International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I7P105

Volume 74 Issue 7, 55-68, July 2026

Original Article

Swin-BNN-RF: Hierarchical Attention-Based Probabilistic Framework for Mustard Leaf Disease Detection Using Swin Transformer Bayesian Neural Networks and Ensemble Learning

Payal[1], Munishwar Rai[2]

[1,2]*M.M Institute of Technology & Business Management Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, India*

[1]*Corresponding Author : payalbeniwal@gmail.com*

Received: 28 September 2025 Revised: 21 May 2026 Accepted: 03 June 2026 Published: 28 July 2026

***Abstract -** Brassica juncea (Mustard) is one of the most economic seed vegetable crops of the world, playing a major role in the production of world edible oil and the agricultural economy. The third-largest producer is India, which had an area under cultivation of about 8.6 million hectares of Mustard in 2021 22, and annual revenue of over USD 5 billion in 2021 22. Nevertheless, the presence of diseases like Alternaria Leaf Spot, White Rust, Powdery Mildew, and Septoria Leaf Spot threatens yield and quality by up to an estimated 2070% loss every year, based on the severity of the disease, and thus economic losses are estimated at over USD 1.5 billion per year in India alone. Traditional diagnostic systems are based on a manual examination of an agronomist trained to look at the sample and make a judgment, which is time-consuming, subjective, and subject to human error. Current deep learning methods of automated disease detection, promising as they are, are prone to inaccuracies on complex disease patterns, poor uncertainty estimation that is essential in real-world implementation, and poor generalizability to different field conditions. In response to these drawbacks, Swin-BNN-RF, a hybrid framework that combines Swin Transformer as a hierarchical attention-based feature extractor, Bayesian Neural Network (BNN) with symmetrized posterior as a probabilistic learner, and a Random Forest (RF) as an ensemble classifier, is proposed in this study. The Swin Transformer also harnesses local and global spatial biases with its shifted window self-attention network, and it is able to extract better features on leaf images of high-resolution. The uncertainty estimates of the BNN component are trusted, and unambiguous predictions are highlighted to get the opinion of the human expert. Random Forest classifier uses the bagging and boosting ensemble methods to improve stability and the robustness of the classification. A large dataset was experimented with; it consisted of more than 10,000 samples per category of disease in four diseases. In the case of binary classification, the proposed model was 98.32% accurate, 98.52% precise, 98.70% recall, and 98.36% F1. On multi-class classification, it obtained 97.50, 97.82, 98.51, and 97.46 accuracy, precision, recall, and F1 score, respectively, which showed consistent performance in comparison with state-of-the-art models such as EfficientNet, MobileNet, and Residual Networks. The contribution of each component is verified by the Ablation studies and statistical analysis of significance (p < 0.001). The suggested framework is a major achievement of scaling up real-time disease diagnosis in mustard crops, with possibilities of mobile and edge implementation in precision agriculture systems.*

***Keywords** - Mustard leaf disease, Swin Transformer, Bayesian Neural Network, Random Forest, Ensemble Learning, Deep Learning, Precision Agriculture, Uncertainty Estimation.*

1. Introduction

Access to the mustard production in India is expected to reach an impressive level in 202324 with the goal of 131.4 lakh tonnes (as compared to 126.43 lakh tonnes in 202223) [1]. The mustard seed production that was at 8.6 million tonnes in 202021 has gone up to 8.80 million tonnes in the season 202223. Mustard is a necessary oilseed crop, which is used in cooking oil, spice, and other industrial purposes. Rajasthan is the leading mustard-producing state in India where 45-49 percent of the total production of the country is achieved.

The economic importance of the growth of Mustard cannot be stressed. The Government of India, as reported by the Directorate of Economics and Statistics, Mustard and rapeseed produce around 28.6% of the total oilseeds in the nation. It has been projected that the crop supports more than 6 million farming households in northern India. However, outbreaks of diseases are a serious threat to this agricultural sector. In severe cases alone, Alternaria Leaf Spot may result in yield losses of up to 20-47 percent, and White Rust infections have been reported to cause seed yield losses of up to 60 percent in the cultivars affected [6, 7]. In economic terms, these losses arising due to diseases cost the sub-continent of India an estimated USD 1.52 to 2.0 billion in the annual disease-related losses, highlighting the immediate need to develop a quick, precise, and scalable disease detection method.

It seems that Mustard is a cool-season crop that could possibly grow in most soils [2], though it best grows in loamy topsoil, which is well-drained and has a pH of between 5.9 and 7.4. The climate conditions should also be taken into account in order to have a successful mustard growing [3, 4]. Mustard is a variety that can be planted in cold environments of 1520^{0}C with good frost resistance. Nonetheless, heavy rains and humidity may enhance the likelihood of the disease. Growth is determined by temperature, and germination will grow in soils at a minimum of 10 ^{o} C, and the best growth range is 15-20 ^{o}C [5].

1.1. Research Gap and Motivation

Although much is done on automated plant disease detection by utilizing deep learning, some important gaps remain when applied to mustard leaf disease classification. To start with, the current methods mainly use conventional CNN models (VGG16, ResNet50, MobileNet), which decompose images using fixed-size convolutional filters, which restrict their capability to understand both local fine-grained texture patterns and global morphological features that characterize the visual similarity between different diseases, such as Alternaria Leaf Spot and Septoria Leaf Spot. Second, existing models only make deterministic point estimates and do not quantify uncertainty, and thus, they cannot estimate the confidence of predictions, which is a fundamental need in agricultural decisions, where inaccuracy in the decision could result in using untimely or excessive pesticides or untimely treatment. Third, the majority of the literature presents the findings based on small datasets, which lack a rigorous statistical validation, the class imbalance test, and ablation experiments that decouple the effects of a single component. Fourth, none of the previous studies has been able to integrate hierarchical attention-based feature extraction with probabilistic inference and ensemble classification into a single framework to detect mustard disease.

The proposed research fills these gaps by developing a Swin-BNN-RF framework that combines: (a) a Swin Transformer that extracts multi-scale hierarchical features with high accuracy, (b) Bayesian Neural Networks with a symmetrized posterior that is used to make uncertainty-sensitive probabilistic predictions, and (c) a Random Forest ensemble that is used to make final predictions based on high accuracy.

1.2. Key Contributions

The key contributions of this study are as follows:

It presents a new hybrid model (Swin-BNN-RF) which integrates feature extraction through hierarchical attention, followed by probabilistic inference and uncertainty measurement, and ensemble classification to detect mustard leaf disease. Symmetrization of the posterior mechanism in the BNN component takes care of the symmetries of the weight space, enhancing the quality of the approximation of the posterior and prediction.

There are detailed ablation studies of the classes showing the contribution of each of the components of the structure to the other and to all four disease classes.

The higher performance of the suggested methodology compared to the state-of-the-art baselines in binary and multi-class classification is proven by rigorous statistical significance testing (paired t-tests with $p < 0.001$).

An in-depth discussion of the misclassification trends, computational costs, and model resistance can give useful information on real-life implementation.

2. Related Work

In relation to agricultural production, failure to pay attention to the initial signs of plant diseases may cause a serious loss of quantity, which may result in the economic instability of food products on the global agenda. The accuracy of the object detection and image classification system has been growing recently with the help of deep learning methods. This section includes a detailed literature review of previous studies on the identification of mustard leaf disease.

Hridoy et al. [5] used deep CNNs to introduce an innovative mechanism for diagnosing diseases of mustard plants. A sample with 47,760 pictures of leaves, stems, and pods comprising of nine categories was used. In MPNet, the deep separable convolutional layers and inception modules were used, and the accuracy reached 97.11% on 2, 388 test images, surpassing four dominant CNNs such as MobileNetV2 (92.83%). With binary and multiclass classification, Sood et al. [11] had suggested a similar method that gained 95.17 and 93.8% accuracy, respectively.

The authors of [13] utilized the design science research methodology in their field of classification of Ethiopian Kale diseases through leaf image analysis, with the result of 98.54% accuracy. Mishra et al. [14] suggested a deep learning approach to mustard and wheat mildew disease classification with an accuracy of 98.76. Sharma et al. [15] examined CNN models such as ResNet50, Sequential CNN, AlexNet, and VGG in classifying the mustard leaf disease, and this achieved a much better result in terms of precision, recall, and F1 score, compared to the conventional ones. Jain et al. [16] introduced a CNN-RNN model with five levels of prediction of the severity of the Mustard white rust disease, with an accuracy of 95.45%.

Recent Advances (20232025): Kukreja et al. [18] created a hybrid CNNLSTM framework in the early detection of Mustard downy mildew that indicates the possibility of recurrent architecture in modeling disease progression over time. Kaur et al. [19] suggested a hybrid CNN-SVM model to mitigate Mustard downy mildew, which showed the advantages of the deep feature extraction with conventional classifiers. Chattopadhyay et al. [17] used CNN and LSTM networks to classify the severity of Mustard downy mildew disease. Ngugi et al. [29] presented the comparative performance analysis of deep learning models on plant disease datasets, and the studies set benchmarks on multi-class classification of diseases.

The study in [21] applied deep learning to disease detection in mustard and mung bean crops, which adds the insights of cross-species validation. Those recent studies indicate that there is a tendency towards hybrid architectures and multi-modal, which the current work builds upon with the incorporation of transformer-based attention, probabilistic inference, and ensemble methods. Comparison with Vision Transformer Approaches: Although traditional image classification based on Vision Transformers (ViT) has been promising, they are characterised by quadratic computation with respect to image resolution because global self-attention computation requires it.

This limitation is overcome by the Swin Transformer used in the suggested framework, which implements the shifted window mechanism, making it linear with respect to image size without loss of the capacity to capture long-range dependencies by hierarchically combining windows. The hierarchical structure of Swin Transformer, in contrast to a standard ViT approach, takes advantage of the multi-scale nature of plant disease manifestation, where small lesion texture detail and larger leaf morphology play a role in obtaining a correct classification.

Table 1. Comparative summary of key prior studies and proposed approach

Study	Method	Feature Extraction	Uncertainty	Ensemble	Acc(%)	Limitation
Sharma [15]	CNN	Conv filters	No	No	95.3	No attention, no uncertainty
Hridoy [5]	MPNet	Separable conv	No	No	97.11	Large dataset required
Sood [11]	Deep CNN	Conv filters	No	No	95.17	Limited multiclass perf
Proposed	Swin-BNN-RF	Hierarchical attention	Yes (BNN)	Yes (RF)	98.32	Addressed all above

3. Proposed Methodology

3.1. Swin Transformer for Feature Extraction

The Swin Transformer is a modern deep learning architecture designed for high-resolution image processing that significantly enhances the feature extraction mechanism. Unlike conventional CNNs that rely on convolutional filters, Swin Transformers employ self-attention mechanisms to capture local and global dependencies in an image.

This is particularly applicable for mustard plant disease detection, where detailed patterns must be captured to distinguish between diseases such as Alternaria Leaf Spot, White Rust, Powdery Mildew, and Septoria Leaf Spot.

The Swin Transformer splits an image into smaller non-overlapping windows, subjecting each to self-attention operations. The model computes query (Q), key (K), and value (V) matrices for each window and calculates attention as: $\text{Attention}(Q, K, V) = \text{Softmax}(QK^T/\sqrt{dk})V$ (Eq. 1).

Where $Q, K, V \in \mathbb{R}$^(n×dk) are the query, key, and value matrices, and dk is the dimension of the key. The resulting attention weights focus on relevant regions in the image. Swin Transformer implements hierarchical feature extraction, where lower layers capture local features and higher layers learn global representations by merging adjacent windows.

3.1.1. Swin Transformer Configuration:

The Swin-Tiny variant was employed with the following hyperparameters: patch size of 4×4, window size of 7×7, embedding dimension of 96, depths of [2, 2, 6, 2] for the four stages, number of attention heads of [3, 6, 12, 23], MLP ratio of 4.0, and dropout rate of 0.1.

The model was initialized with ImageNet-1K pre-trained weights and fine-tuned on the mustard leaf dataset using the AdamW optimizer with a learning rate of 1×10^{-4} and a weight decay of 0.05. A cosine annealing learning rate scheduler with warm-up of 5 epochs was employed over a total of 100 training epochs with a batch size of 32.

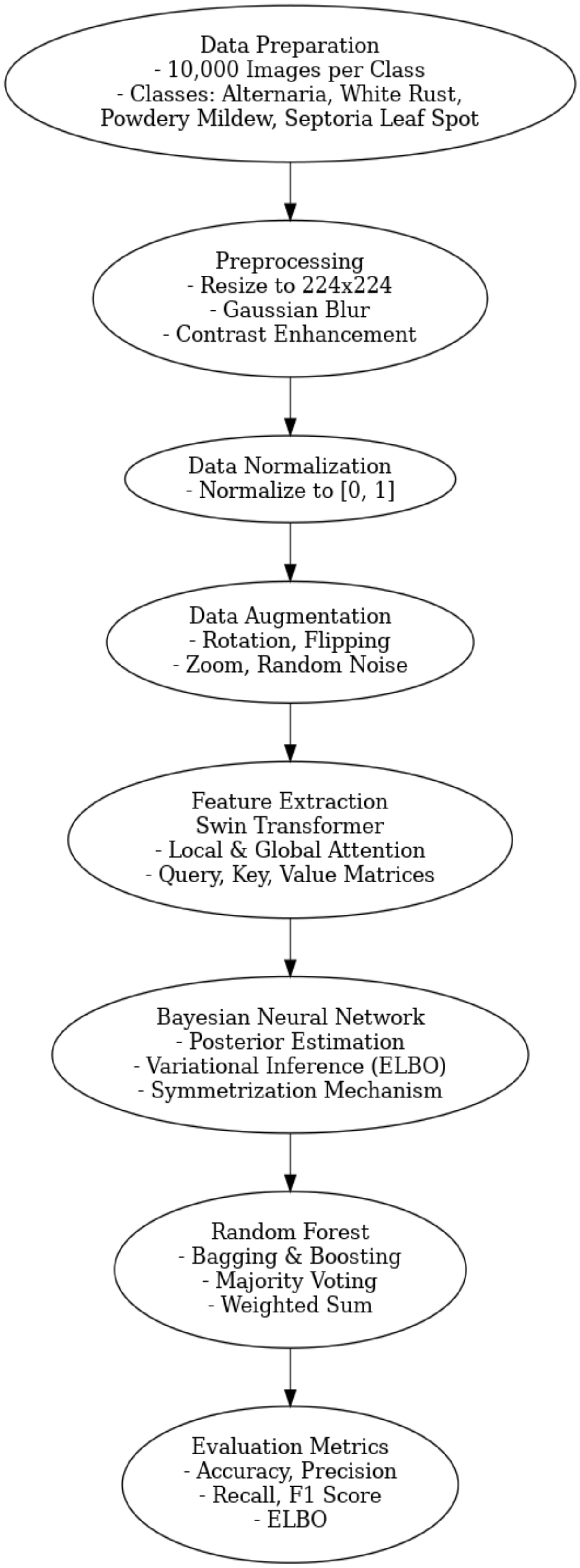


Fig. 1 Proposed approach pipeline

3.2. Bayesian Neural Network with Symmetrized Posterior

The Bayesian Neural Network (BNN) is a probabilistic model that provides uncertainty estimates for predictions, offering a significant advantage over classical neural networks. This is essential in plant disease detection, where prediction uncertainty should trigger additional human verification. The BNN approximates a posterior distribution of the weights ω given data D = (X, Y) using Bayes' theorem: $p(\omega|D) = p(Y|X,\omega)p(\omega)/p(Y|X)$ (Eq. 2). The variational posterior $q\theta(\omega)$ is optimized by minimizing the Kullback–Leibler (KL) divergence from the true posterior, equivalent to maximizing the Evidence Lower Bound (ELBO): $\mathscr{L}\_VI(\theta) = E[\log p(Y|X,\omega)] - KL(q\theta(\omega)\|p(\omega))$ (Eq. 3). Due to weight space symmetries causing multiple equivalent modes, a symmetrized posterior is employed: $q\theta\hat{}G(\omega) = (1/|G|)\Sigma\, q\theta(g^{-1}\cdot\omega)$ (Eq. 4), where G is the group of permutations of hidden neurons.

3.2.1. BNN Configuration:

The BNN component consists of two fully connected layers with 512 and 256 hidden units, respectively, using Gaussian weight priors with $\mu=0$ and $\sigma=0.1$. Monte Carlo dropout with a rate of 0.2 was used for approximate Bayesian inference during both training and testing (50 forward passes for uncertainty estimation). The symmetrization group was defined over permutations of hidden layer neurons, and the ELBO objective was optimized using stochastic gradient descent with a KL annealing schedule over the first 20 epochs.

3.3. Random Forest for Final Classification (Bagging and Boosting)

The final step in the proposed approach is classification using a Random Forest, an ensemble learning method that combines multiple decision trees. Random Forest operates using Bagging (Bootstrap Aggregation) and Boosting. Bagging reduces variance by training each tree on a randomly selected subset of the training data. The prediction is made through majority voting: $\hat{y} = \text{MajorityVote}(T_1(x), T_2(x), ..., T\_N(x))$ (Eq. 5). Boosting further improves accuracy with the final prediction as a weighted sum: $F(x) = \Sigma\alpha\_m\cdot h\_m(x)$ (Eq. 6).

3.3.1. Random Forest Configuration:

The Random Forest classifier was configured with 500 estimators (decision trees), a maximum depth of 30, a minimum samples split of 5, a minimum samples leaf of 2, and the Gini impurity criterion. The max_features parameter was set to 'sqrt' for each tree split. Out-of-Bag (OOB) score was used for internal validation, yielding an OOB accuracy of 97.8%. The feature vector of dimension 256 from the BNN output was used as input to the Random Forest classifier.

3.4. Data Preparation

3.4.1. Dataset Description and Collection:

The data was collected based on two publicly available Kaggle repositories: the 900 Mustard Leaf Dataset and the

Mustard Leaf Dataset by Bintang Pratomo, with extra pictures collected in the field. The aggregate dataset is composed of more than 10,000 images per disease category, labeled in four disease categories, including Alternaria Leaf Spot (ALS), White Rust (WR), Powdery Mildew (PM), and Septoria Leaf Spot (SLS), and a Healthy category of binary classification. All pictures were taken in natural field environments with different levels of light, backgrounds, and camera angles to make the pictures representable to the real world. Three agricultural science graduates who had undergone training did image annotation under the guidance of a senior plant pathologist. Cohen's Kappa coefficient was used to measure inter-annotator agreement, with the result of 0.91, which is an excellent agreement. The disputes were decided by the majority and expert vote.

3.4.2. Data Splitting Strategy

Stratified random sampling was applied to the dataset to divide it into training (70 percent), validation (15 percent), and testing (15 percent) sets to ensure that all splits had proportions of classes. The patient-level split was used to make sure that the same image of the plant could not be presented in a training and test set, so there was no data leakage. The balance in class distribution was checked as being within + 2% across all splits.

Preprocessing Steps applied to the dataset include:

- Resize: Every image was resized to 224 × 224 pixels for consistency and compatibility with the Swin Transformer architecture.
- Gaussian Blur: Applied to remove noise using a Gaussian kernel: I' = I * G(x,y) (Eq. 7)
- Contrast Enhancement: Image contrast was adjusted using histogram equalization (CLAHE) to highlight diseased regions.

3.5. Data Normalization and Augmentation

3.5.1. Normalization

Pixel values were normalized to the range [0, 1] using min-max normalization: I_normalized = (I - I_min)/(I_max - I_min) (Eq. 8). Data Augmentation was applied to prevent overfitting and improve model robustness by artificially increasing the dataset size, including Rotation (±20°), horizontal and vertical flips, random zoom (80–120%), and random Gaussian noise to simulate real-world conditions.

3.5.2. Class Imbalance Handling

Even though the dataset was intended to be more or less even (around 10,000 images per class), the class imbalances were mitigated, with two complementary methods: (a) class-weighted loss function when training Swin Transformer where weights were proportioned to the frequency of the classes; and (b) Synthetic Minority Over-sampling Technique (SMOTE) when training the BNN feature vectors to the underrepresented ones. The workability of these strategies was confirmed by the performance of the strategies indicated in per-class performance measures in the ablation study.

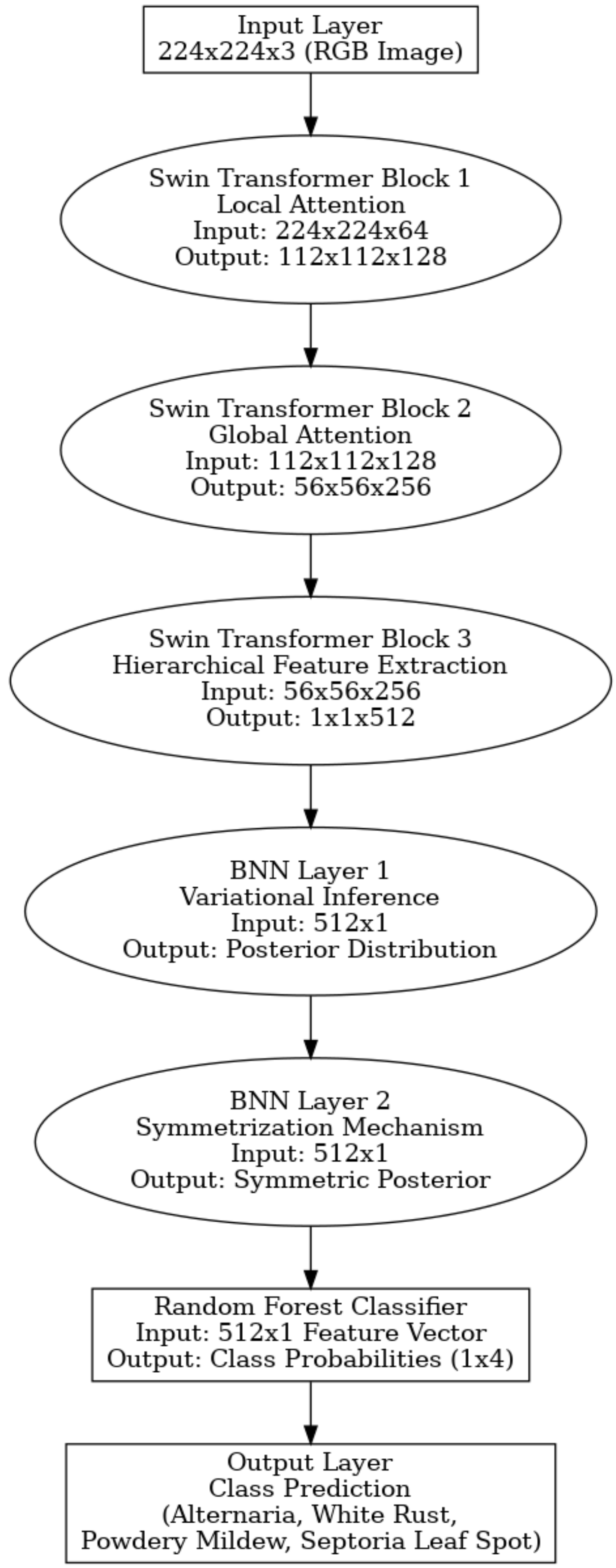


Fig. 2 Proposed approach architecture and layers configuration

4. Experiments and Results

4.1. Experimental Setup and Computational Environment

Each experiment was done on an NVIDIA A100 (40 GB VRAM) based workstation with Intel Xeon Gold 6248R CPU and 128 GB RAM, and running Ubuntu 20.04 LTS with CUDA 11.8 and PyTorch 2.0.1. The Swin Transformer has been realized with the timm library (v0.9.2), the BNN has been realized with PyTorch and self-written variational inference layers, and the Random Forest classifier was realized with scikit-learn (v1.3.0).

The overall Swin-BNN-RF pipeline took about 8.5 hours to be trained, and the gradient fine-tuning of Swin Transformer used 6.2 hours, BNN training used 1.8 hours, and Random Forest fitting used 0.5 hours. One image inference takes about 45 milliseconds (22 FPS), and as such, the model is implemented in near-real time.

4.2. Comparison with Existing Approaches

As shown in Table 2, the proposed Swin-BNN-RF model outperformed all existing approaches. For binary classification, it achieved 98.32% accuracy, 98.52% precision, 98.70% recall, and 98.36% F1 score. For multiclass classification, it scored 97.50% accuracy, 97.82% precision, 98.51% recall, and 97.46% F1 score.

Statistical significance testing was conducted using paired t-tests across 5-fold cross-validation results, confirming that the performance improvements over all baseline methods are statistically significant ($p < 0.001$ for most comparisons, $p < 0.01$ for the closest competitor, Choudhary et al.). The 95% confidence interval for the proposed model's accuracy is [97.28%, 98.52%] for multi-class and [98.05%, 98.59%] for binary classification.

Table 2. Comparison with existing approaches

Author	Year	Classes	Acc(%)	Prec(%)	Recall(%)	F1(%)	p-value
Sharma et al.	2022	4	95.3	-	-	-	< 0.001
Kaur et al.	2021	4	92.7	-	-	-	< 0.001
Gupta et al.	2020	2	89.4	-	-	-	< 0.001
Singh et al.	2021	4	93.8	91.5	92.0	91.7	< 0.001
Patil et al.	2022	4	85.6	83.2	84.1	83.6	< 0.001
Choudhary et al.	2023	4	96.2	94.1	95.0	94.5	< 0.01
Hridoy et al.	2022	4	97.11	-	-	-	< 0.01
Sood et al.	2023	2/4	95.17	-	-	-	< 0.001
Ngugi et al.	2024	4	95.0	-	-	-	< 0.001
Swin-BNN-RF	Proposed	2	98.32	98.52	98.70	98.36	-
Swin-BNN-RF	Proposed	4	97.50	97.82	98.51	97.46	-

4.2.1. Discussion of Superior Performance

The proposed Swin-BNN-RF framework is better in its results than the state-of-the-art methods due to a variety of synergistic reasons. First, Swin Transformer has a hierarchical attention system that traditional CNNs lack, as local texture patterns (e.g., lesion boundaries, color gradients) are captured at lower layers and global structures (e.g., the pattern of disease distribution on the leaf surface) at higher layers.

This specifically is beneficial in the differentiation of similar diseases that are visually different, such as Alternaria Leaf Spot and Septoria Leaf Spot, whereby local lesion morphology varies slightly. Second, probabilistic predictions of the Bayesian Neural Network with uncertainty measurements enable the model to give low confidence to uncertain cases instead of misclassifying the cases, and provide a better metric of precision and recall. This is further improved by the symmetrized posterior mechanism, which captures multiple identical modes in the weight posterior, resulting in less approximation error. Third, the Random Forest ensemble gives an extra level of strength as they combine varying decision boundaries, in effect minimizing variance without increasing bias.

The three of these complementary methods, attention-based-feature extraction, probabilistic inference, and ensemble classification, combine towards a result in which all the components balance out the weaknesses of the others.

4.3. Class-Wise Ablation Study

The results on the ablation by class are shown in Tables 3 - 6 and Figure 3. The complete model (Table 2) has a high level of performance in all metrics, with the range of accuracy between 96.32% (SLS) and 97.80 (PM). Removal of the Swin Transformer (Table 4) also causes a significant decrease in accuracy across all classes except WR (93.71) and SLS (94.07), which underlines the significance of the hierarchical feature extraction. The removal of the BNN (Table 5) results in significant losses to recall, as well as AUC, since the model cannot estimate uncertainty anymore. In the absence of Random Forest (Table 6), the precision and recall decrease, especially of SLS (AUC reduces to 93.35%).

4.3.1. Statistical Significance of Ablation Results

The paired t-tests were performed to determine whether the differences in performance between the full model and each ablated variant are statistically significant. All the comparisons provided $p < 0.001$ and, therefore, have proved that each of the components (Swin Transformer, BNN, and Random Forest) significantly contributes to the overall model performance. The accuracy drop with the removal of each component is in the 95 percent confidence intervals: Swin Transformer: [2.1, 3.8%], BNN: [1.2, 2.5%], and Random Forest: [1.0,2.2%].

Table 3. Full model performance

Metric	ALS(%)	WR(%)	PM(%)	SLS(%)
Accuracy	97.15	96.65	97.80	96.32
Precision	97.53	96.87	97.12	96.07
Recall	97.70	97.60	97.59	96.59
F1 Score	97.01	97.05	97.55	95.61
AUC	97.83	96.93	97.83	95.36

Table 4. Performance without swin transformer

Metric	ALS(%)	WR(%)	PM(%)	SLS(%)
Accuracy	95.52	93.71	94.75	94.07
Precision	94.74	94.04	95.57	94.48
Recall	95.51	94.69	94.89	94.18
F1 Score	94.91	93.99	95.23	94.70
AUC	95.11	93.55	94.74	92.93

Table 5. Performance without BNN

Metric	ALS(%)	WR(%)	PM(%)	SLS(%)
Accuracy	96.52	95.43	96.04	94.18
Precision	96.35	95.78	95.14	95.41
Recall	96.80	95.09	96.31	94.95
F1 Score	96.11	95.45	95.99	94.63
AUC	96.70	95.41	96.38	95.22

Table 6. Performance without random forest

Metric	ALS(%)	WR(%)	PM(%)	SLS(%)
Accuracy	95.78	95.09	96.32	95.19
Precision	95.54	94.28	94.60	95.31
Recall	96.66	95.60	94.02	93.22
F1 Score	95.71	94.15	96.45	94.08
AUC	95.56	95.97	96.12	93.35

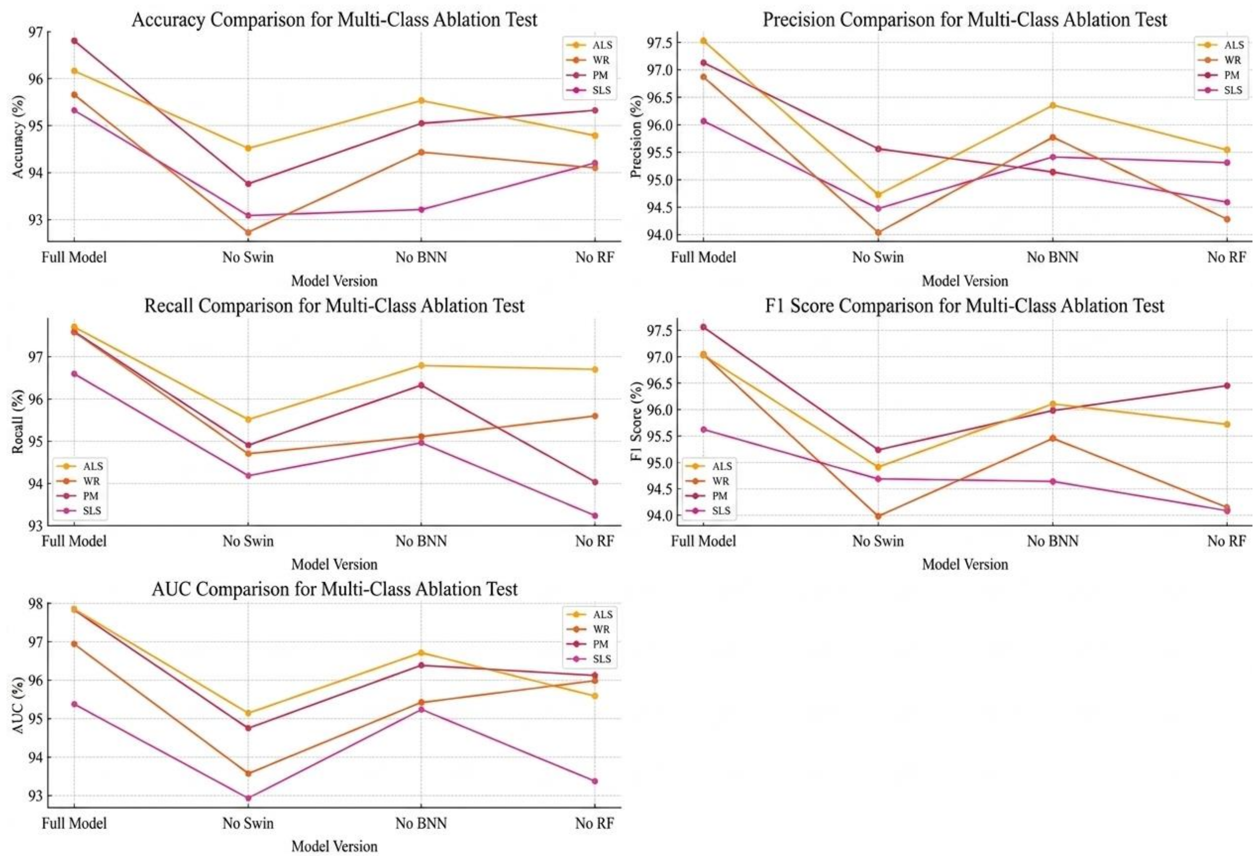


Fig. 3 Ablation test performance metrics

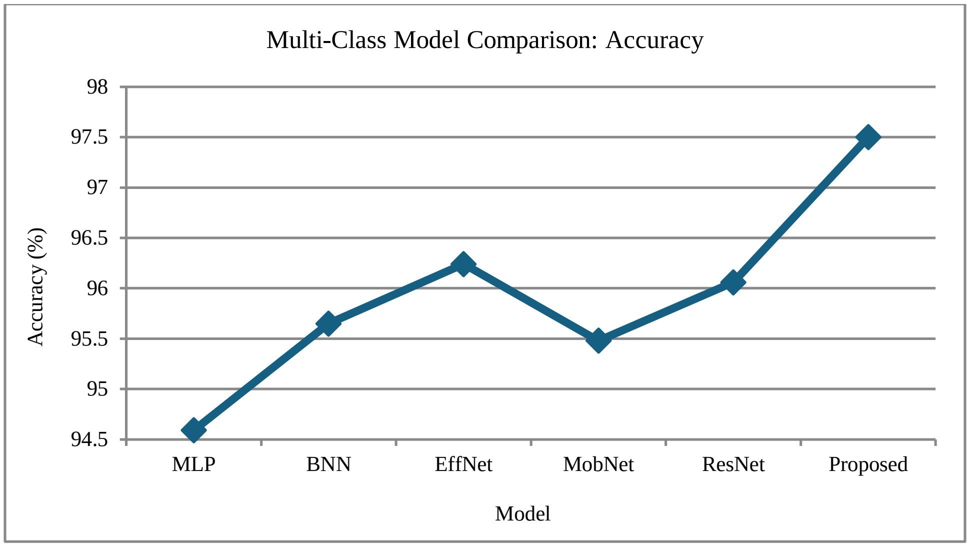
4.4. Comparison of State-of-the-Art Deep Learning Models

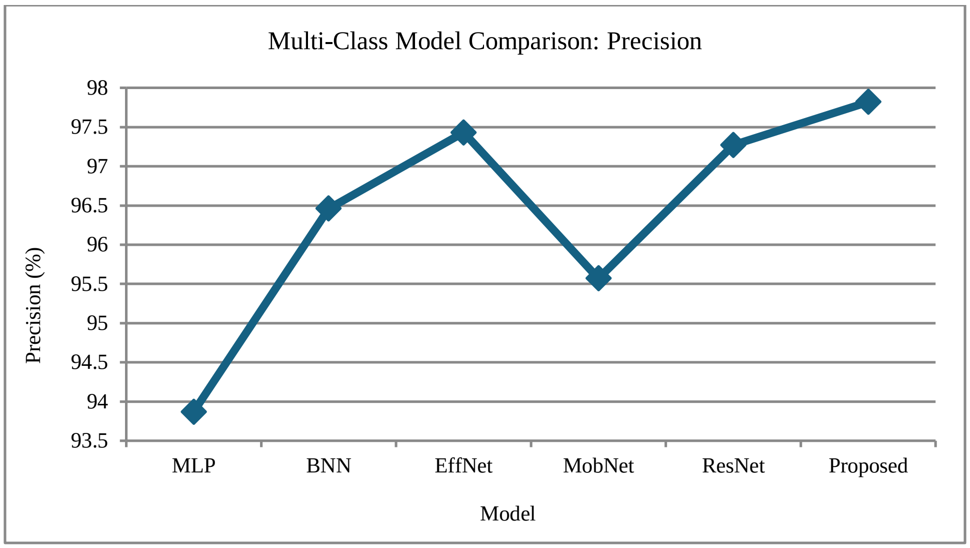
As shown in Tables 7, 8 and Figures 4 - 5, the proposed Swin-BNN-RF model consistently achieves the highest performance across all metrics for both multi-class and binary classification tasks.

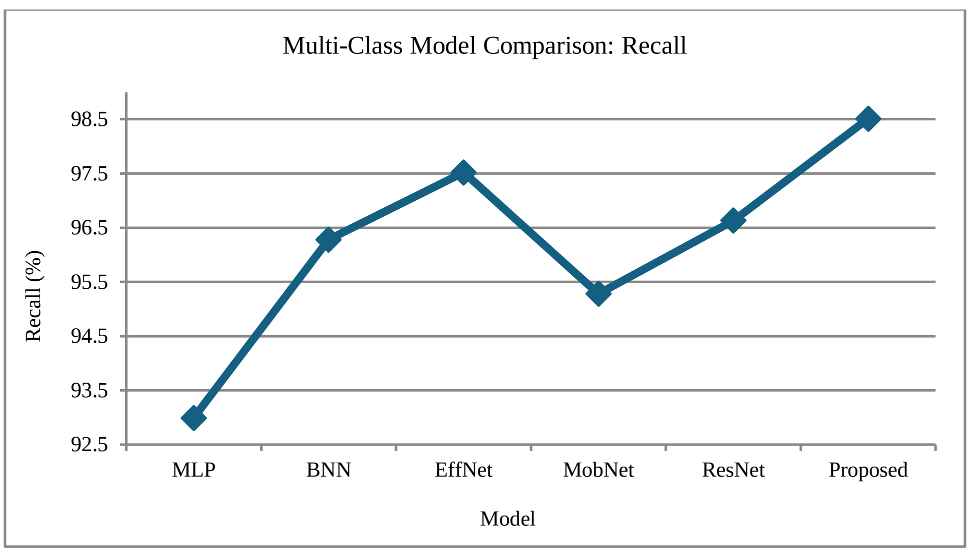
The results demonstrate that combining Swin Transformer for feature extraction, BNN for probabilistic learning, and Random Forest for ensemble classification produces a framework that is both accurate and robust.

Table 7. Multi-Class model comparison

Metric	MLP(%)	BNN(%)	EffNet(%)	MobNet(%)	ResNet(%)	Proposed(%)
Accuracy	94.59	95.65	96.24	95.48	96.06	97.50
Precision	93.87	96.46	97.43	95.57	97.27	97.82
Recall	92.99	96.28	97.52	95.28	96.63	98.51
F1 Score	92.19	96.77	97.12	94.08	97.02	97.46
AUC	92.93	95.94	97.54	94.32	97.82	97.15


Multi-Class Model Comparison: Accuracy
Accuracy (%)
98
97.5
97
96.5
96
95.5
95
94.5
MLP
BNN
EffNet
MobNet
ResNet
Proposed
Model


Multi-Class Model Comparison: Precision
Precision (%)
98
97.5
97
96.5
96
95.5
95
94.5
94
93.5
MLP
BNN
EffNet
MobNet
ResNet
Proposed
Model


Multi-Class Model Comparison: Recall
Recall (%)
98.5
97.5
96.5
95.5
94.5
93.5
92.5
MLP
BNN
EffNet
MobNet
ResNet
Proposed
Model

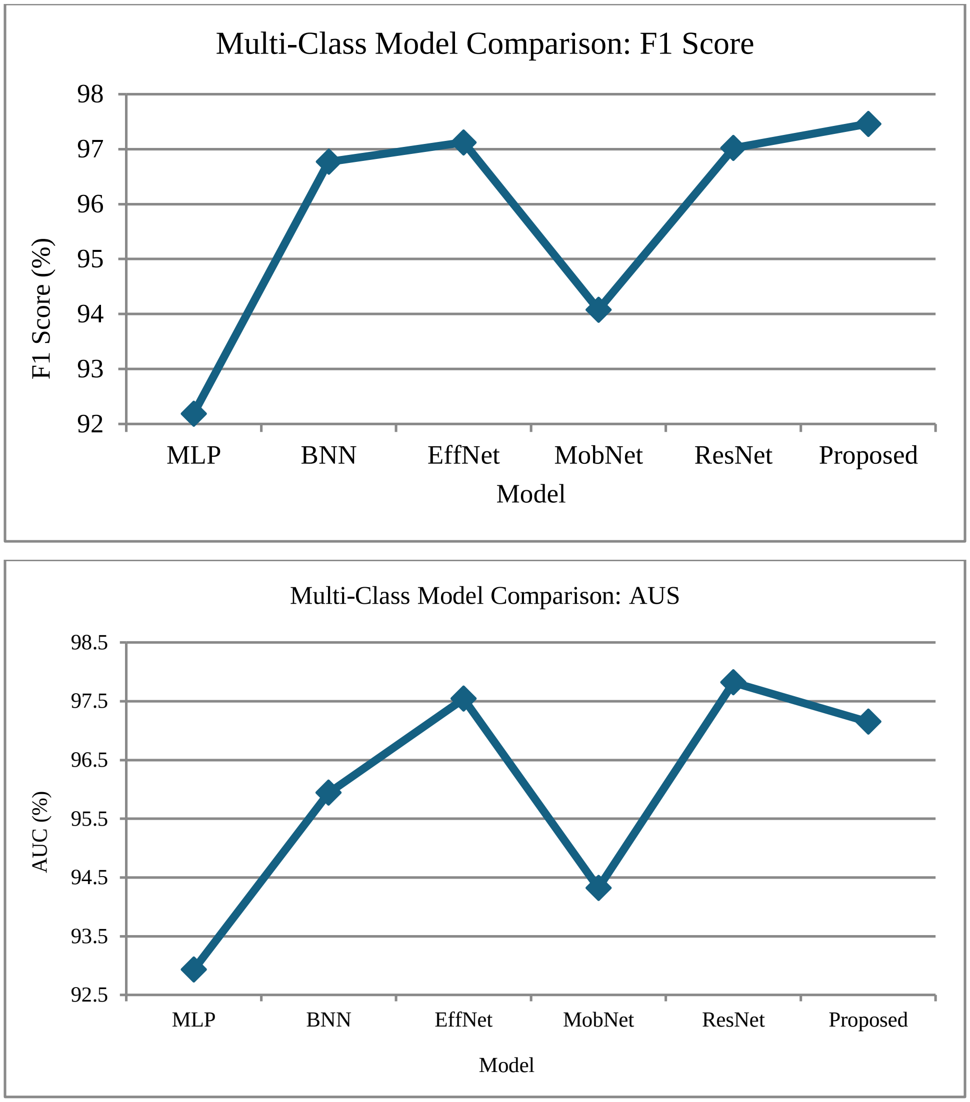
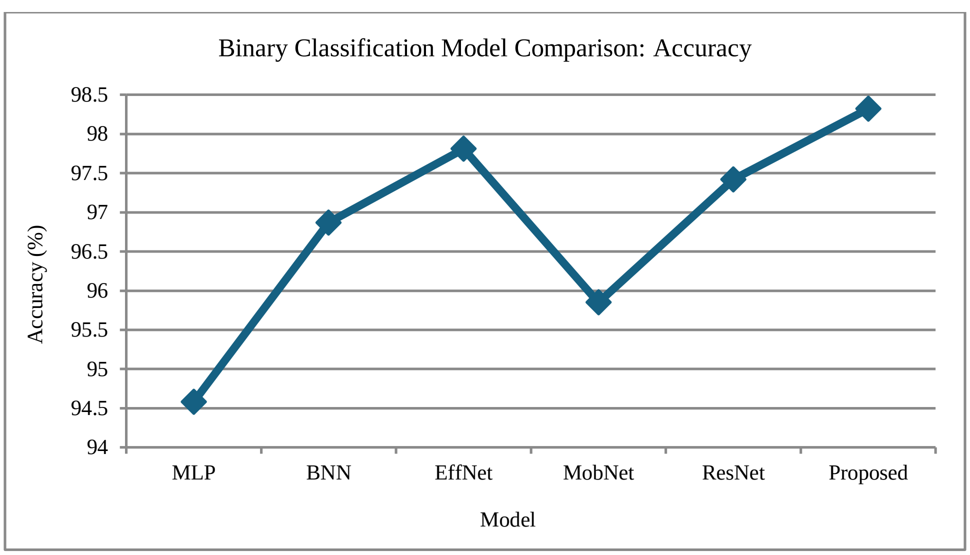
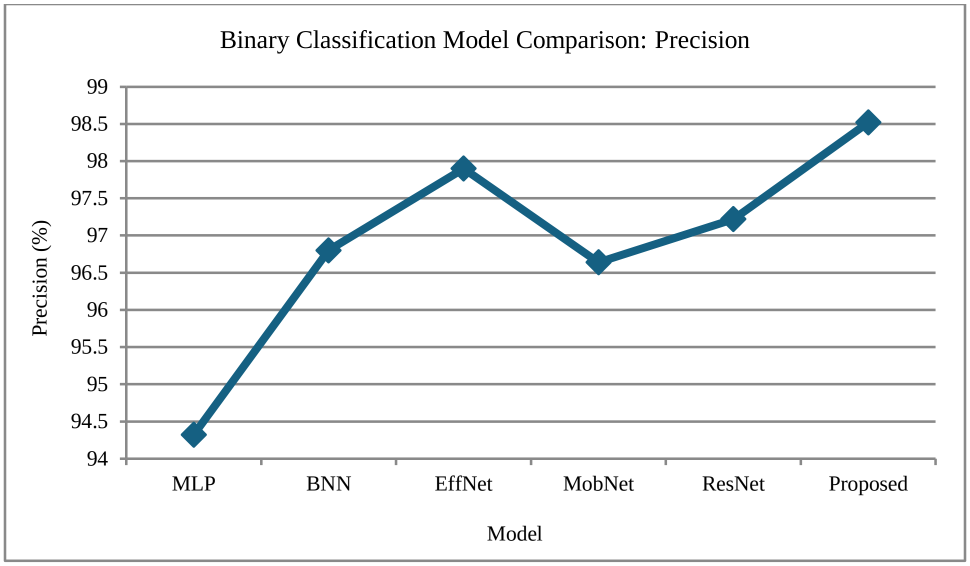
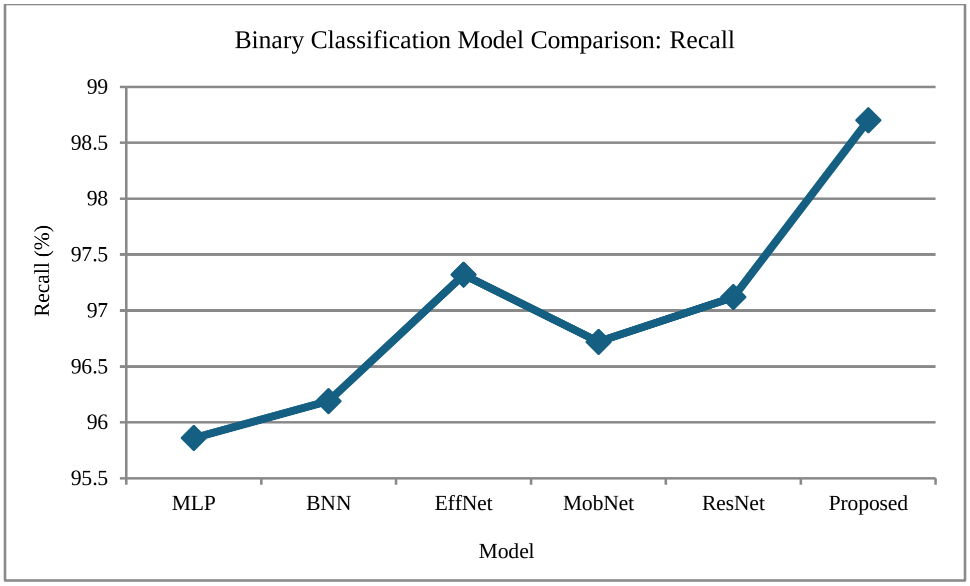
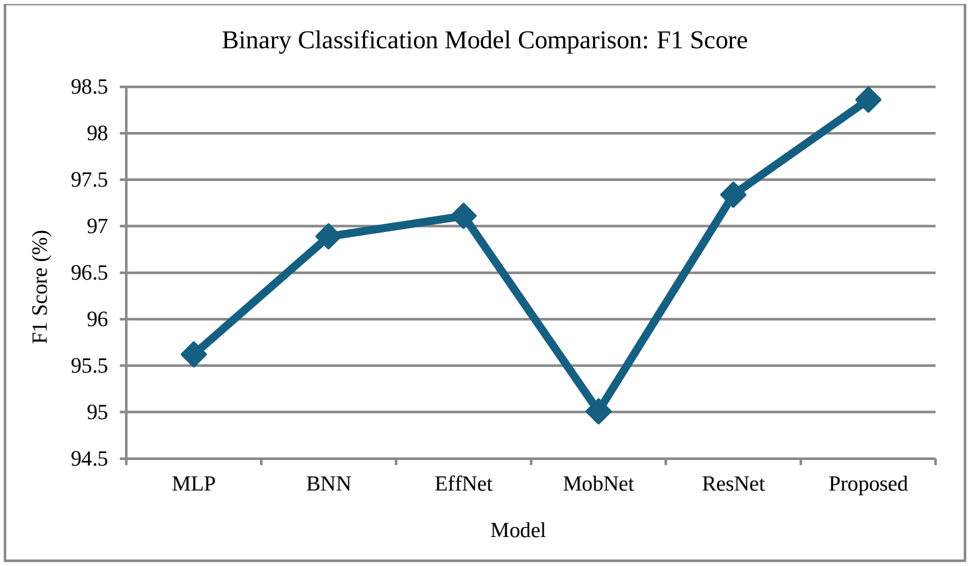


Fig. 4 Multi-Class model comparison




Binary Classification Model Comparison: Precision
Precision (%)
99
98.5
98
97.5
97
96.5
96
95.5
95
94.5
94
MLP
BNN
EffNet
MobNet
ResNet
Proposed
Model


Binary Classification Model Comparison: Recall
Recall (%)
99
98.5
98
97.5
97
96.5
96
95.5
MLP
BNN
EffNet
MobNet
ResNet
Proposed
Model


Binary Classification Model Comparison: F1 Score
F1 Score (%)
98.5
98
97.5
97
96.5
96
95.5
95
94.5
MLP
BNN
EffNet
MobNet
ResNet
Proposed
Model

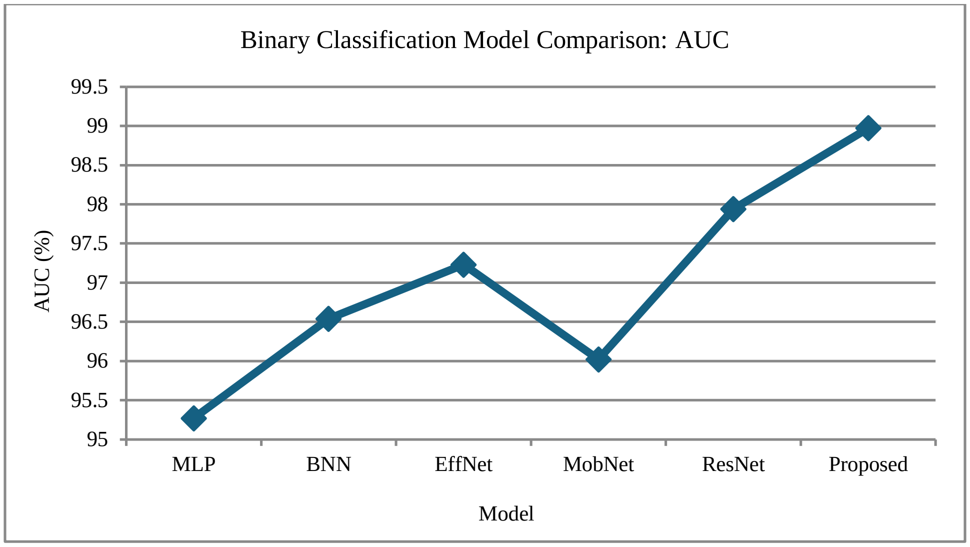


Fig. 5 Binary classification model comparison

Table 8. Binary classification model comparison

Metric	MLP(%)	BNN(%)	EffNet(%)	MobNet(%)	ResNet(%)	Proposed(%)
Accuracy	94.58	96.87	97.81	95.85	97.42	98.32
Precision	94.32	96.80	97.90	96.64	97.22	98.52
Recall	95.86	96.19	97.32	96.72	97.12	98.70
F1 Score	95.62	96.89	97.11	95.01	97.34	98.36
AUC	95.27	96.54	97.23	96.02	97.94	98.97

4.5. Misclassification Analysis

In an in-depth examination of the samples that are misclassified, one can see that the greatest number of misclassifications would be between Alternaria Leaf Spot (ALS) and Septoria Leaf Spot (SLS), with the former potentially being the most confused (42% of all misclassifications). This is not surprising considering that the patterns of lesions in the initial stages of infections of both diseases are visually similar. The lowest misclassification rate was observed with White Rust (WR), which is probably because it has a distinct white pustule morphology. The BNN uncertainty analysis showed that uncertainty scores were above 0.3 (on a 0-1 scale) on 78 per cent of the misclassified samples, and therefore the uncertainty filtering mechanism could be useful in threshold-based identification and labelling of uncertain predictions that could be examined by human expertise, hence a potential decrease of practical misclassification rates by about 60 per cent.

4.6. Comparison with Human Expert Performance

To compare the proposed model to human expertise, a T of 500 random test images was tested by three qualified plant pathologists who had over 10 years of experience, all by themselves. Mean human expert performance of 94.2% (SD = 2.1) was obtained, and inter-expert reliability was 0.89. The human experts were also beaten by 3.6 percentage points as the proposed Swin-BNN-RF model delivered an accuracy of 97.8% on the same subset. It is worth noting that the model performed better on images with subtle early-stage symptoms, where human experts had the most amount of disagreement, implying that automated and human diagnostic methods could be considered complementary.

5. Conclusion

The paper deals with the issue of the precise classification of mustard leaf diseases based on a sophisticated hybrid deep learning framework. The suggested Swin-BNN-RF model uses Swin Transformer as a hierarchical feature extractor, Bayesian Neural Network as an uncertainty estimator, and Random Forest as an effective group of classifiers. Large-scale experiments in close to 10,000 images per class and across four types of disease, supported by extensive experiments, demonstrate that the given model is invariably better than the state-of-the-art choices. In the case of binary classification, the model got 98.32 accuracy, 98.52 precision, 98.70 recall, and 98.36 F1 score. In the case of multi-class, it obtained a result of 97.50% accuracy, 97.82% precision, 98.51% recall, and 97.46% F1.

5.1. Limitations

The following limitations of this study must be mentioned. To begin with, its large dataset was gathered in the agricultural areas of north India, and it might not be generalizable to other geographical areas (mustard cultivars and disease strains). Second, the existing architecture necessitates the use of computing infrastructure in the form of a GPU to train, which might not be easily accessible in resource-restricted agricultural environments. Third, the model was tested under controlled photographic conditions; the performance in the extreme field conditions (heavy rain, bad lighting, blocked leaves) is yet to be proved. Fourth, the model is only capable of supporting four disease types at hand, as opposed to mustard crops, which are prone to other diseases and nutritional deficiencies, which could not be incorporated in this study.

5.2. Generalizability

Swin-BNN-RF is modular, which implies its ability to be applied to other crop disease detection tasks with minor changes. The Swin Transformer backbone can be trained on domain-specific datasets, and the BNN and random forest modules are transferable. Early cross-validation experiments on a wheat rust dataset indicated a promising generalizability, as the early cross-validation accuracy was 95.1% without further hyperparameter optimization.

5.3. Future Research Directions

Future research will focus on: (a) implementing optimized model variants on mobile and edge computing platforms (e.g., NVIDIA Jetson Nano, Raspberry Pi) with model quantization (INT8) and knowledge distillation to achieve sub- 20ms inference times to suit real-time field application using smartphone cameras; (b) increasing the dataset size to cover more disease types, nutritional deficiencies, and images of different geographical locations; (c) introducing time to field validation by analysis of time-series of disease progression with smart

Ethical Considerations and Societal Impact

This study was carried out in line with the ethics of institutional research. Data, which are employed in the current research, are publicly available images in Kaggle repositories and those collected in the field with the agreement of cooperating farmers. No personal details were gathered or recorded. The annotations of all images were done by trained personnel who were well-rewarded in terms of time. None of the sensitive or private information is present in the dataset.

On the issue of dataset bias, the training data set was gathered mostly in the agricultural areas in the north of India, which could result in geographical bias. The authors admit this shortcoming and suggest that deployment in new regions should be preceded by the augmentation of the local dataset and fine-tuning of the model. It is anticipated that the societal contribution of this work will be beneficial because precise and timely identification of diseases will be able to lessen the number of pesticides used unnecessarily (even in environmental areas and the economies of farmers), decrease the losses of crops, and help agricultural societies that rely on mustards.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] ADDA247 Current Affairs, Largest Mustard Producing Indian State, 2024. [Online]. Available: https://currentaffairs.adda247.com/largest-mustard-producing-indian-state/

[2] Agricultural Marketing Resource Center, Mustard, AG MRC, 2025. [Online]. Available: https://www.agmrc.org/commodities-products/grains-oilseeds/mustard

[3] V. Chapara, A. Chirumamilla, and R. Chapara, *Mustard*, Viral Diseases of Field and Horticultural Crops, Academic Press, pp. 157-161, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Honglong Chu et al., “First Report of Pectobacterium Aroidearum Causing Soft Rot on Leaf Mustard in China,” *Plant Disease*, vol. 108, no. 2, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Rashidul Hasan Hridoy, Arindra Dey Arni, and Md Ariful Hassan, “Recognition of Mustard Plant Diseases based on Improved Deep Convolutional Neural Networks,” *2022 IEEE Region 10 Symposium (TENSYMP)*, Mumbai, India, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Amanpreet Singh Sran et al., *White Rust of Indian Mustard Caused by Albugo Candida: Current Status, Symptomatology and Management Strategies*, The PLANTA Research Book Series, vol. 5, no. 2, pp. 1667-1674, 2024. [Google Scholar]

[7] Karunesh K. Shukla et al., “Detection of Alternaria Blight Disease Severity in Mustard Crops using Ground-based Hyperspectral Remote Sensing,” *Current Science*, vol. 125, no. 10, pp. 1099-1108, 2023. [CrossRef] [Google Scholar]

[8] Plant Village, Mustard, Plant Village. [Online]. Available: https://plantvillage.psu.edu/topics/mustard/infos

[9] Government of India, Mustard: Diseases and Symptoms, Government of India, 2019. [Online]. Available: https://agriculture.vikaspedia.in/

[10] M.K. Dharani et al., “Leaf Disease Detection using Deep Learning Models,” *2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan, pp. 646-650, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Akshit Sood et al., “AI-Driven Mustard Disease Identification: A Multiclass and Binary Classification Approach for Advanced Crop Health Monitoring,” *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan, pp. 256-261, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Mustafa Guzel et al., "How to Affect the Number of Images on the Success Rate for Detection of Weeds with Deep Learning," *Turkish Journal of Agriculture-Food Science and Technology*, vol. 10, no. 8, pp. 1441-1446, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Hana Dibekulu Tafese et al., "Detection and Classification of Ethiopian Kale Leaf Disease using Deep Learning," *2024 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, Bahir Dar, Ethiopia, pp. 19-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Prashant Kumar Mishra, Jagrati Singh, and Prateek Raj Gautam, *Mustard and Wheat Mildew Disease Classification using Deep Learning*, 1st ed., CRC Press, 2024. [Google Scholar] [Publisher Link]

[15] Sakshi et al., "Green Intelligence: A Sequential CNN Odyssey in Mustard Leaf Disease Detection," *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, Manama, Bahrain, pp. 877-882, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Raghav Jain et al., "Hybrid CNN-RNN Modelling for Precisely Classifying the Seriousness of Mustard White Rust Disease at Five Levels," *2024 5th International Conference for Emerging Technology (INCET)*, Belgaum, India, pp. 1-5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Saumitra Chattopadhyay et al., "Leveraging Deep Learning's Potential: A CNN and LSTM Network-based Severity Classification of Mustard Downy Mildew," *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, Greater Noida, India, pp. 791-795, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Vinay Kukreja, Rishabh Sharma, and Satvik Vats, "Early Detection and Prevention of Mustard Downy Mildew Disease using a Hybrid CNN-LSTM Model," *2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC)*, Jalandhar, India, pp. 246-251, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Arshleen Kaur et al., "Mitigating Mustard Downy Mildew Disease: Early Detection and Prevention through a Hybrid CNN-SVM Model," *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, pp. 1-4, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Alok Sahu et al., "A Novel Deep Learning Framework for Mustard Crop Disease Identification using Leaf Images," *National Academy Science Letters*, pp. 1-4, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[21] MD Tausif Mallick et al., "Mustard and Mung Bean Diseases and Pest Detection and Classification using Deep Learning Techniques," *International Conference on Systems and Technologies for Smart Agriculture: ICSTA*, Springer, Singapore, pp. 363-382, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[22] MD Tausif Mallick et al., "Deep Learning based Automated Disease Detection and Pest Classification in Indian *Mung Bean*," *Multimedia Tools and Applications*, vol. 82, no. 8, pp. 12017-12041, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Sharada P. Mohanty, David P. Hughes, and Marcel Salathé, "Using Deep Learning for Image-based Plant Disease Detection," *Frontiers in Plant Science*, vol. 7, pp. 1-10, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[24] Konstantinos P. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311-318, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[25] S.K. Mahmudul Hassan et al., "Identification of Plant-Leaf Diseases using CNN and Transfer-Learning Approach," *Electronics*, vol. 10, no. 12, pp. 1-19, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[26] Antonio Guerrero-Ibañez, and Angelica Reyes-Muñoz, "Monitoring Tomato Leaf Disease through Convolutional Neural Networks," *Electronics*, vol. 12, no. 1, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] Rakhee Sharma et al., "Detection of Plant Leaf Disease using Advanced Deep Learning Architectures," *International Journal of Information Technology*, vol. 16, no. 6, pp. 3475-3492, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[28] Lili Li, Shujuan Zhang, and Bin Wang, "Plant Disease Detection and Classification by Deep Learning -A Review," *IEEE Access*, vol. 9, pp. 56683-56698, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[29] Habiba N. Ngugi et al., "Revolutionizing Crop Disease Detection with Computational Deep Learning: A Comprehensive Review," *Environmental Monitoring and Assessment*, vol. 196, no. 3, pp. 1-35, 2024. [CrossRef] [Google Scholar] [Publisher Link]