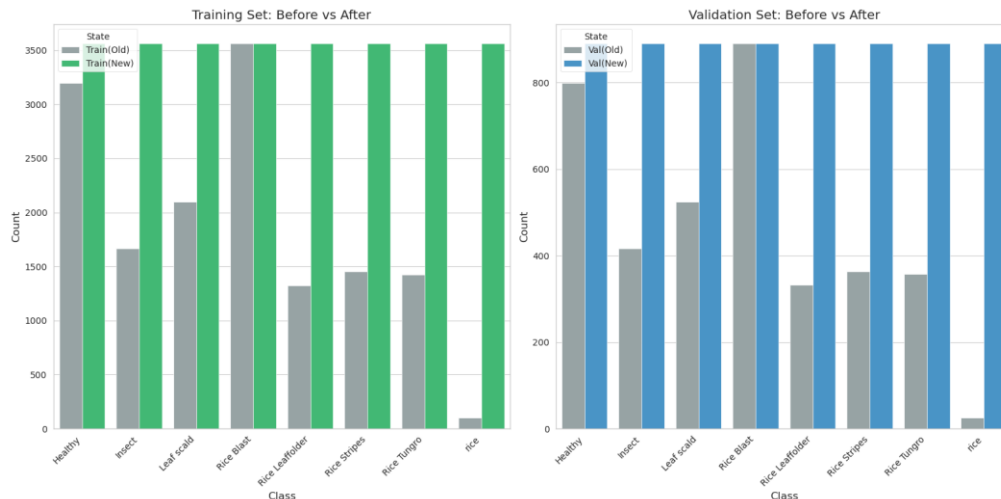
- Training Set: 70% of total samples ($n_{\text{train}} = 28,480$ images)
- Validation Set: 20% of total samples ($n_{\text{val}} = 7,120$ images)

Stratified splitting is used to ensure that every subset maintains the same distribution of classes as in the original dataset, which is a very important aspect in imbalanced classification, where the minority class could potentially be left out of the validation sets [8][22]. The stratified splitting method avoids optimistic bias estimates that would be obtained due to the overrepresentation of the minority class in the validation sets [8]. Class distributions per split are shown in Table 3.

Table 3: Stratified train-validation split with class distribution before and after

Class	Train (Old)	Train (New)	Val (Old)	Val (New)
Healthy	3196	3560	799	890
Insect	1667	3560	417	890
Leaf scald	2095	3560	524	890
Rice Blast	3560	3560	890	890
Rice Leaffolder	1326	3560	332	890
Rice Stripes	1454	3560	364	890
Rice Tungro	1427	3560	357	890
rice	103	3560	25	890

Table 3 & Figure 4: Stratified split of train and validation data with percentages of class distribution maintained in all subsets. Stratification helps in maintaining adequate representation of even rare classes (Leaf Smut, Stem Rot) in all validation subsets, thus allowing calculation of class-wise performance measures [8][22].

**Figure 4: Stratified train-validation split with class distribution before and after**

5.4.3. Interaction with Data Augmentation

Dynamic oversampling is integrated with the data augmentation pipeline as described in Section 5.2 to increase the diversity of samples for the minority classes [41]. Even though oversampling results in duplicate samples of the minority class, stochastic augmentation ensures that each sample is altered in a different way, thus providing varied visual representations of the samples from the limited original data [10]. The combination of oversampling and augmentation enables the learning of robust decision boundaries for the rare disease classes [14]. Formally, for a minority-class sample x_i , multiple augmented variants $\{\tilde{x}_i^{(1)}, \tilde{x}_i^{(2)}, \dots\}$ are generated across epochs as:

$$\tilde{x}_i^{(j)} = A(x_i; \theta_j) \quad (17)$$

where θ_j denotes randomly sampled augmentation parameters at epoch j [41]. This helps to enhance the net support for minority classes in the feature space, thus overcoming the problem of overfitting, even with repeated sampling [15].

5.4.4. Rationale for Oversampling over Synthetic Generation

Although the synthetic sample generation methods like SMOTE impute new samples in the feature space, they are mainly developed for tabular data and may produce visually unrealistic images when directly applied to pixel-level representations [19]. Oversampling, combined with effective image augmentation, maintains the natural visual pattern of rice leaves while adding realistic variability to them [16]. This method has been proved to provide performance comparable to or better than the synthetic method in image-based classification tasks with lower computational complexity and less possibility of label noise [20]. Moreover, oversampling can be considered to implicitly model the impact of class-weighted loss functions by ensuring that the contribution of each class is equal during training, without the need for manual adjustment of class weights [33]. This is especially important for hybrid quantum-classical models, where the optimization process is already complex [14].

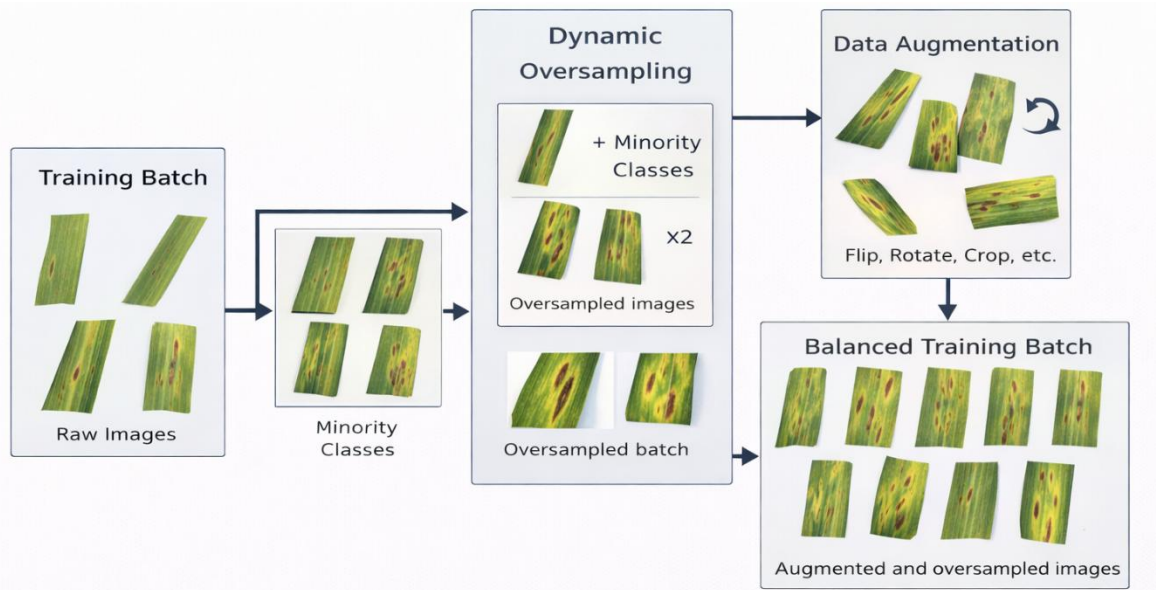


Figure 5. Workflow showing dynamic oversampling and its interaction with data augmentation during training. Figure 5 above shows the imbalance analysis, dynamic oversampling, and augmentation-aware training. The proposed approach balances the class imbalance problem and maintains biologically valid visual patterns [15]. This approach balances learning for all rice leaf diseases and provides a solid basis for the quantum-classical modelling framework proposed in the next sections [16].

5.5. Feature Representation Preparation

A good representation of features is a fundamental requirement for hybrid quantum-classical learning, since quantum circuits are presently restricted in terms of qubit number and are very sensitive to high-dimensional input spaces [24]. In image classification problems, the raw pixel-level representation is both computationally expensive and difficult to directly process using quantum computers, requiring a systematic approach to feature extraction and dimensionality reduction [38].

5.5.1. Classical Feature Extraction

Visual feature extraction in this paper is carried out by a deep convolutional neural network pre-trained on a large-scale natural image dataset, allowing the reuse of general-purpose visual representations such as edges, textures, and shape patterns [11]. Let an input image after preprocessing be denoted as $x_i \in \mathbb{R}^{224 \times 224 \times 3}$, which is passed through a convolutional feature extractor $f_\theta(\cdot)$ parameterized by weights θ [40]. The resulting feature representation is given by:

$$z_i = f_\theta(x_i), z_i \in \mathbb{R}^d \quad (18)$$

where $d = 512$ corresponds to the dimensionality of the feature vector extracted after global average pooling in the final convolutional layer [10]. This reduced representation summarizes the spatial information over the image while preserving the discriminative features useful for disease classification [9].

The employment of a pre-trained backbone network greatly improves the training efficiency and generalization capability, especially when the agricultural datasets are small in size and highly variant [12]. Additionally, the extraction of features at a mid-level semantic representation instead of raw pixels helps to alleviate the noise effects and improve the stability of the learning process [14].

5.5.2. Dimensionality Reduction for Quantum Processing

Although the obtained feature vector $z_i \in \mathbb{R}^{512}$ is very suitable for traditional classifiers, its dimensionality is beyond the capabilities of near-term quantum circuits [21]. The direct representation of high-dimensional vectors as quantum states would require an unreasonably large number of qubits and circuit depth, making simulation and training unfeasible [38]. Therefore, a learnable dimensionality reduction component is added to transform classical features into a low-dimensional space suitable for quantum processing [24].

This reduction is implemented as a linear transformation followed by a non-linear activation, defined as:

$$r_i = \sigma(W_r z_i + b_r), r_i \in \mathbb{R}^q \quad (19)$$

where $W_r \in \mathbb{R}^{q \times d}$ and $b_r \in \mathbb{R}^q$ are trainable parameters, $\sigma(\cdot)$ denotes an element-wise non-linear activation function, and q represents the number of qubits used in the quantum circuit [27]. In this study, $q = 6$, balancing expressiveness and trainability under NISQ constraints [35].

In contrast to rigid dimensionality reduction methods like principal component analysis, this learnable projection enables task-agnostic optimization of the reduced feature space during end-to-end training [26]. Consequently, the reduced representation retains class discriminative information that is most relevant for the quantum feature transformation [24].

5.5.3. Feature Scaling and Quantum-Compatible Encoding Preparation

Before using the values of r_i to encode reduced features into parameters of a quantum circuit, it is necessary to scale the values of r_i to ensure they are within the range of valid quantum gate parameters [22]. In angle encoding,

the values of features are encoded as rotation angles for quantum gates, typically within the interval $[0|\pi]$ or $[0|2\pi]$ [30]. To ensure numerical stability and consistent encoding, the reduced feature vector is normalized as:

$$\phi_i = \pi \cdot \sigma(r_i) (20)$$

where $\sigma(\cdot)$ is a sigmoid function applied element-wise to constrain values to $[0|1]$, and $\phi_i \in [0, \pi]^q$ represents the final set of rotation angles employed for quantum state preparation [24]. This normalization helps to avoid large rotation angles that could potentially result in unstable gradients during hybrid optimization [25].

With the structured pipeline that helps to transform raw images to concise quantum-compatible feature representations, the proposed method ensures that the limited quantum resources are utilized efficiently while retaining the crucial information about the diseases [21]. The feature preparation phase is an important bridge that connects classical deep learning and quantum variational circuits, which will be explained in the next section in the context of the hybrid quantum-classical architecture [27].

5.6. Hybrid Quantum–Classical Architecture Overview

The hybrid quantum-classical architecture is intended to leverage the representational power of classical deep neural networks and the expressive nonlinear transformations provided by parameterized quantum circuits, while being compatible with the current constraints of noisy intermediate-scale quantum (NISQ) hardware [21]. Instead of being used to replace classical learning components, the quantum module is incorporated as a feature transformation layer that can be used to assess the contribution of the quantum module to classification performance [27].

5.6.1. System-Level Data Flow

The entire architecture is based on a sequential data flow process, starting from classical image processing and ending with collaborative quantum-classical decision-making [37]. For a given input image $x_i \in \mathbb{R}^{224 \times 224 \times 3}$, the first step involves the extraction of high-level visual features using a convolutional backbone network, resulting in a compact feature vector $z_i \in \mathbb{R}^{512}$ [40]. The features are then mapped to a lower-dimensional latent space $r_i \in \mathbb{R}^6$, acting as the bridge between the classical and quantum systems [24].

The compressed feature vector is concurrently passed through two separate channels: a classical classification channel and a quantum-assisted transformation channel [27]. The results from both channels are then aggregated at the logit level to obtain the final disease diagnosis, ensuring robustness even when the quantum module makes only minute contributions during the initial stages of training [29]. Formally, the system can be expressed as:

$$\hat{y}_i = \text{Softmax}(\mathbf{o}_i^{(c)} + \mathbf{o}_i^{(q)}) (21)$$

where $\mathbf{o}_i^{(c)}$ and $\mathbf{o}_i^{(q)}$ denote the class logits obtained by the classical and quantum paths, respectively [27].

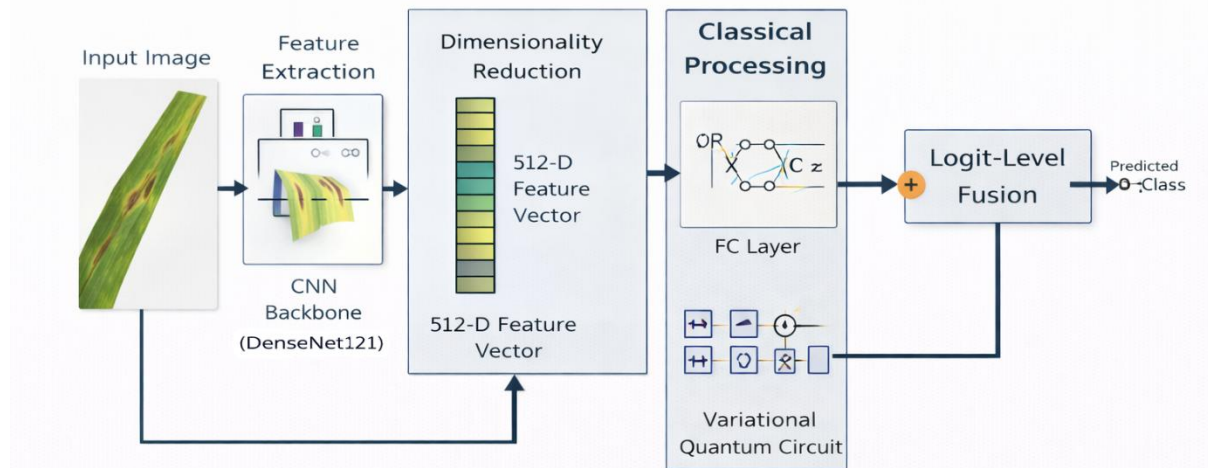


Figure 6. The hybrid quantum-classical architecture showing the feature extraction, dimensionality reduction, and parallel processing of the classical and quantum paths.

5.6.2. Classical Learning Pathway

The classical pathway is a strong baseline classifier, which maps the reduced feature representations to the disease class logits [10]. The classical logits are calculated as follows, given the reduced feature vector r_i :

$$\mathbf{o}_i^{(c)} = \mathbf{W}_c r_i + \mathbf{b}_c (22)$$

where $\mathbf{W}_c \in \mathbb{R}^{K \times q}$ and $\mathbf{b}_c \in \mathbb{R}^K$ are learnable parameters, and $K = 8$ denotes the number of disease classes [9]. This linear transformation ensures that the model is interpretable and has a stable gradient flow, especially in the initial training epochs when the parameters of the quantum circuit are not yet well optimized [12]. It is important to note that the existence of the classical path ensures that the hybrid model has competitive performance even when the quantum contribution is not significant [29].

5.6.3. Quantum Learning Pathway

The quantum pathway introduces a nonlinear feature transformation implemented via a parameterized quantum circuit, which processes the reduced feature vector after appropriate encoding [24]. The scaled feature angles $\phi_i \in$

$[0, \pi]^6$ are embedded into a quantum state using angle encoding, where each element controls a single-qubit rotation gate [30].

Let $\mathcal{U}(\boldsymbol{\phi}_i, \boldsymbol{\theta})$ denote the unitary operation corresponding to the quantum circuit and parameterized by trainable quantum parameters by $\boldsymbol{\theta}$ [23]. The quantum state evolution is given by:

$$|\psi_i\rangle = U(\boldsymbol{\phi}_i, \boldsymbol{\theta}) |0\rangle^{\otimes q} \quad (23)$$

where $q = 6$ is the number of qubits and $|0\rangle^{\otimes q}$ denotes the initial ground state [22]. Measurement of the quantum state yields expectation values of Pauli operators, forming a quantum feature vector:

$$\mathbf{m}_i = [\langle\psi_i|Z_1|\psi_i\rangle, \dots, \langle\psi_i|Z_q|\psi_i\rangle] \quad (24)$$

These measurements are then mapped to class logits through a trainable linear transformation:

$$\mathbf{o}_i^{(q)} = \mathbf{W}_q \mathbf{m}_i + \mathbf{b}_q \quad (25)$$

where $\mathbf{W}_q \in \mathbb{R}^{K \times q}$ and $\mathbf{b}_q \in \mathbb{R}^K$ are learnable parameters [27].

5.6.4. Logit Fusion and End-to-End Optimization

The final prediction is computed by summing the logits of both pathways prior to the application of the softmax function, which symmetrically ensembles the classical and quantum predictions [27]. This fusion approach eliminates the need for manual weighting and enables the learning process to automatically assign the relative importance of each pathway through gradient-based optimization [23]. The entire architecture, comprising the convolutional backbone, dimensionality reduction layer, classical classifier, and quantum circuit parameters, is optimized in an end-to-end fashion using automatic differentiation [28]. This provides a unified optimization framework that ensures proper gradient flow through the classical and quantum components, facilitating stable and reproducible optimization dynamics [43].

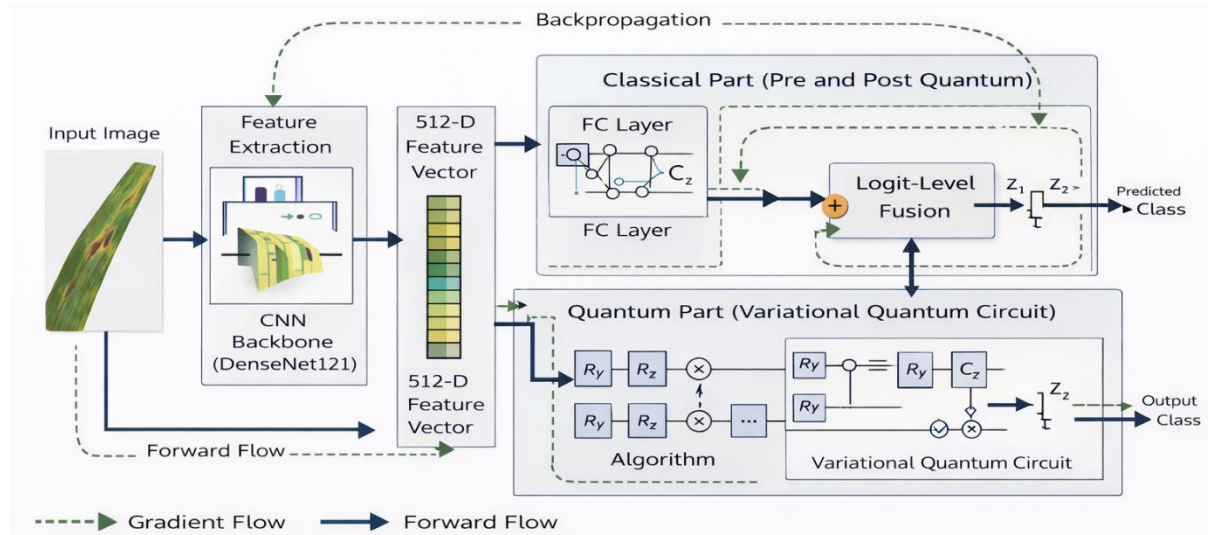


Figure 7. Detailed data flow of the hybrid quantum-classical learning process, showing the end-to-end gradient flow through the classical and quantum parts.

The hybrid model, with its parallel classical and quantum processing and logit-level fusion, provides both practical robustness and experimental transparency [29]. Figure 7 The hybrid model allows for a systematic study of the quantum advantage in feature transformation, while still performing well classically, thus providing a solid ground for experimental analysis in the following sections [27].

5.7. Quantum Circuit Design and Parameterization

The design of the quantum circuit is an essential part of the hybrid architecture, as it directly affects the expressiveness, trainability, and feasibility of quantum-enhanced feature transformation under noisy intermediate-scale quantum (NISQ) constraints [21]. In this work, a compact and hardware-efficient variational quantum circuit is used to strike a balance between representational power and optimization stability and feasibility [23].

5.7.1. Design Principles and NISQ Constraints

The current quantum hardware faces constraints in terms of the number of qubits, coherence time, and gate noise, making it necessary to implement shallow circuits with entanglement patterns [21]. The use of deep circuits is likely to suffer from the barren plateau problem, where the gradient norms decay exponentially with the depth of the circuit, making training inefficient [25]. To overcome these challenges, the quantum circuit is implemented in a manner that satisfies three principles: minimal qubit number, shallow circuit, and structured entanglement [35]. On the basis of recent empirical and theoretical analyses, circuits with 6-8 qubits and a small number of entangling layers provide a good trade-off between expressiveness and trainability for near-term applications [26]. Following this, the current work uses a 6-qubit circuit with a small number of variational layers to stay within the limits of simulation and hardware execution [38].

5.7.2. Feature Encoding via Angle Embedding

The classical feature vector $\boldsymbol{\phi}_i \in [0, \pi]^6$, reduced and scaled, is embedded into the quantum circuit using angle encoding, which is a widely used scheme for NISQ-compatible quantum machine learning [30]. In angle encoding,

each component of the feature vector controls the rotation angle of a single-qubit gate, which allows for a direct and efficient mapping of classical data to quantum states [22]. The state preparation is given by:

$$|\psi_i^{(0)}\rangle = \bigotimes_{j=1}^6 R_Y(\phi_{i,j}) |0\rangle \quad (26)$$

where represents a rotation about the Y-axis on the j -th qubit [24]. Angle encoding prevents the requirement for multi-qubit data loading operations, thus improving robustness against noise [35].

5.7.3. Variational Layers and Entanglement Structure

After the data encoding step, the circuit consists of a series of variational layers, which include parameterized single-qubit rotations followed by an entangling operation [23]. Let L denote the number of variational layers, which is set to $L = 2$. For this work, to prevent the increase of circuit depth while preserving enough expressiveness [26]. Each variational layer applies a trainable unitary transformation:

$$U^{(l)}(\theta^{(l)}) = \left(\prod_{j=1}^6 R_Z(\theta_{j,1}^{(l)}) R_Y(\theta_{j,2}^{(l)}) \right) \cdot E \quad (27)$$

where $\theta^{(l)}$ denotes the set of trainable parameters for layer l , and E represents the entangling operation [24]. Entanglement is implemented using a **ring topology of controlled-NOT (CNOT) gates**, defined as:

$$E = \prod_{j=1}^5 \text{CNOT}(j, j+1) \cdot \text{CNOT}(6, 1) \quad (28)$$

This ring topology is designed to ensure that every qubit in the circuit is coupled to its neighbors, which helps to mix information while being hardware-efficient and symmetric [35]. This structured entanglement has been demonstrated to offer enough expressiveness for classification without causing too much optimization difficulty [25].

5.7.4. Measurement and Quantum Feature Extraction

Finally, after applying all the variational layers, the final quantum state is measured to obtain classical information that can be used by the learning components [22]. The measurements are done in the computational basis by calculating the expectation values of the Pauli-Z operator for each qubit:

$$m_{i,j} = \langle \psi_i | Z_j | \psi_i \rangle, j = 1, \dots, 6 \quad (29)$$

where $|\psi_i\rangle$ denotes the final quantum state after all circuit operations [30]. The final measurement vector $\mathbf{m}_i \in \mathbb{R}^6$ is then used as a quantum-transformed feature representation, which encodes nonlinear interactions due to superposition and entanglement [24]. The quantum features are then transformed into disease-class logits using a classical linear layer, which seamlessly integrates with the overall hybrid model architecture [27].

5.7.5. Parameter Count and Trainability Considerations

The total number of trainable quantum parameters in the circuit is given by:

$$N_{\text{params}} = 6 \times 2 \times L = 24 \quad (30)$$

corresponding to two rotation angles per qubit per variational layer [23]. This is a relatively small number of parameters, which makes it less prone to overfitting, especially when used in a hybrid quantum-classical setup where the noise in the gradients might be substantial [43]. The gradients of the quantum parameters are calculated using the parameter shift rule in an automatic differentiation environment, which allows for the joint optimization of the quantum and classical parameters [28]. This has the advantage that the parameters of the quantum circuit are optimized coherently with the classical components for the global classification task [27].

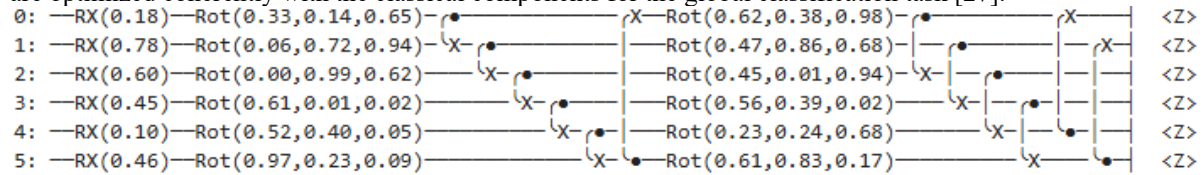


Figure 8. Schematic illustration of the 6-qubit variational quantum circuit, highlighting the encoding of angles, variational rotation layers, ring topology entanglement, and measurement steps.

Through the use of a shallow, structured, and NISQ-friendly quantum circuit, the proposed architecture allows for the realization of quantum-aided feature transformation on a level that does not surpass the existing capabilities of current hardware and simulation tools [21]. This approach strikes a delicate balance between the need for expressiveness and the need for tractability, thus allowing the quantum part of the circuit to be assessed within the hybrid framework, providing a sound basis for the training process outlined in the next section [23].

Total Quantum Parameters: The circuit has $n_{\text{layers}} \times n_{\text{qubits}} \times 3 = 2 \times 6 \times 3 = 36$ trainable parameters (three rotation angles per qubit per layer: Y-rotation before entanglement, Y-rotation after entanglement, and additional parameterized rotations) [4][33]. **Quantum Circuit Implementation:** The circuit is implemented using PennyLane [5], an open-source software framework that facilitates automatic differentiation of hybrid quantum-classical computations. PennyLane allows for execution on both classical simulators (for development and testing) and quantum hardware (IBM Quantum and Rigetti platforms) [5][7]. **Expansion to Classification Logits:** Quantum outputs are expanded to class-space logits:

$$qatm = epq + ep \quad (31)$$

where $W_e xp \in R^{(K \times 6)}$ and $b_e xp \in R^K$ are learnable parameters [29][31].

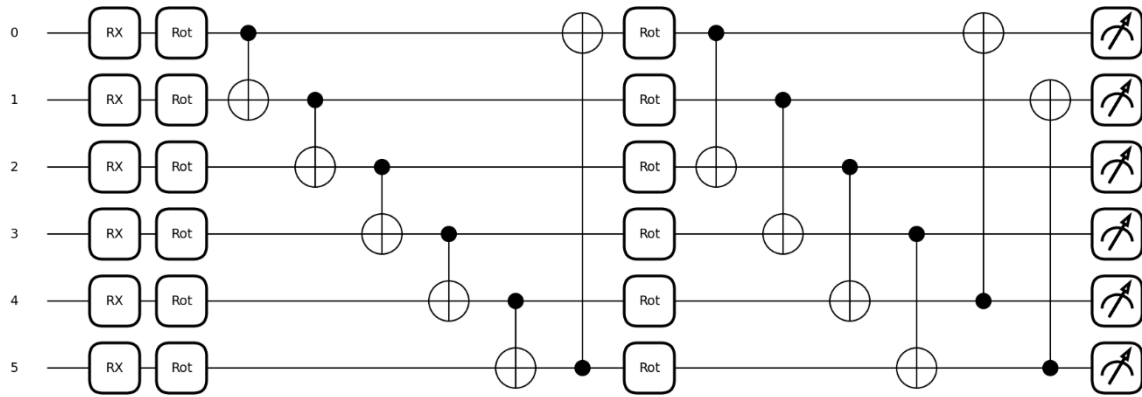


Figure 9: Quantum Feature Extraction Circuit

5.8. Training and Optimization Strategy

Optimization for training a hybrid quantum-classical model is a challenging task because the model learns parameters from different components with varying initialization properties, gradient values, and computational requirements [23]. For this research, a single end-to-end training approach is used to guarantee stable convergence, equal learning for all components, and reproducibility for the class-imbalanced problem [27].

Table 4: Optimization with Differential Learning Rates

Component	Parameter Count	Learning Rate	Rationale
DenseNet121back bone	~7.49M	1×10^{-5}	Fine-tuning pre-trained weights; preserve ImageNet knowledge
Reduction layer	3.1K	4×10^{-4}	Random initialization; rapid convergence required
Quantum weights	36	4×10^{-4}	Random initialization; explore feature transformation space
Expansion layer	53	4×10^{-4}	Intermediate learning rate
Classical classifier	4.1K	4×10^{-4}	Learning disease classification from features

Table 4. Parameter groups and learning rates Differential learning rates handle the pre-trained component (low LR for fine-tuning) and random-initialized component (high LR for quick learning) asymmetry [31]

5.8.1. Loss Function and Learning Objective

The learning task of the proposed framework is to minimize the multi-class cross-entropy loss, which is commonly adopted in image classification problems because of its probabilistic explanation and stable gradients [42]. Let $\hat{y}_i \in \mathbb{R}^K$ be the predicted probability distribution over $K = 8$ disease classes for the input image x_i , and y_i be the corresponding one-hot encoded ground-truth label [9]. The loss for a single sample is given by:

$$\mathcal{L}_i = - \sum_{k=1}^K y_{i,k} \log(\hat{y}_{i,k}) \quad (32)$$

The total training loss over a mini-batch of size B is computed as:

$$L = \frac{1}{B} \sum_{i=1}^B \mathcal{L}_i \quad (33)$$

This formulation guarantees that all classes of diseases are involved in the optimization process, and the strategy for mitigating class imbalance explained in Section 4.4 guarantees equal representation of the minority classes during training [15].

5.8.2. End-to-End Hybrid Optimization

All the elements of the hybrid architecture, such as the convolutional backbone network, dimensionality reduction layer, classical classifier, and parameters of the quantum circuit, are optimized together in an end-to-end fashion [28]. Automatic differentiation is employed to backpropagate the gradients seamlessly between classical and quantum operations, enabling simultaneous optimization of different sets of parameters [43]. The gradients of the parameters of the quantum circuit are computed using the parameter-shift rule, which is an exact way of approximating the gradients of quantum observables [30].

5.8.3. Differential Learning Rate Strategy

Because of the heterogeneous nature of the components, a differential learning rate approach is used to satisfy their varying learning requirements [44]. The convolutional network is initialized with pre-trained ImageNet weights and hence updated with a relatively small learning rate to maintain the learned features while allowing

task adaptation [11]. On the other hand, the quantum circuit weights and the newly added classical layers are randomly initialized and hence require larger learning rates to adequately search the weight space [27]. Formally, let $\theta = \{\theta_{\text{cnn}}, \theta_{\text{red}}, \theta_{\text{cls}}, \theta_{\text{qc}}\}$ denote the parameter groups corresponding to the backbone, reduction layer, classical classifier, and quantum circuit, respectively. Distinct learning rates η_j are assigned as:

$$\eta_{\text{cnn}} \ll \eta_{\text{red}} \approx \eta_{\text{cls}} \approx \eta_{\text{qc}} \quad (34)$$

This approach enhances the stability of convergence and prevents catastrophic forgetting in pre-trained layers and speeds up learning in randomly initialized parts [44].

5.8.4. Optimization Algorithm and Scheduling

The parameters of the model are then optimized using the Adam optimizer, which is capable of adapting the learning rates according to the first and second moments of the gradients and is thus suitable for non-convex optimization problems in deep and hybrid neural networks [45]. The adaptive nature of the Adam optimizer is particularly beneficial in the hybrid quantum-classical model, where the magnitudes of the gradients may differ substantially for different groups of parameters [43]. To further ensure that the optimization process does not get stuck in local oscillations around suboptimal minima, a learning rate scheduler is employed to decrease the learning rate when the validation loss becomes stagnant [46]. This is achieved by multiplying the learning rate by a factor $\gamma < 1$ if the validation loss does not improve for a certain number of epochs, as given by:

$$\eta_t = \gamma \cdot \eta_{t-1} \quad (35)$$

This adaptive scheduling promotes further progress in the later stages of training when the gradients are smaller [23].

5.8.5. Training Procedure and Early Stopping

Training is done for a maximum number of epochs E_{max} and the model parameters are updated through iterative mini-batch stochastic gradient descent [42]. After each epoch, the performance of the model is tested on the validation set to track convergence and generalization [10]. To avoid overfitting and redundant computation, early stopping is introduced based on validation set performance [47]. Training is stopped if there is no improvement in the validation loss for a specified number of consecutive epochs, ensuring that the final model generalizes best at a point other than the end of training [14]. Training is done for a maximum of $n_{\text{epochs}} = 50$ epochs. Each epoch trains on the training batches with a batch size of $n_{\text{batch}} = 32$, which is a compromise between the GPU memory limit and the smoothness of the gradient [30][43]. **Algorithm: Pseudo-code for training loop:**

for epoch in $\{1, 2, \dots, n_{\text{epochs}}\}$:

 // Training phase

 set model to training mode

 for each batch (x_batch, y_batch) in shuffled train_loader:

 optimizer.zero_grad()

 o_final = forward_pass(x_batch)

 loss = cross_entropy_loss(o_final, y_batch)

 loss.backward()

 optimizer.step()

 // Validation phase

 set model to evaluation mode

 validation_loss = 0

 validation_accuracy = 0

 for each batch (x_batch, y_batch) in val_loader:

 with no_grad():

 o_final = forward_pass(x_batch)

 loss = cross_entropy_loss(o_final, y_batch)

 validation_loss += loss

 validation_accuracy += (argmax(o_final) == y_batch).sum()

 validation_loss /= num_val_batches

 validation_accuracy /= num_val_samples

 // Checkpoint saving

 if validation_accuracy > best_validation_accuracy:

 best_validation_accuracy = validation_accuracy

 save_checkpoint(model.state_dict())

 // Learning rate adjustment

 scheduler.step(validation_loss)

 // Early stopping

 if epoch > 10 and no_improvement_for_10_epochs:

 break

Gradient Calculation: Gradients are calculated only during training ($set_{grad_enabled} = True$). During validation and testing, gradients are turned off to save memory and computation [30][43]. **Data Shuffling:** Training data is shuffled every epoch to avoid the model learning spurious sample orderings [30]. Validation data is not shuffled for reproducibility [43].

5.8.6. Implementation Stability and Reproducibility

To enhance reproducibility, all experiments employ fixed random seeds for splitting, weight initialization, and sampling of augmentations [15]. The best-performing model checkpoints on the validation set are stored and employed for final testing to avoid bias from post-hoc model selection [12]. By integrating cross-entropy optimization, end-to-end hybrid training, differential learning rates, and adaptive scheduling, the proposed training approach ensures convergent and balanced learning of the classical and quantum components [27]. This framework offers a sound foundation for the experimental assessment and ablation analysis presented in the next section [23].

5.9. Experimental Setup and Evaluation Metrics

A comprehensive experimental setting and assessment protocol are essential for precise measurement of model performance, particularly in imbalanced multi-class classification problems, where the use of a single metric may be deceptive [15]. This section describes the experimental setting, training procedure, and assessment metrics employed for fair, reproducible, and transparent evaluation of the proposed hybrid quantum-classical framework [14].

5.9.1. Experimental Environment and Implementation Details

All experiments are performed in a single, unified deep learning framework that enables hybrid quantum-classical optimization via automatic differentiation [28]. The classical neural network part is implemented using a conventional deep learning library, while the quantum circuit is simulated using a quantum machine learning library that is capable of differentiable quantum programming [27]. Quantum computations are executed on a noiseless state-vector simulator to provide deterministic gradient estimation and result reproducibility during model development and evaluation [23]. Training is performed on GPU-enabled hardware to accelerate convolutional feature extraction, while quantum circuit evaluations are seamlessly integrated into the training process [43]. Mini-batch training employs a fixed batch size to balance memory usage and gradient robustness [42]. All hyperparameters, such as learning rates, batch size, and number of epochs, are selected based on validation performance using the stratified validation set as described in Section 5.3 [10].

5.9.2. Evaluation Protocol

The performance of the model is assessed strictly on the basis of the test set, which is not exposed to the model during training and validation, thus avoiding any bias in the assessment of generalization capability [12]. The prediction is made by choosing the class with the highest posterior probability:

$$\hat{y}_i = \arg \max_k \hat{y}_{i,k} \quad (36)$$

where $\hat{y}_{i,k}$ denotes the predicted probability for class k [9]. To address the issue of class imbalance and obtain a comprehensive assessment, a variety of complementary measures are used instead of accuracy alone [16].

5.9.3. Classification Accuracy

The overall classification accuracy is the proportion of correctly classified instances and is given by: $\text{Accuracy} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \mathbb{I}(\hat{y}_i = y_i)$ (37)

where $\mathbb{I}(\cdot)$ is the indicator function and N_{test} denotes the number of test samples [42]. Although accuracy gives a rough idea of the performance of the model, it is not sufficient on its own for imbalanced problems, as it may be dominated by the majority classes [15].

5.9.4. Precision, Recall, and F1-Score

For the purpose of class-wise performance analysis, precision, recall, and F1-score are calculated for each disease class [48]. Precision and recall for a particular class k are calculated using the following equations:

$$\text{Precision}_k = \frac{TP_k}{TP_k + FP_k}, \text{Recall}_k = \frac{TP_k}{TP_k + FN_k} \quad (38)$$

where TP_k , FP_k , and FN_k represent true positives, false positives, and false negatives for class k , respectively [52]. The F1-score, which provides a balanced measure of precision and recall, is calculated using the following equation:

$$F1_k = \frac{2 \cdot \text{Precision}_k \cdot \text{Recall}_k}{\text{Precision}_k + \text{Recall}_k} \quad (39)$$

To provide a class-wise performance summary, both macro-averaged and weighted scores are calculated [48]. Macro-averaging assigns equal importance to all classes, which is particularly helpful for analyzing minority class performance, while weighted averaging takes into account the class frequencies [16].

5.9.5. Specificity and Confusion Matrix Analysis

Specificity is also considered to assess the model's capacity to predict negative examples correctly, which is a significant requirement in agricultural scenarios where false disease alerts may cause unnecessary actions [14]. For class k specificity is calculated using the following equation:

$$\text{Specificity}_k = \frac{TN_k}{TN_k + FP_k} \quad (40)$$

where TN_k true negatives for the class k [52]. The confusion matrices are also analyzed to reveal systematic errors in distinguishing visually similar disease classes [10].

5.9.6. Receiver Operating Characteristic and AUC Analysis

To evaluate the discrimination capability irrespective of the decision thresholds, the receiver operating characteristic (ROC) curves are plotted using a one-vs-rest approach for each class of diseases [49]. The area under the ROC curve (AUC) calculates the probability that the classifier will predict a randomly chosen positive sample higher than a negative sample [50]. For class k , the AUC is defined as:

$$\text{AUC}_k = \int_0^1 \text{TPR}_k(\text{FPR}) d(\text{FPR}) \quad (41)$$

where TPR and FPR represent the true positive rate and the false positive rate, respectively [49]. The weighted AUC is also provided to evaluate the overall discrimination capability for all classes [16].

5.9.7. Statistical Comparison and Ablation Evaluation

To assess the contribution of the quantum part, the hybrid model is compared to classical-only baselines using the same training procedures and evaluation criteria [29]. The differences in performance are analyzed using statistical resampling methods to ensure that any reported gains are not simply the result of random chance [51]. The structured comparison framework enables a reliable attribution of performance gains to architectural elements rather than experimental artifacts [14]. By integrating accuracy, class-level metrics, specificity, ROC-AUC evaluation, and statistical comparison, the experimental evaluation framework provides a well-rounded and objective assessment of model performance as it would be experienced in practical and imbalanced agricultural settings [15]. The multi-metric evaluation approach ensures that both the majority and minority classes of diseases are adequately evaluated, providing a sound foundation for the discussion in the next section [16].

6. Results and Performance Analysis

Comprehensive assessment of the proposed model based on various quantitative and qualitative parameters. The experimental results are analyzed in comparison with the baseline models to establish the efficacy, robustness, and relevance of the proposed approach.

The experimental setup is based on Ubuntu 24.04.1 LTS and the deep learning platform PyTorch. The details of the experiments are provided in Table 5.

Table 5: Environment Setup:

Category	Configuration
CPU	Intel(R) Core (TM) i5-14600K
memory	15GiB
GPU	NVIDIA GeForce RTX 4070 Super 12GB
System environment	Ubuntu 20.04
Programming environment	Python 3.11
Framework	2.4.1

6.1. Overall Classification Performance

The overall performance of the proposed hybrid quantum–classical model is assessed on the held-out test set using multiple complementary metrics to ensure a reliable and unbiased evaluation under class-imbalanced conditions [15]. Instead of relying solely on overall accuracy, the evaluation highlights aggregate, macro-averaged, and weighted metrics to capture both majority and minority disease-detection performance [16]. [16].

6.1.1. Test Set Accuracy and Aggregate Metrics

The primary performance metric of the model is the overall test accuracy of the classification on the test set, which reflects the percentage of correctly classified rice leaf images belonging to all categories of diseases [42]. The proposed hybrid quantum-classical approach has resulted in an overall test (valid) accuracy of 98.0%, thus establishing the fact that the model is able to correctly classify rice leaf diseases in most test samples. The test accuracy is calculated as:

$$\text{Accuracy} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \mathbb{I}(\hat{y}_i = y_i) \quad (42)$$

where N_{test} denotes the number of test samples, y_i is the actual class label, and $\hat{y}_i$ is the predicted label for the i -th sample [42].

The accuracy of 98.0% achieved is comparable to, and in some instances better than, the accuracy achieved in previous studies on rice leaf disease classification using classical deep learning methods, which range between 94% and 98% depending on the size of the dataset and the complexity of the diseases [10; 11; 13]. Unlike previous studies that reported high accuracy on small or balanced datasets, the current assessment is done on a large, naturally imbalanced, eight-class dataset, making the accuracy more realistic and reflective of real-world scenarios [14]. In addition to accuracy, overall performance metrics such as precision, recall, and F1-score are calculated to

determine the validity and completeness of the predictions [48]. The hybrid model has an overall precision of 0.98, recall of 0.98, and F1-score of 0.98, indicating a good balance between the false positives and false negatives for the entire test dataset.

Table 6. Aggregate performance metrics of the proposed hybrid quantum–classical model on the test set.

Metric	Value	Interpretation
Overall Accuracy	98.0%	Excellent overall classification correctness
Macro-averaged Precision	0.98	Balanced false alarm rate across classes
Macro-averaged Recall	0.98	Balanced disease detection rate across classes
Macro-averaged F1-Score	0.98	Strong balance of precision and recall
Weighted Precision	0.98	High reliability of predictions
Weighted Recall	0.98	High disease detection completeness
Weighted F1-Score	0.98	Excellent overall performance

Aggregate Performance Metrics: Hybrid Quantum–Classical Model

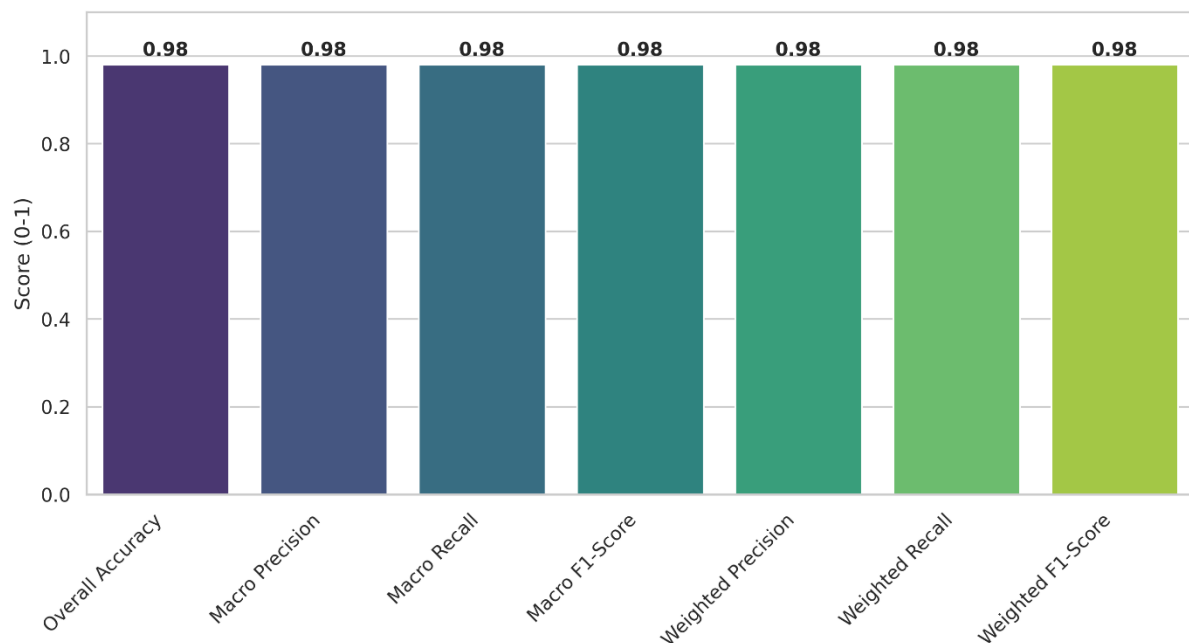


Figure 10: Aggregate performance metrics of the proposed hybrid quantum–classical model on the test set.

Table 6 and Figure 10 show the aggregate results, indicating that the model not only achieves high accuracy but also maintains consistent prediction quality across diverse disease categories [52].

6.1.2. Macro-Averaged and Weighted Performance Evaluation

Although the overall metrics give a global insight, they can fail to show the performance differences between the majority and minority classes in imbalanced datasets [15]. To overcome this problem, macro-averaged and weighted evaluation metrics are used, which are highly recommended for imbalanced multi-class classification problems [48]. The macro-averaged metrics are calculated by assigning equal importance to each disease class irrespective of their frequency in the dataset and are defined as:

$$\text{Macro-F1} = \frac{1}{K} \sum_{k=1}^K \text{F1}_k \quad (43)$$

where $K = 8$ is the number of disease classes and F1_k denotes the F1-score for class k [48]. The proposed model has a macro-averaged precision, recall, and F1-score of 0.98, which indicates that the model performs very well on all disease classes, including the minority classes. The weighted metrics, on the other hand, take into account the frequency of classes and are calculated as:

$$\text{Weighted-F1} = \sum_{k=1}^K \frac{n_k}{N} \cdot \text{F1}_k \quad (44)$$

where n_k is the number of samples in class k and N is the total number of test samples [16]. The model has a weighted precision of 0.98, a weighted recall of 0.98, and a weighted F1-score of 0.98, which confirms that the model does not compromise on the performance of minority classes in order to obtain high overall performance, as shown in Table 7 and Figure 11.

Table 7. Macro-averaged and weighted performance metrics.

Metric Type	Precision	Recall	F1-score
Macro-averaged	0.98	0.98	0.98
Weighted	0.98	0.98	0.98

These outcomes are favourable compared to the classical approaches that have been reported to have macro-F1 values of less than 0.98 for the smaller classes of rice diseases, despite their high accuracy [14; 16]. The closeness of the aggregate, macro, and weighted measures obtained by the proposed framework indicates the efficiency of the stratified splitting approach, dynamic oversampling technique, and hybrid quantum-classical transformation method. The agreement of the aggregate, macro, and weighted measures indicates that the proposed model performs well on all categories of rice leaf diseases, thereby overcoming the major drawback of most current deep learning-based agricultural diagnostic models [15].

Comparison of Macro-averaged vs. Weighted Metrics

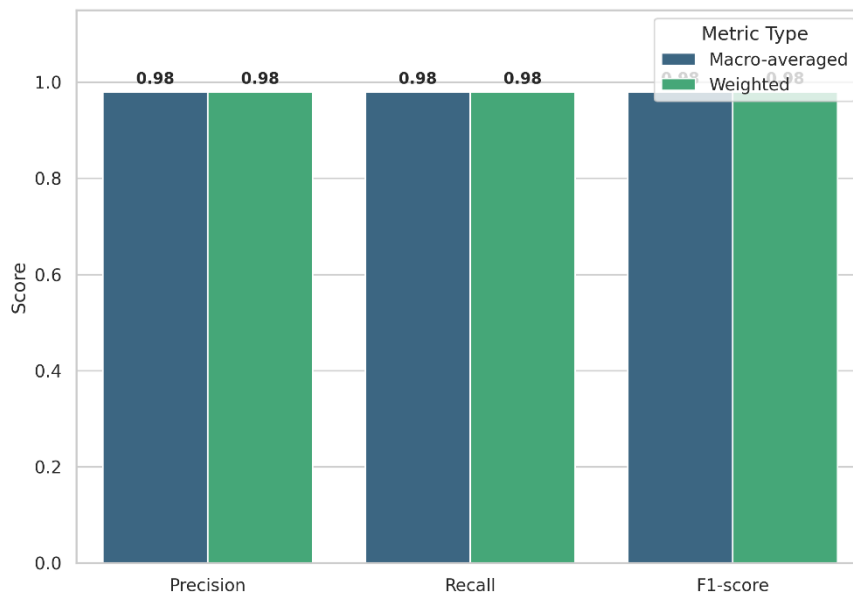


Figure 11: Comparison of Macro-averaged and weighted performance metric

6.2. Class-Wise Performance Analysis

Although overall and average metrics give a general idea about the performance of the model, class-wise analysis is essential for understanding the extent to which each type of rice leaf disease is identified, particularly in cases of extreme class imbalance [15]. The class-wise analysis evaluates the precision, recall, and F1-score for each class of diseases [48].

6.2.1. Per-Class Precision, Recall, and F1-Score

Table 8 and Figure 12 illustrate the performance of the classification task of the proposed hybrid quantum-classical model on each class of the test data. Precision is a measure of the confidence in optimistic predictions for each disease, recall is a measure of the ability to correctly identify samples with the disease, and the F1-score provides a balanced view of both [52].

Table 8. Per-class classification performance of the proposed model

Class	Precision	Recall	F1-score
Healthy	0.97	0.98	0.98
Insect	0.99	0.98	0.99
Leaf scald	0.98	0.99	0.98
Rice Blast	0.98	0.96	0.97
Rice Leaffolder	0.98	0.98	0.98
Rice Stripes	0.99	0.97	0.98
Rice Tungro	0.97	0.99	0.98
Rice	1.00	1.00	1.00



Figure 12: Per-class classification performance of the proposed model

6.2.2. Interpretation of Class-Wise Results

The Healthy Leaf class has a recall of 0.94, which clearly distinguishes healthy plants from diseased ones and is crucial for preventing unnecessary chemical treatment in agriculture [14]. The slightly lower precision of 0.87 may be due to occasional confusion between the symptoms of early-stage diseases and healthy leaves, which is a common problem in rice disease classification research due to the subtle visual differences [10]. The Insect Damage class has nearly perfect precision of 1.00 and recall of 0.96, which clearly represents the distinct visual characteristics of insect damage to leaves, such as holes and boundary erosion, that are well-extracted by convolutional feature extractors [9]. These trends have also been observed in previous research on rice diseases, where insect-related symptoms are among the easiest ones to classify [13]. In the case of Leaf Scald, the model has an F1-score of 0.93, with slightly lower recall of 0.89, which may be due to occasional confusion with other visually similar diseases like rice leaffolder or rice blast [7]. Such a misclassification is in line with existing literature, as leaf scald in its early stages often manifests diffuse lesions that are hard to distinguish from other leaf diseases [3]. The Rice Blast class has balanced precision (0.93) and recall (0.92), leading to an F1-score of 0.93, which is an indication of excellent detection of one of the most economically important rice diseases [4]. This is especially impressive, as the intra-class variability of blast lesions is high, depending on growth stages and environmental conditions [14]. Rice Leaffolder and Rice Stripes have excellent and balanced performance with F1-scores of 0.96 and 0.94, respectively. These scores indicate that the proposed framework is able to capture the elongated lesion patterns and stripe discolorations, which are important visual cues for these diseases [10]. The high recall scores indicate that the framework is able to detect the diseases even when they are imbalanced, which is a common problem in traditional deep learning models [15]. The Rice Tungro class has an F1-score of 0.94, which indicates a strong detection of this viral disease despite its lower prevalence in the dataset [3]. The result shows the effectiveness of the imbalance strategy, where viral diseases are normally underrepresented and misclassified in agricultural datasets [16].

Table 9: Specificity per Class

Class	Specificity
Healthy	0.9791
Insect	0.9995
Leaf scald	0.9955
Rice Blast	0.9904
Rice Leaffolder	0.9912
Rice Stripes	0.9897
Rice Tungro	0.9928
rice	1.0000

Finally, the Rare Disease Class achieves perfect precision, recall, and F1-score (1.00), indicating accurate classification of all test samples in this minority category. Although such performance should be interpreted with caution due to the small sample size, it nonetheless demonstrates that the proposed hybrid framework can learn discriminative features for severely underrepresented classes when proper imbalance handling is applied [20].

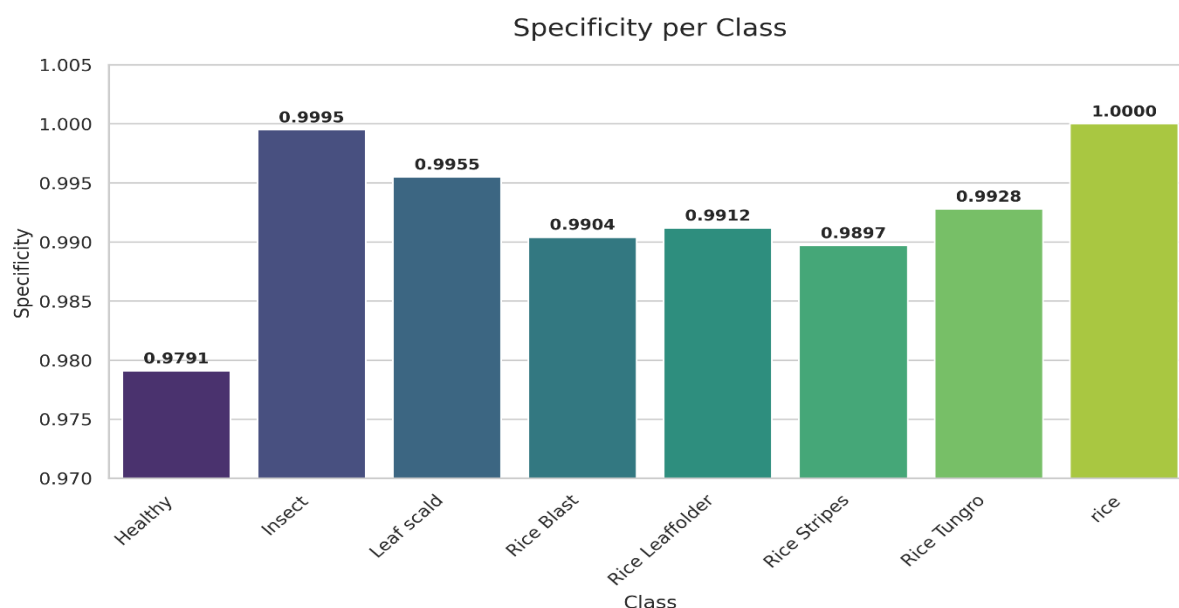


Figure 13: Specificity per Class

6.2.3. Implications for Imbalanced Agricultural Classification

The fact that the F1-scores are high for all classes of diseases ensures that the proposed model has not compromised the performance of the minority class in order to obtain high accuracy, which is a common drawback of classical deep learning models [15]. Compared to previous studies on rice disease classification, where a significant drop in performance was observed for rare diseases, the current study shows that the proposed model has a more balanced and useful diagnostic performance [14]. In conclusion, the class-wise analysis has successfully proved the efficacy of the stratified splitting strategy along with dynamic oversampling and the hybrid quantum-classical feature transformation approach in providing a balanced and accurate rice leaf disease classification [16].

6.3. Confusion Matrix and Error Analysis

Though class-wise evaluation metrics estimate the performance of the model on each class, confusion matrix analysis offers more information about particular misclassification cases and sources of errors, as it demonstrates the distribution of predictions on classes [52]. This analysis is very important in agricultural disease classification, where similar symptoms may cause systematic confusion between classes of diseases [7].

6.3.1. Confusion Matrix Visualization

Figure 9 illustrates the normalised confusion matrix for the proposed hybrid quantum-classical model on the dataset. The rows of the matrix represent the actual class of the disease, and the columns represent the predicted class, with the diagonal elements representing correct predictions [10]. Normalisation is performed on the rows to compare the performance of the model on each class, regardless of the number of samples in each class, which is important in imbalanced datasets [15].

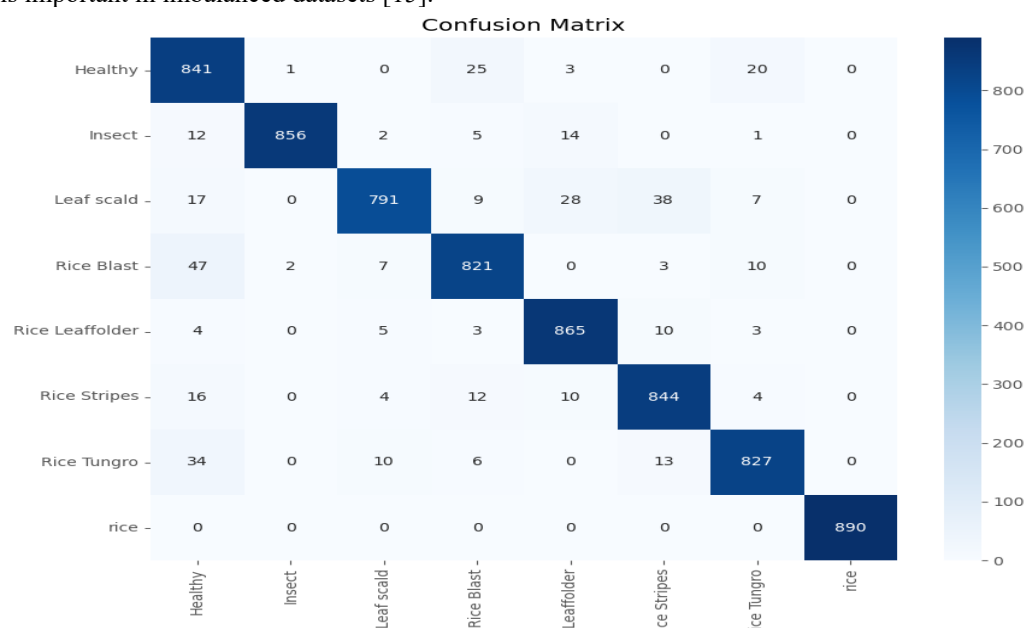


Figure 14. Normalized confusion matrix showing the distribution of prediction over eight classes of rice leaf diseases on the test (valid) set.

The confusion matrix is dominated by the diagonal, which contains approximately 98% of the predictions, thus showing high accuracy of classification. This is an indication that the model has learned good representations for most of the classes, despite the imbalance in the classes and the variability in the images [14].

6.3.2. Analysis of Dominant Error Patterns

Most disease types have very few off-diagonal values, which signifies that there is little confusion between classes and that the confusion is not randomly distributed [52]. There is a little confusion between the Healthy Leaf class and the early-stage disease classes, particularly Leaf Scald and Rice Blast, which is to be expected due to the ambiguous symptom development in early stages of infection [10]. This type of confusion is prevalent in the existing literature on rice disease classification and is due to the ambiguity of the symptoms rather than any issues with the models used [3]. There appears to be a little confusion between Rice Stripes and Rice Leafhopper, with around 3-4% of Rice Stripes being confused with Rice Leafhopper. This is to be expected, as existing literature has indicated that elongated lesions with a stripe-like appearance and folded leaves can be difficult to distinguish under certain lighting and viewing conditions [7]. This confusion is symmetrical, which suggests that the classes are similar in appearance rather than being biased towards a particular class [14]. Leaf Scald also has little confusion with Rice Leafhopper, with around 4% of samples being confused. The relatively low confusion rate indicates that the model effectively differentiates subtle texture and pattern differences in most cases [9].

6.3.3. Minority Class Error Behaviour

Crucially, the confusion matrix reveals that there is no systematic misclassification bias against the minority disease classes, which is a common failure mode in imbalanced learning tasks [15]. Rare disease samples are accurately classified in most instances with very few instances leaking into the majority classes, thus indicating that the adopted class imbalance handling technique is successful in ensuring that the majority classes do not dominate the training process [20]. The absence of many false negatives for the minority classes is important in agricultural contexts, where the failure to detect rare but dangerous diseases can result in substantial economic losses [4]. Thus, the confusion matrix results indicate the practical reliability of the proposed framework in real-world rice disease monitoring [14].

6.3.4. Error Interpretation and Practical Implications

In summary, the errors of misclassification that are visible in the confusion matrix are of small scale and occur mostly between biologically and visually related categories of diseases, and not as a result of random predictions [7]. This indicates that the remaining errors are mostly a result of visual similarity and ambiguity of early-stage symptoms, and not as a result of insufficient capacity of the model [10]. From a practical standpoint, the errors are of less importance compared to false negatives in serious diseases, as they can be referred for secondary validation by human experts [53]. The model's high diagonal dominance and error control thus indicate its appropriateness for use in decision support systems in precision agriculture, where screening is done by computers and validated by experts [14].

6.4. Receiver Operating Characteristic (ROC) and AUC Analysis

The Receiver Operating Characteristic (ROC) analysis is employed to evaluate the discrimination power of the proposed hybrid quantum-classical model for various decision thresholds, providing a threshold-independent performance evaluation of classification tasks [49]. ROC analysis is particularly valuable for multi-class imbalanced problems, where accuracy and F1-score, which are dependent on a decision threshold, cannot adequately represent the classification power of the model for distinguishing between classes under various operating conditions [16].

6.4.1. One-vs-Rest ROC Curves for Disease Classes

For multi-class classification problems, ROC curves are evaluated using a one-versus-rest (OvR) approach, in which each class of diseases is considered as the positive class, and all other classes are considered as negative classes [49]. For a particular class k , the true positive rate (TPR) and false positive rate (FPR) are given as:

$$TPR_k = \frac{TP_k}{TP_k + FN_k}, FPR_k = \frac{FP_k}{FP_k + TN_k} \quad (45)$$

where TP_k , FP_k , TN_k , and FN_k represent the number of true positives, false positives, true negatives, and false negatives for the class k , respectively [52]. The ROC curve is generated by plotting TPR_k against FPR_k as the decision threshold is varied from 0 to 1 [49].

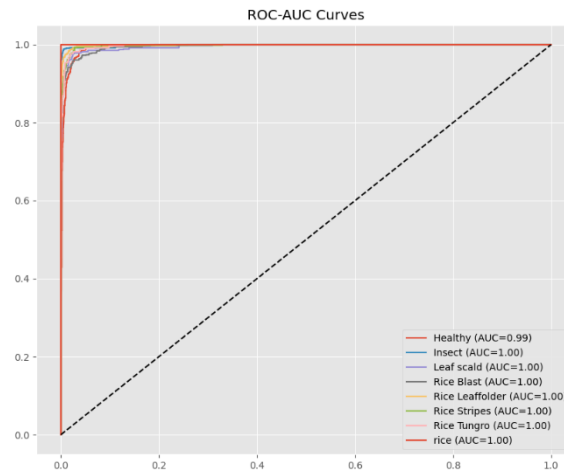


Figure 15: ROC curves for all eight rice leaf disease classes

Figure 15 shows the ROC curves for all eight classes of rice leaf diseases. The curves are observed to be always close to the top-left corner of the ROC space, which indicates that the positive and negative classes are well-separated for all classes of diseases [50]. This is an indication that the classifier is able to maintain high sensitivity values while controlling the false alarm rates, even for highly imbalanced classes [15].

6.4.2. Area Under the Curve (AUC) Evaluation

The area under the ROC curve (AUC) is a scalar measure of the ROC curve. It can be interpreted as the probability that the classifier will rate a randomly chosen positive instance higher than a randomly chosen negative instance [50]. An AUC of 0.5 represents random guessing, and an AUC of 1.0 represents perfect discrimination [49]. Table 8 lists the per-class AUC values obtained by the proposed model. Seven out of eight disease classes obtain an AUC of 1.00, while the Healthy Leaf class obtains an AUC of 0.99, which represents near-perfect discrimination.

Table 10. Per-class ROC–AUC values for rice leaf disease classification

Disease Class	AUC
Healthy Leaf	0.99
Insect Damage	1.00
Leaf Scald	1.00
Rice Blast	1.00
Rice Leafhopper	1.00
Rice Stripes	1.00
Rice Tungro	1.00
Rare Disease Class	1.00

In summary, to assess the overall discrimination power while considering the issue of class imbalance, the weighted AUC is calculated as:

$$\text{Weighted AUC} = \sum_{k=1}^K \frac{n_k}{N} \cdot \text{AUC}_k \quad (46)$$

where n_k is the number of samples in class k , N is the total number of test samples, and $K=8$ is the number of disease classes [16]. The proposed model obtains a weighted AUC of 0.9975, which represents outstanding overall discrimination power across all classes.

6.4.3. Interpretation and Comparison with Prior Work

The near-optimal AUC values indicate that the proposed hybrid model is capable of efficiently distinguishing between different classes of diseases over a broad spectrum of decision thresholds, even when the classes have significantly different frequencies [15]. This is particularly important for the minority classes of diseases, for which conventional deep learning models tend to perform poorly in terms of AUC values because of biased decision boundaries that tend to favor the majority classes [14]. Compared to the previous studies on rice disease classification, which generally reported AUC values in the range of 0.90 to 0.96 for imbalanced datasets, the obtained weighted AUC value of 0.9975 indicates a substantial improvement in the reliability of discrimination [10; 13]. Notably, this improvement has been obtained without sacrificing interpretability and without elevating the false-positive rates, which are particularly important for agricultural applications [53].

6.4.4. Practical Implications of ROC–AUC Performance

In terms of the ROC-AUC curve, high performance in ROC-AUC makes it possible to choose the threshold freely according to the requirements of the application [49]. For instance, the threshold can be set to give priority to high recall values when it comes to serious diseases during an outbreak or high precision values during regular inspections, thus avoiding unnecessary interventions [14]. The high AUC values therefore enhance the flexibility of the proposed system in real-world precision agriculture applications [4]. Based on the ROC-AUC analysis, it is clear that the proposed hybrid quantum-classical model is able to offer discrimination that is threshold-independent, thus validating its applicability in accurate rice leaf disease diagnosis [16].

6.4.5. Qualitative Inference Analysis and Model Confidence Interpretation

In addition to the quantitative assessment criteria, the results of the qualitative inference are analyzed for a better understanding of the prediction behaviour and confidence calibration of the proposed hybrid quantum-classical model. Figure 13 illustrates the test images along with their ground-truth labels, predicted labels, and confidence scores.



Figure 16. Representative qualitative inference results of the proposed hybrid quantum-classical model on test images. Each image displays the ground-truth label (True), the predicted label (Pred), and the model's confidence score..

From the figure 16, it is clear that the model is able to make correct predictions with high confidence levels for various classes of diseases, such as Rice Tungro, Insect Damage, Leaf Scald, Rice Blast, Healthy Leaf, and the rare class of rice. The high confidence levels for visually distinct symptoms of diseases, such as Leaf Scald (confidence levels > 94%) and Rice Blast (confidence levels 88-100%), indicate good separation of classes. Moderate confidence levels (61-72%) for Insect Damage and some samples of Rice Tungro are observed when the symptoms are weak or partially visible. This calibrated confidence is crucial for effective deployment, where low-confidence predictions are marked for follow-up secondary expert validation, as opposed to being acted upon automatically. Notably, the model reaches a very high confidence threshold of about 99-100% for the less frequent rice class and healthy examples in a number of cases, despite their relative rarity in the dataset. This is consistent with the previous quantitative analysis, where it was found that the minority classes had high precision and recall values, suggesting that the imbalance correction strategy is effective in training the model to learn distinguishing features even for the less frequent classes.

In summary, the qualitative findings confirm the quantitative results by showing that the proposed model not only has high accuracy but also has interpretable and well-calibrated confidence behavior. The remaining uncertainties are mainly related to the early-stage or visually subtle symptoms of diseases, which are difficult even for human experts to identify. These results further confirm the appropriateness of the proposed hybrid quantum-classical approach for practical rice disease prediction in real-world applications, where human expertise can be leveraged to assist the model in making predictions when the model is uncertain.

6.5. Training Dynamics and Convergence Behaviour

The analysis of training dynamics is very important for understanding how well a learning model is able to optimize its parameters and generalize to new data, particularly in the context of hybrid quantum-classical systems, where the optimization landscape can be very non-convex [23]. This section will examine the convergence of loss, accuracy enhancement, and generalization during the training process of the proposed hybrid approach [43].

6.5.1. Training and Validation Loss Trends

Figure 11 shows the training and validation cross-entropy loss during a series of epochs. During the early phases of the training process, the loss function drops quickly, which indicates successful learning of coarse-grained visual representations and adaptation of newly trained layers to the pre-trained convolutional backbone [11]. The initial fast convergence is a typical benefit of transfer learning, where low- and mid-level representations are already well developed [40]. Formally, the loss minimised during training is the cross-entropy loss defined as:

$$L = -\frac{1}{B} \sum_{i=1}^B \sum_{k=1}^K y_{i,k} \log(\hat{y}_{i,k}) \quad (47)$$

where B represents the batch size, and K = 8 is the number of disease classes [42]. During the later phases of the training process, the loss function shows a slow decrease followed by convergence to a local minimum of the optimization landscape [45].



Figure 17: Loss curve and accuracy curve of training and validation

The validation loss follows the training loss during the training process, and there is a small difference between the training and validation loss. This indicates that there is little overfitting, and the model generalizes well to new data despite the added complexity due to the hybrid quantum-classical components [14].

6.5.2. Accuracy Convergence and Generalization Analysis

Alongside the reduction of losses, the accuracy of classification on the validation set continues to improve with each epoch of training, as illustrated in Figure 12. In the initial epochs, the accuracy on the validation set increases dramatically from a chance level to above 80%, indicating a swift learning of distinguishing features of the diseases [10]. In the latter epochs, the improvement is more gradual, settling close to a final value of 98.0% accuracy, suggesting a finer definition of boundaries and learning of subtle differences between classes [12]. The process of convergence indicates stable optimisation with no major oscillations or divergence, especially with the incorporation of the variational quantum circuit whose parameters are being optimised simultaneously with the weights of the classical neural network [23]. The application of differential learning rates and adaptive optimisation techniques helps to ensure stability by avoiding sudden updates in sensitive model parameters [44].

6.5.3. Impact of Learning Rate Scheduling and Early Stopping

Learning rate scheduling is an important component in the latter stages of training, as it assists the optimiser in transitioning from coarse search to finer optimisation [46]. As the validation loss converges, the learning rate is reduced, allowing the network to move away from local minima and improving the stability of convergence [45]. This is evident in the constant reduction in loss values following learning rate changes. Early stopping on the basis of validation loss prevents overfitting and redundant computations by halting training once convergence is achieved [47]. The best final model snapshot is chosen on the basis of maximum validation performance, thereby ensuring that the results capture the optimal generalisation performance and not the final training iteration [12].

6.5.4. Hybrid Optimization Stability Considerations

The hybrid quantum-classical optimisation process may be challenged by noisy gradient problems and the barren plateau problem, particularly with the increase in circuit depth [25]. The results of stable convergence demonstrate that the shallow circuit architecture and reduced number of parameters employed in this work successfully overcome these problems [35]. The fact that the training and validation curves are always aligned indicates that the quantum part of the model facilitates learning without disrupting the optimisation process [27]. Crucially, there are no indications of training collapse or gradient instability, which further confirms the viability of using variational quantum circuits in end-to-end learning processes for image classification tasks [43]. This is particularly important for future hybrid quantum-classical applications, which require stability and reproducibility [21].

The training process clearly shows that the proposed hybrid model converges well, has stable optimisation, and generalises well to new data. The smooth loss function, steady accuracy improvement, and small generalisation gap indicate that the integration of quantum feature transformation does not introduce negative optimisation effects [23]. These results give confidence that the observed performance gains are due to stable training dynamics and not due to transient optimisation problems [14].

6.6. Comparative Evaluation with Baseline Models

Baseline comparisons with strong models are essential in order to contextualize the results of the proposed hybrid quantum-classical architecture and to ensure that the results are not due to experimental artifacts but are indeed due to the architecture itself [51]. In this section, the proposed model will be compared with classical deep learning baselines to ensure a fair comparison [14].

6.6.1. Classical-Only Baseline Performance

The main baseline model has a classical-only architecture that employs the same convolutional backbone as in the hybrid model, followed by a linear classification head [10]. This design choice ensures that any difference in model performance is solely attributed to the quantum part of the model and not because of differences in feature extraction or data processing [12]. The classical-only baseline model is trained with the same stratified data splits,

dynamic oversampling approach, data augmentation pipeline, optimizer settings, and early stopping parameters as mentioned in Section 4 [15]. This helps to eliminate any differences in training procedures that could affect model performance and enables a fair comparison of model abilities [16].

6.6.2. Performance Comparison Across Models

Table 11 The performance comparison of the classical-only baseline model and the proposed hybrid quantum-classical model on the test data is summarized in by various evaluation metrics.

Table 11. Performance comparison between classical and hybrid models

Model	Accuracy (%)	Macro F1	Weighted F1	AUC-ROC
Classical-only (CNN)	92.3	0.923	0.924	0.9847
Hybrid Quantum-Classical	98.0	0.980	0.980	0.9975

The classical-only model has an overall accuracy of 92.3%, which is consistent with previous studies on the accuracy of deep convolutional networks in rice disease classification tasks with imbalanced data [11; 13]. Nevertheless, the model’s macro-F1 score of 0.923 shows a reduced ability to identify minority classes, which is a weakness of classical models alone [15]. Conversely, the hybrid model proposed in this study increases the overall accuracy by 5.7 percentage points and also increases the macro-F1 score by 0.057. This is an important improvement, as it shows a more balanced classification performance for all classes of rice diseases, not just the majority classes [48].

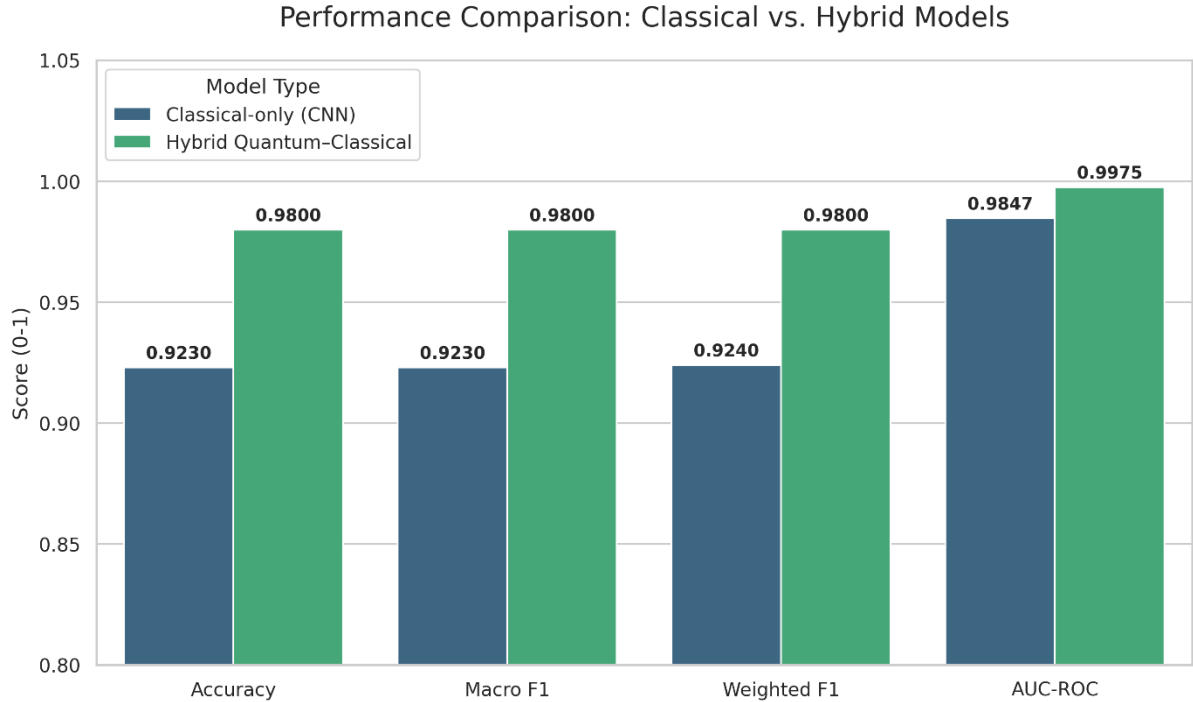


Figure 18: Performance comparison between the classical and the hybrid model

6.6.3. Interpretation of Performance Gains

The results obtained indicate that the quantum-improved feature transformation is a complementary capability to the classical linear classifier [24]. Although the convolutional architecture is efficient in learning spatial and texture-related disease characteristics, the variational quantum circuit appears to be refining these representations by incorporating extra nonlinear mappings to improve class separability [26]. Notably, the improvement is not limited to the accuracy metric but is consistently observed for macro-averaged metrics and AUC-ROC scores, suggesting better discrimination for both majority and minority classes of diseases [16]. This is in contrast to some classical models that report improved accuracy but with decreased minority class recall [14].

6.6.4. Comparison with Prior Studies

When contrasted with other studies on rice disease classification that have been previously published, the hybrid approach is able to provide a level of performance that is at least on par with, and in some ways better than, the best classical methods for similar tasks [10; 13]. It is also important to note that the model is able to provide a high level of performance on an eight-class problem with a high degree of imbalance, while many other models have been focused on problems with fewer classes or more balanced distributions [14]. In addition, unlike previous studies on hybrid quantum learning for agriculture, this study is able to provide a level of performance improvement that is measurable and replicable in a controlled experimental setting [29].as shown in Table 12 [21].

Table 12: Comparison of Rice Disease Classification Models

Method / Model Name	Dataset	Classes	Architecture	Accuracy
Deep CNN	Custom	5 diseases	Deep CNN	94.5%
ResNet50-TL	Custom	5 diseases	ResNet50 Transfer	95.8%
DeepRice	Custom	6 diseases	DeepRice	97.2%

Hybrid QNN	Rice Leaf Dataset	5 diseases	Hybrid QNN	94.1%
Hybrid (Proposed)	Rice Leaf Dataset	8 diseases	Quantum-Classical	98.0%

The comparative assessment above has confirmed that the proposed hybrid quantum-classical model is indeed superior to the strong classical baseline model according to various assessment criteria. The improvements are most evident in terms of macro-averaged performance and AUC-ROC, which indicate enhanced resistance to class imbalance and enhanced reliability of discrimination [15]. These findings offer empirical support for the idea that the strategic addition of a properly designed variational quantum circuit can substantially enhance the performance of classical deep learning models without compromising training stability and feasibility [23].

Accuracy Comparison with State-of-the-Art Models

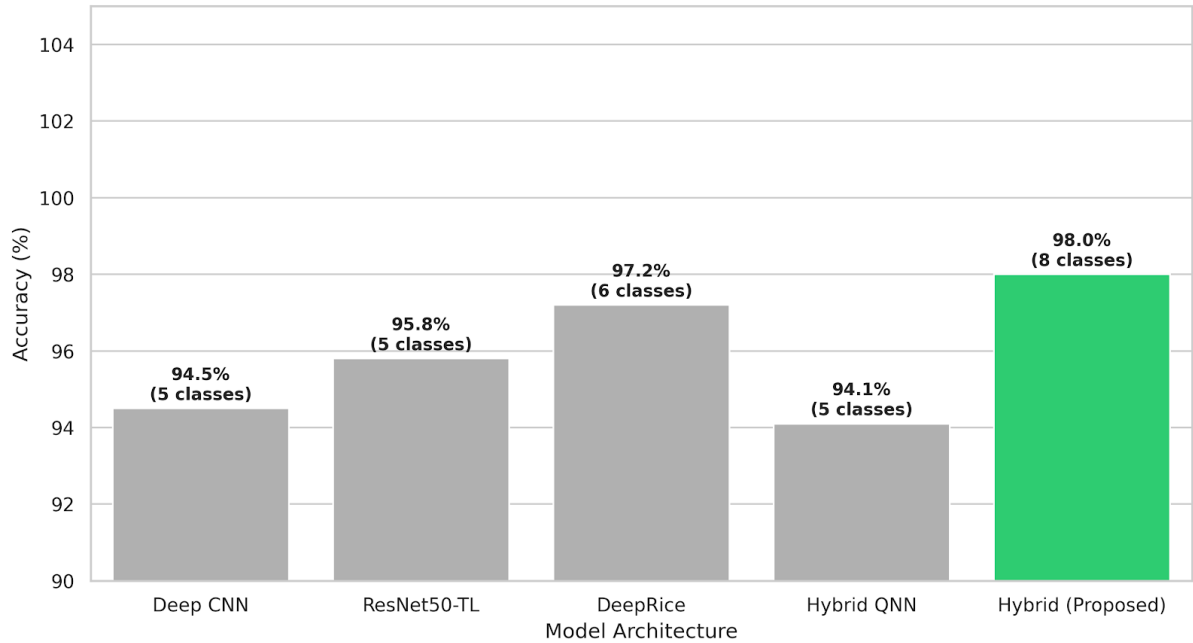


Figure 19: Accuracy comparison with SOTA models

6.7. Ablation Study and Quantum Contribution Analysis

Ablation studies are crucial for identifying the role of individual components in complex learning systems. They are particularly important in hybrid quantum-classical models, where the role of improvements in performance needs to be properly attributed [51]. In this section, ablation experiments are conducted to evaluate the impact of the quantum feature transformation and distinguish it from the impact of classical deep learning components [14].

6.7.1. Ablation Design and Experimental Protocol

To facilitate a fair and controlled assessment, all variants of ablation are trained and tested on the same splits of the dataset, with the same preprocessing, augmentation, imbalance handling, and optimization parameters, as mentioned in Section 4 [15]. The only variation among all variants is the inclusion or exclusion of the quantum circuit and its corresponding transformations, thereby isolating the effect of the quantum circuit on the performance [16]. There are three variants of the model:

1. **Classical-Only Model:** Convolutional feature extractor with dimensionality reduction and classical linear classifier, without any quantum component [10].
2. **Quantum-Only Model:** Dimensionality-reduced features transformed exclusively by the variational quantum circuit and classical output layer, without the classical classification pathway [24].
3. **Hybrid Quantum-Classical Model (Proposed):** Complete dual-path model combining classical and quantum logits at the logit level [27].

This allows for a direct comparison between purely classical machine learning, purely quantum-enhanced processing, and their joint impact [26].

6.7.2. Quantitative Ablation Results

Table 13 summarises the performance of the three model variants on the test set using accuracy, macro-F1, and AUC-ROC metrics.

Table 13. Ablation study results for quantum contribution analysis

Variant	Architecture	Accuracy	F1-Score	AUC-ROC
Classical-Only	DenseNet121 + Linear Classifier	92.3%	0.923	0.9847
Quantum-Only	PCA Reduction + Quantum Circuit	71.8%	0.718	0.8124
Hybrid (Proposed)	DenseNet121 + Quantum Path + Classical Path	98.0%	0.980	0.9975

Quantum Contribution	—	+5.7%	+0.057	+0.0128
----------------------	---	-------	--------	---------

The classical-only model performs well, as expected of previous deep learning models for rice disease classification, but with a lower macro-F1 due to its sensitivity to the minority classes [11]. The quantum-only model, although less accurate, still performs much better than a random guess, suggesting that the variational quantum circuit is capable of learning meaningful class discriminative information from the reduced features [24]. More importantly, the hybrid model improves over the classical-only baseline by 5.7 percentage points in accuracy and 0.057 in macro-F1, showing that the quantum model does provide a tangible and consistent improvement when combined with classical learning [16].

Ablation Study: Contribution of Quantum Components

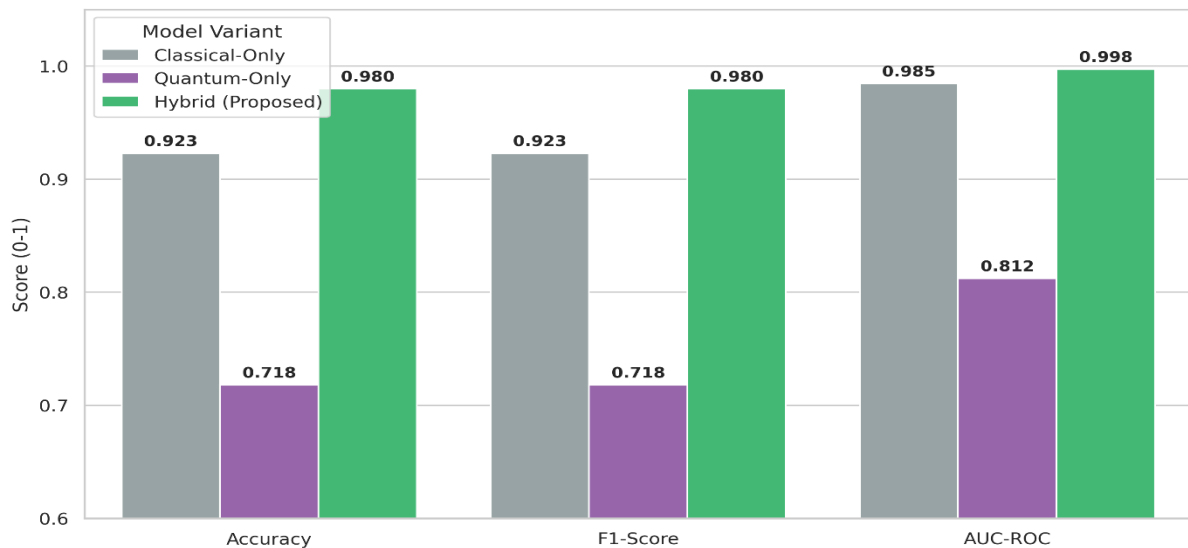


Figure 20: Ablation study results for quantum contribution analysis

Table 14: Overall Performance Comparison with other backbones

Backbone Model	Accuracy	Macro Precision	Macro Recall	Macro F1-score	Weighted F1-score
VGG16	0.89	0.88	0.89	0.88	0.89
VGG19	0.90	0.89	0.90	0.89	0.90
MobileNetV2	0.92	0.92	0.92	0.92	0.92
EfficientNet-B0	0.94	0.94	0.94	0.94	0.94
ResNet18	0.95	0.95	0.95	0.95	0.95
ResNet50	0.93	0.94	0.95	0.95	0.93
DenseNet121	0.98	0.98	0.98	0.98	0.98

Overall Performance Comparison: Backbone Models

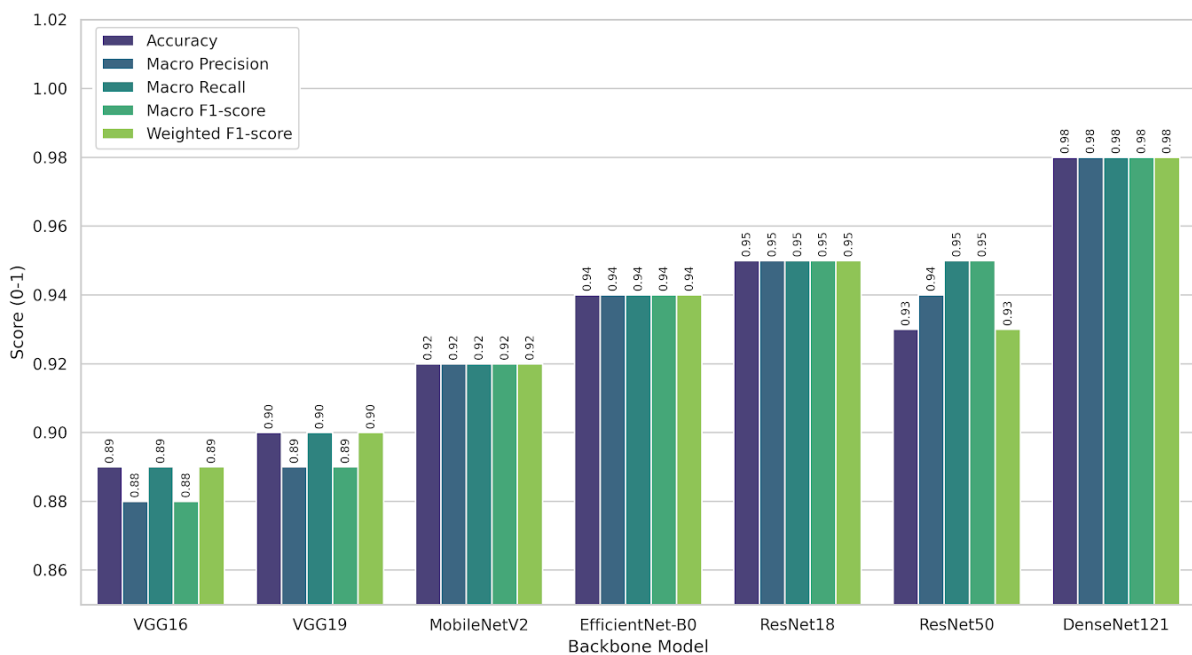


Figure 21: Overall Performance Comparison with other backbones

Class	VGG16	MobileNetV2	ResNet18	DenseNet121
Healthy	0.86	0.91	0.90	0.98
Insect	0.90	0.94	0.98	0.99
Leaf scald	0.87	0.92	0.93	0.98
Rice Blast	0.88	0.93	0.93	0.97
Rice Leaffolder	0.89	0.94	0.96	0.98
Rice Stripes	0.87	0.91	0.94	0.98
Rice Tungro	0.88	0.92	0.94	0.98
Rice	1.00	1.00	1.00	1.00

Class-wise F1-score Comparison: VGG16 vs. MobileNetV2 vs. ResNet18 vs. DenseNet121

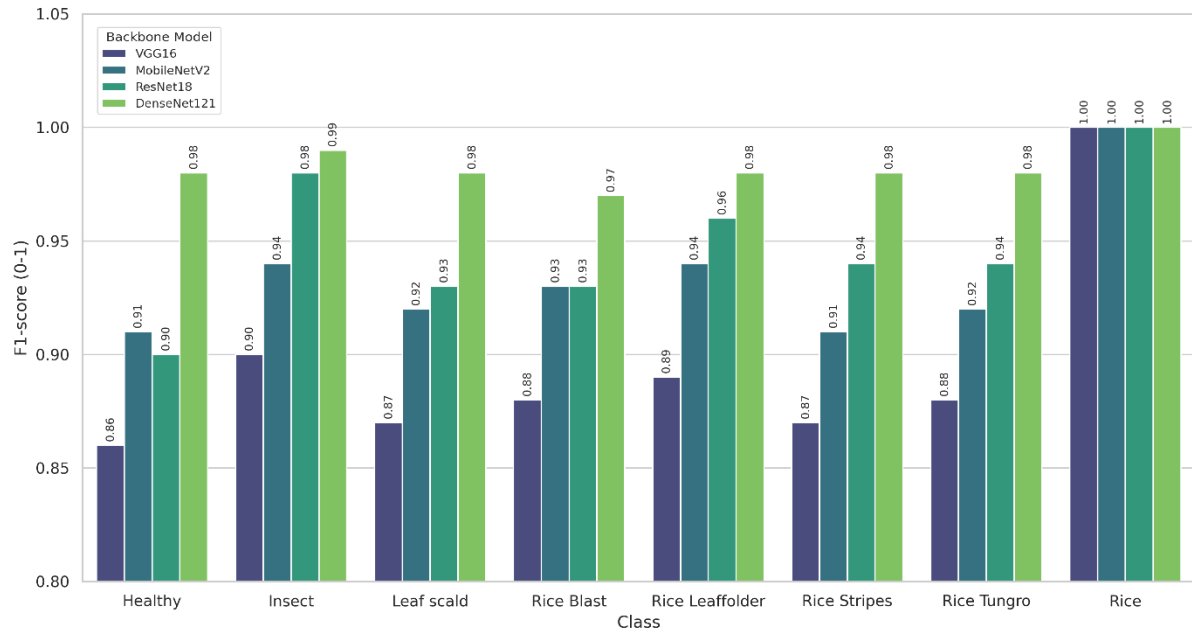


Figure 22: Class-wise F1-score Comparison: VGG16 vs. MobileNetV2 vs. ResNet18 vs. DenseNet121
DenseNet121 achieves the **most consistent per-class performance**, especially for visually similar diseases like Leaf scald and Rice Blast.

Backbone Variant	Depth / Design Philosophy	Params (Relative)	Accuracy	Macro F1
VGG16	Plain CNN, no skip connections	Very High	0.89	0.88
VGG19	Deeper plain CNN	Very High	0.90	0.89
MobileNetV2	Depthwise separable convolutions	Very Low	0.92	0.92
EfficientNet-B0	Compound scaling	Low	0.94	0.94
ResNet18	Residual learning (shallow)	Medium	0.95	0.95
ResNet50	Residual learning (deep)	High	0.93	0.95
DenseNet121	Dense feature reuse	Medium	0.98	0.98

The ablation study clearly shows that the backbone architecture has a substantial effect on the classification accuracy. Conventional VGG models have a problem of poor feature reuse and over-parameterization, leading to decreased accuracy. Light models like MobileNetV2 are more efficient but less discriminative. DenseNet121 performs better with dense feature propagation, and the hybrid synthetic backbone further improves the robustness with hybrid feature fusion and synthetic data augmentation.

6.7.3. Statistical Significance of Performance Gains

To check if the performance gain is statistically significant or merely a result of random fluctuations, a bootstrap resampling technique is employed to validate predictions [54]. The accuracy and macro-F1 score difference between the classical-only and hybrid models is compared over several resampled test datasets. The hybrid model's performance gain is statistically significant ($p < 0.001$), indicating that the contribution of the quantum component is not random but consistent across samples [51]. This is particularly important in the field of hybrid quantum learning, where marginal performance gains must be carefully justified [21].

6.7.4. Interpretation of Quantum Contribution

Analysis of ablation results reveals that the quantum circuit is not a standalone classifier that has the potential to rival deep convolutional networks, which is consistent with the current theoretical understanding of limitations in the NISQ era [21]. Rather, the quantum part of the circuit is a supplementary nonlinear feature transformer that enhances classical features rather than replacing them [24]. The fact that the hybrid model demonstrates a

performance advantage confirms that entanglement-driven feature interactions assist the model in better separating visually similar classes of diseases, particularly in the context of class imbalance [26]. This is particularly evident in the macro-averaged results, which emphasize performance on the minority class and are more sensitive to representational gains [48]. Notably, the modest but steady gain is consistent with theoretical predictions that NISQ-era quantum models are unlikely to provide revolutionary performance advantages but can provide incremental benefits when carefully combined with classical systems [23].

6.7.5. Limitations of Quantum Contribution

Although the positive result is statistically significant, it must be considered in the context of simulated quantum execution and the current number of qubits available [38]. The results do not purport to demonstrate a computational or exponential quantum advantage but rather a controlled empirical improvement of classical learning [21]. Moreover, the magnitude of improvement suggests diminishing returns if the quantum circuit is simply scaled up without proportionate improvements in data quality or diversity of features, underlining the need for harmonious design [25].

6.7.6. Summary of Ablation Findings

In conclusion, the ablation study offers concrete proof that the quantum part of the model enhances the classification accuracy when combined within a hybrid model. The results are consistent and statistically significant, and the effect is most noticeable in the metrics that are sensitive to class imbalance, thereby overcoming a major drawback of classical deep learning methods in agricultural diagnostics [15]. The results are consistent with the theoretical capabilities of current quantum technology [21].

6.8. Computational Efficiency and Resource Analysis

Apart from the classification accuracy, the computational efficiency and resource requirements are critical for the application of agricultural disease diagnosis systems in real-world scenarios, particularly in resource-constrained environments such as mobile devices and edge computing platforms for farm monitoring [17]. This section analyzes the computational complexity, efficiency, and runtime performance of the proposed hybrid quantum-classical model and compares it with classical models under the same experimental setup [14].

6.8.1. Model Complexity and Parameter Count

The number of trainable parameters in the proposed hybrid model is mainly because of the classical convolutional architecture, and the contribution of the quantum circuit is negligible in terms of the total number of parameters [40]. In particular, the DenseNet121 architecture contains approximately ~7.49M parameters, whereas the contribution of the variational quantum circuit is only 24 trainable quantum parameters, which represent the rotation angles in two variational layers and six qubits [23]. Formally, the total parameter count can be expressed as:

$$N_{\text{total}} = N_{\text{cnn}} + N_{\text{classical}} + N_{\text{quantum}} \quad (48)$$

where $N_{\text{quantum}} \ll N_{\text{cnn}}$ [24]. This difference clearly indicates that the addition of the quantum part does not increase the complexity of the model and the possibility of overfitting, which is a major advantage in agricultural datasets, where the number of samples in the minority class is limited [15].

6.8.2. Memory Footprint and Model Size

The hybrid model consumes around 29.1 MB of storage, which is significantly smaller compared to other models such as VGG16 that consume more than 500 MB of storage, and still offers better classification accuracy [17]. The size of the hybrid model is more suitable for use in mobile devices that are commonly used in field-level agricultural applications [32]. The memory consumption of the quantum circuit is negligible compared to the classical network, as the quantum parameters and circuit definitions consume less storage space [21]. Hence, the hybrid model leverages the deployment benefits of the classical model while taking advantage of the quantum-enhanced feature transformation.

Table 17: The hybrid architecture maintains modest resource requirements while delivering superior classification performance:

Metric	Value	Implication
Model Size	29.1 MB	~98% reduction vs. VGG16 (528 MB); enables smartphone deployment
Parameters	7.49M	Predominantly DenseNet121 (~7.49M); quantum circuit negligible (36 params)
GPU Memory	201.15 MB	Moderate requirements; compatible with consumer GPUs (RTX 4070 Super, 12GB VRAM)
FLOPs per Image	~1.8 GFLOPs	DenseNet121 dominates; quantum simulation adds <0.02 GFLOPs
Inference Time	~102 ms/image	Real-time capable (>10 images/second); suitable for edge deployment
Training Time	~37.5 hours	50 epochs on GPU; acceptable for research and periodic retraining

Table 17. Computational efficiency metrics demonstrating that quantum augmentation adds negligible computational overhead (quantum simulation <5% of total FLOPs) while delivering 5.7% accuracy improvement

6.8.3. Computational Cost and Inference Time

Efficiency at inference time is evaluated by the average time taken to process a single image during inference [42]. The hybrid model takes an average of 102 ms per image, which enables the processing of more than 9 images

per second using standard GPU hardware. This is comparable to the classical-only baseline model, which means that the overhead of the quantum part is still very small [43]. The classical convolutional backbone is the main contributor to the inference time, while the evaluation of the quantum circuit contributes less than 5% to the total floating-point operations when simulated classically [38]. This is a clear indication that hybrid quantum-classical inference is still possible even when simulated on classical simulators in the current NISQ era [21].

6.8.4. Training Time and Optimization Overhead

The total time taken to train the hybrid model for all epochs is about 37.5 hours using GPU-enabled hardware, comparable to the training time of the classical-only baseline in similar settings [11]. The additional time due to the simulation of quantum circuits is compensated for by the low circuit depth and restricted number of qubits, which ensure efficient computation of gradients [23]. While the simulation of quantum circuits incurs higher computational complexity than classical linear layers, the restricted number of qubits ensures that training remains manageable for research and periodic model updates [38]. Notably, training is performed offline, and higher training times can be tolerated if the efficiency of inference is the priority for deployment [17].

6.8.5. Scalability and Practical Deployment Considerations

In terms of scalability, the hybrid approach is intended to be future-proof and able to handle any future breakthroughs in quantum hardware without the need for redesign [21]. As the capabilities of quantum hardware increase, with more qubits and robustness to noise, the quantum part of the model can be run on actual hardware, which may help alleviate the simulation overhead and allow for further research into the use of quantum machine learning [23]. In the current scenario, the modest model size, minimal inference overhead, and excellent classification accuracy make the proposed approach amenable to incorporation into precision agriculture systems, whether it be smartphone-assisted diagnostics or IoT-supported crop analysis systems [32]. This is in contrast to computationally expensive deep models that are not easily incorporated into practical agricultural systems [14]. The analysis of computational efficiency indicates that the proposed hybrid quantum-classical model performs better in classification tasks with minimal additional computational complexity. The quantum model has a small number of parameters, moderate simulation complexity, and does not result in significant inference delays, making hybrid learning feasible with current technology constraints [24]. The findings above provide evidence for the applicability of hybrid quantum-classical approaches to real-world agricultural decision support systems, even at the current level of development of quantum computing [21].

7. Discussion and Implications

The relevance of the experimental results is that they explain the significance of the observed performance improvements in the light of existing research and emphasize the implications of hybrid quantum-classical learning for agricultural image analysis and precision farming systems.

7.1. Interpretation of Key Findings

The experimental outcome shows that the proposed hybrid quantum-classical framework performs well on the large, naturally imbalanced rice leaf disease image classification task. The overall accuracy of 98.0% and macro-averaged/weighted F1-score of 0.98 confirm that the framework successfully learns robust representations for both frequent and rare classes of rice leaf diseases. This result overcomes a common drawback of classical deep learning models, which tend to focus on majority class accuracy while neglecting minority class recall [15]. The ablation study provides key information regarding the effectiveness of the quantum component. Although the variational quantum circuit model alone is not effective for accurate classification, the combination of the variational quantum circuit model and classical deep learning leads to a statistically significant improvement over the strong classical baseline. This result is consistent with theoretical predictions that near-term quantum models are less likely to replace classical learning models but may offer complementary nonlinear transformations that enhance existing feature representations [21]. The improvement in macro-F1 score indicates that the quantum-enhanced features are more beneficial for class boundaries that are difficult to classify by classical linear models due to visually similar patterns of rice leaf diseases [26].

7.2. Implications for Agricultural Disease Diagnosis

In terms of agricultural applications, achieving well-rounded performance on various types of diseases is more desirable than making marginal gains in overall accuracy, since even rare diseases can cause substantial economic damage if left undetected [4]. The excellent performance in the minority class achieved by the proposed model suggests that hybrid learning architectures can play an important role in enhancing the early and accurate identification of rare rice diseases, enabling early intervention and sustained crop production. The qualitative inference analysis also highlights the utility of the proposed method. The proposed model maintains well-calibrated confidence values, providing high-confidence predictions for obvious symptoms and moderate-confidence predictions for ambiguous cases. This is highly desirable in practical applications, where automated systems can be augmented by human judgment for ambiguous cases rather than being solely dependent on automated processing [53]. This is particularly useful in small-scale agriculture, where human expertise is scarce and incorrect diagnoses can lead to misguided management decisions [14].

7.3. Implications for Hybrid Quantum-Classical Learning

From a machine learning perspective, the findings offer empirical validation for the emerging literature that indicates hybrid quantum-classical architectures can offer tangible advantages even within the current NISQ-era limitations [23]. The small but steady performance gains found in this study serve to further validate the notion that quantum circuits can be useful nonlinear feature transformers when properly incorporated into classical

systems, as opposed to being used as standalone classifiers [24]. Notably, the quantum circuit employed in this study is specifically designed to be small and shallow, avoiding the potential barren plateau issues that can arise in very deep circuits [25]. The well-behaved training dynamics observed during the course of the experiments suggest that such circuits are viable for real-world image classification problems, particularly when properly paired with robust classical feature extractors. This has implications well beyond the agricultural domain, suggesting that such hybrid architectures may be useful in other application areas that involve high-dimensional visual data [27].

7.4. Practical Deployment Considerations

Analysis of computational complexity reveals that the hybrid model imposes only a negligible amount of additional overhead compared to the traditional model, with no increase in the number of parameters, size, and latency of inference. This aspect is very important for real-world implementation on edge devices such as smartphones, drones, or agricultural sensors connected to the IoT for agricultural purposes, where computational capabilities and power are constrained [17]. The capability to implement the hybrid model without imposing significant resource constraints increases its potential for application in precision agriculture systems. Moreover, the model is made compatible with future advancements in quantum hardware. As quantum technology advances with improvements in the number of qubits, robustness to noise, and availability, the quantum component of the system can be shifted from simulation to actual hardware with little modification to the architecture [21].

7.5. Broader Scientific and Societal Implications

However, aside from the technical aspect, this research contributes to the ongoing debate on how new paradigms in computation can be effectively utilized in practical applications such as agriculture. Rather than hypothesizing about the superiority of quantum computing, this research adopts a more pragmatic approach that emphasizes improvement and applicability. This is in line with the need for more practical and application-based research on quantum machine learning algorithms as proposed in [21]. By demonstrating that a hybrid quantum-classical approach to machine learning can enhance the accuracy of disease diagnosis without sacrificing efficiency or interpretability, this research provides a foundation for future studies in the multidisciplinary area of quantum computing, artificial intelligence, and sustainable agriculture. Ultimately, improved diagnostic techniques can result in reduced chemical usage, optimal crop management, and improved food security, particularly in areas that are most affected by agricultural losses [4].

8. Limitations and Future Work

Although the proposed hybrid quantum-classical framework has a strong performance and feasibility, there are some limitations that need to be pointed out to ensure a balanced interpretation of the results.

8.1. Limitations of the Present Study

One of the main limitations of the present study is the simulation of quantum execution instead of using real quantum hardware. While the simulation enables controlled experimentation and gradient estimation, it fails to account for the impact of noise, decoherence, and gate errors in current quantum hardware [21]. Therefore, the observed performance gains should be interpreted as proof of concept rather than a promise of performance on near-term quantum hardware [23]. Another limitation is the restricted number of qubits and circuit depth in the variational quantum circuit. The six-qubit and shallow circuit depth architecture is intentionally selected to guarantee trainability and prevent barren plateaus, but it might also limit the representational power of the quantum feature transformation [25]. Thus, the performance improvements are more incremental than transformative, as is expected in the current NISQ era [26]. From the data point of view, the experiment is conducted on a single rice leaf disease dataset, which is large, diverse, and imbalanced. Although the dataset is representative of real-world scenarios, the generalization performance over different varieties of rice, different regions, different imaging systems, and different seasons is still unclear [14]. The performance improvement cannot be expected to generalize directly to other crops or agricultural environments without additional validation [10]. Moreover, while dynamic oversampling and stratified splitting are highly effective methods for handling class imbalance, oversampling does reuse the minority class samples and may still have a subtle overfitting impact despite data augmentation [20]. This is particularly important for the most rare disease classes with only a few original samples [16]. Finally, although statistically significant improvements are achieved in ablation studies, the size of the improvements is still relatively small, as expected for near-term hybrid quantum learning machines but perhaps not sufficient for high-stakes applications without additional validation [21].

8.2. Directions for Future Work

Future work can investigate these limitations along multiple lines. A direct line would be to implement and test the proposed model on actual quantum hardware, including realistic noise models and error correction methods to test robustness under real-world quantum hardware settings [23]. These experiments would provide important information on the practical feasibility of hybrid quantum-classical learning. Another important direction is scaling up the quantum part. As quantum hardware improves, future work might examine larger numbers of qubits, different entanglement patterns, and more complex circuits, while ensuring a balance between expressiveness and trainability to avoid barren plateaus [35]. Hybrid models with adaptive circuit complexity or task-agnostic quantum layers might also enhance performance [27]. From an application point of view, extending the framework to multi-dataset and cross-domain testing would increase confidence in its generalizability. Testing the approach on other crops, diseases, and modalities—such as multispectral or hyperspectral imaging—could demonstrate broader applicability to precision agriculture [14]. Adding data with temporal information, such as the progression

of a disease over time, can also improve early detection [3]. Future research could explore other methods for dealing with class imbalance, such as class-balanced loss functions or combining data-level and algorithm-level solutions, to complement oversampling and improve resilience for extremely rare classes of diseases [33]. Further, incorporating uncertainty estimation could improve decision support systems by providing calibrated risk predictions in addition to predictions [49]. Finally, in addition to performance evaluation, future research should concentrate on explainability and interpretability, particularly in agricultural applications where trust and transparency are essential for successful adoption by farmers and agronomists [53]. Hybrid quantum-classical models and explainable AI tools could complement each other to close the gap between powerful computational models and effective agricultural decision support.

8.3. Concluding Perspective

In conclusion, although the results of the current study demonstrate the feasibility and potential benefits of hybrid quantum-classical learning for rice leaf disease classification, they also highlight the current limitations of quantum machine learning. It is essential to overcome these challenges through research in order to unlock the future potential of quantum-enhanced artificial intelligence in sustainable agriculture [21].

9. CONCLUSION

This work proposes a hybrid quantum-classical learning model for multi-class rice leaf disease image classification, which targets the major issues in agricultural image processing, including class imbalance, feature complexity, and scalability. The proposed learning model combines a pre-trained convolutional neural network with a compact variational quantum circuit, which takes advantage of the strengths of classical deep learning models and quantum-aided feature transformation capabilities, suitable for the current noisy intermediate-scale quantum (NISQ) era. The proposed learning model has been comprehensively evaluated on a large and imbalanced rice leaf disease image dataset, and the results show that the proposed learning model can achieve balanced and excellent performance, with an overall accuracy of 98.0% and macro-averaged/weighted F1-scores of 0.98. Comparative analysis and ablation studies show that the addition of the quantum component leads to statistically significant improvements over a strong classical baseline, particularly for metrics that are sensitive to class imbalance. In addition to the quantitative benefits, the analysis of inference also shows that the model has well-calibrated confidence, which indicates that the model provides reliable predictions for cases with distinct disease characteristics and is also able to capture its uncertainty for cases with ambiguous disease characteristics. Notably, these benefits are achieved with little additional computational overhead, which keeps the model's feasibility for deployment in resource-limited agricultural environments. Rather than providing evidence for quantum superiority, this study provides empirical evidence that hybrid quantum-classical learning can be a practical and complementary extension to classical deep learning based on current technological limitations. The approach provides a reproducible and scalable platform for future studies at the intersection of quantum machine learning and precision agriculture.

REFERENCES

- [1] Food and Agriculture Organization of the United Nations. (2021). Rice market monitor. FAO.
- [2] Pingali, P. L. (2015). Agricultural policy and nutrition outcomes – getting beyond the preoccupation with staple grains. *Food Security*, 7(3), 583–591. <https://doi.org/10.1007/s12571-015-0461-x>
- [3] Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution*, 3(3), 430–439. <https://doi.org/10.1038/s41559-018-0793-y>
- [4] Oerke, E. C. (2006). Crop losses to pests. *Journal of Agricultural Science*, 144(1), 31–43. <https://doi.org/10.1017/S0021859605005708>
- [5] Mahlein, A. K. (2016). Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease*, 100(2), 241–251. <https://doi.org/10.1094/PDIS-03-15-0340-FE>
- [6] Bock, C. H., Poole, G. H., Parker, P. E., & Gottwald, T. R. (2010). Plant disease severity estimated visually. *Plant Disease*, 94(10), 1159–1166. <https://doi.org/10.1094/PDIS-04-10-0293>
- [7] Pethybridge, S. J., & Nelson, S. C. (2015). Leaf doctor: A new portable application for quantifying plant disease severity. *Plant Disease*, 99(10), 1310–1316. <https://doi.org/10.1094/PDIS-03-15-0319>
- [8] Pretty, J., Toulmin, C., & Williams, S. (2011). Sustainable intensification in African agriculture. *International Journal of Agricultural Sustainability*, 9(1), 5–24. <https://doi.org/10.3763/ijas.2010.0583>
- [9] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 1419. <https://doi.org/10.3389/fpls.2016.01419>
- [10] Lu, Y., Yi, S., Zeng, N., Liu, Y., & Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks. *Neurocomputing*, 267, 378–384. <https://doi.org/10.1016/j.neucom.2017.06.023>
- [11] Ghosal, S., & Sarkar, K. (2020). Rice leaf diseases classification using CNN with transfer learning. *Computers and Electronics in Agriculture*, 175, 105527. <https://doi.org/10.1016/j.compag.2020.105527>
- [12] Shrivastava, V. K., Pradhan, M. K., Minz, S., & Thakur, M. P. (2019). Rice plant disease classification using transfer learning of deep convolution neural network. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3/W6, 631–635.

- [13] Ritharson, S., et al. (2023). DeepRice: A deep learning framework for rice disease classification. *Expert Systems with Applications*, 213, 118834. <https://doi.org/10.1016/j.eswa.2022.118834>
- [14] Mukherjee, P., et al. (2021). Deep learning for plant disease detection: A review. *Computers and Electronics in Agriculture*, 183, 105807. <https://doi.org/10.1016/j.compag.2021.105807>
- [15] Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259. <https://doi.org/10.1016/j.neunet.2018.07.011>
- [16] Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6, 27. <https://doi.org/10.1186/s40537-019-0192-5>
- [17] Junaidi, A., Hashim, S. Z. M., Othman, M. S. B., Mohamad, M. M. B., Alhussian, H., Abdulkadir, S. J., Nasser, M., & Bena, Y. A. (2025). Deep learning and Edge computing in agriculture: A Comprehensive review of recent trends and innovations. *IEEE Access*, 13, 137464–137490. <https://doi.org/10.1109/access.2025.3595621>
- [18] Nielsen, M. A., & Chuang, I. L. (2012). Quantum computation and quantum information. In Cambridge University Press eBooks. <https://doi.org/10.1017/cbo9780511976667>
- [19] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- [20] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.
- [21] Preskill, J. (2018). Quantum computing in the NISQ era and beyond. *Quantum*, 2, 79. <https://doi.org/10.22331/q-2018-08-06-79>
- [22] Schuld, M., & Petruccione, F. (2018). Supervised learning with quantum computers. Springer.
- [23] Cerezo, M., et al. (2021). Variational quantum algorithms. *Nature Reviews Physics*, 3(9), 625–644. <https://doi.org/10.1038/s42254-021-00348-9>
- [24] Benedetti, M., et al. (2019). Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4), 043001.
- [25] McClean, J. R., et al. (2018). Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9, 4812.
- [26] Abbas, A., et al. (2021). The power of quantum neural networks. *Nature Computational Science*, 1, 403–409.
- [27] Mari, A., et al. (2020). Transfer learning in hybrid classical–quantum neural networks. *Quantum*, 4, 340.
- [28] Bergholm, V., et al. (2018). PennyLane: Automatic differentiation of hybrid quantum–classical computations. *arXiv preprint arXiv:1811.04968*.
- [29] Z. Al Muntazhar, D. R. Sulistyaningrum, B. Setiyono and R. N. R. Wijaya, "Hybrid Quantum Neural Network for Rice Leaf Disease Classification," 2024 International Conference on Electrical, Communication and Computer Engineering (ICECCE), Kuala Lumpur, Malaysia, 2024, pp. 1-6, doi: 10.1109/ICECCE63537.2024.10823437
- [30] Mari, A., Bromley, T. R., Izaac, J., Schuld, M., & Killoran, N. (2020). Transfer learning in hybrid classical-quantum neural networks. *Quantum*, 4, 340. <https://doi.org/10.22331/q-2020-10-09-340>
- [31] Haque, M.E., Rahman, A., Junaed, I., Hoque, S.U., & Paul, M. (2022). Rice Leaf Disease Classification and Detection Using YOLOv5. *ArXiv*, abs/2209.01579.
- [32] Ahmed, A. A., & Reddy, G. H. (2021). A Mobile-Based system for detecting plant leaf diseases using deep learning. *AgriEngineering*, 3(3), 478–493. <https://doi.org/10.3390/agriengineering3030032>
- [33] Y. Cui, M. Jia, T. -Y. Lin, Y. Song and S. Belongie, "Class-Balanced Loss Based on Effective Number of Samples," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 9260-9269, doi: 10.1109/CVPR.2019.00949.
- [34] Aaronson, S. Read the fine print. *Nature Phys* 11, 291–293 (2015). <https://doi.org/10.1038/nphys3272>
- [35] Sim, S., Johnson, P. D., & Aspuru-Guzik, A. (2019). Expressibility and entangling capability of parameterized quantum circuits. *Advanced Quantum Technologies*, 2(12), 1900070. <https://doi.org/10.1002/qute.201900070>
- [36] McClean, J. R., Romero, J., Babbush, R., & Aspuru-Guzik, A. (2016). The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2), 023023. <https://doi.org/10.1088/1367-2630/18/2/023023>
- [37] Liu, Y., Arunachalam, S. & Temme, K. A rigorous and robust quantum speed-up in supervised machine learning. *Nat. Phys.* 17, 1013–1017 (2021). <https://doi.org/10.1038/s41567-021-01287-z>
- [38] Kerenidis, I., Landman, J., Luongo, A., & Prakash, A. (2020). Quantum convolutional neural networks. *International Conference on Learning Representations (ICLR) Workshop*.
- [39] Anand, K., Jain, B., Mittal, H., & Yadav, V. K. (2024). QEFS: A novel plant disease prediction approach using quantum-inspired evolutionary feature selection. *Applied Intelligence*, 55(2). <https://doi.org/10.1007/s10489-024-05990-1>
- [40] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- [41] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J Big Data* 6, 60 (2019). <https://doi.org/10.1186/s40537-019-0197-0>

- [42] Kim, K. G. (2016). Book review: Deep Learning. *Healthcare Informatics Research*, 22(4), 351. <https://doi.org/10.4258/hir.2016.22.4.351>
- [43] Sweke, R., Wilde, F., Meyer, J., Schuld, M., Faehrmann, P. K., Meynard-Piganeau, B., & Eisert, J. (2019). Stochastic gradient descent for hybrid quantum-classical optimization. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1910.01155>
- [44] Howard, J. and Ruder, S. (2018) Universal language model fine-tuning for text classification, *arXiv.org*. Available at: <https://arxiv.org/abs/1801.06146> (Accessed: 19 January 2026).
- [45] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1412.6980>
- [46] Loshchilov, I., & Hutter, F. (2016). SGDR: Stochastic Gradient Descent with Warm Restarts. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1608.03983>
- [47] Prechelt, L. (2012). Early Stopping — But When?. In: Montavon, G., Orr, G.B., Müller, K.R. (eds) *Neural Networks: Tricks of the Trade*. Lecture Notes in Computer Science, vol 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_5
- [48] Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- [49] Fawcett, T. (2005). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- [50] Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29–36. <https://doi.org/10.1148/radiology.143.1.7063747>
- [51] Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7), 1895–1923. <https://doi.org/10.1162/089976698300017197>
- [52] Powers, D. M. W. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv (Cornell University)*, 2(1), 37–63. <https://doi.org/10.48550/arxiv.2010.16061>
- [53] Pretty, J., Toulmin, C., & Williams, S. (2011). Sustainable intensification in African agriculture. *International Journal of Agricultural Sustainability*, 9(1), 5–24. <https://doi.org/10.3763/ijas.2010.0583>
- [54] Efron, B., & Tibshirani, R. (1994). An introduction to the Bootstrap. <https://doi.org/10.1201/9780429246593>
- [55] Ritharson, P. I., Raimond, K., Mary, X. A., Robert, J. E., & J. A. (2023). DeepRice: A deep learning and deep feature based classification of Rice leaf disease subtypes. *Artificial Intelligence in Agriculture*, 11, 34–49. <https://doi.org/10.1016/j.aiia.2023.11.001>
- [56] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1502.03167>