Multimodal Deep Learning Framework for Proactive Plant Disease Diagnosis

Purshottam J. Assudani[1], V. Rama Krishna[2]

[1]*Assistant Professor, School of Computer Science and Engineering, Ramdeobaba University, India*
[2]*Assistant professor, School of Engineering, Anurag University, India*
Email: pjassudani@gmail.com

Abstract: Early diagnosis, the correct diagnosis of plant diseases is important to ensure sustainable agriculture and the minimalization of the loss of production. Traditional approaches of plant disease detection, which involve manual inspection and single modal imaging, are highly cumbersome, erroneous and lack in capturing the niche characteristics of the disease. Some recent achievements of deep learning advocate for possible automatic plant disease diagnosis; however, still most of the current models are plagued from low generalization capability, high computational cost and the issue of real time implementation. To alleviate these difficulties, this article introduces a brand-new multiple-mode deep learning framework, that combines RGB, hyperspectral and thermal imaging to take on the task of setting up precision and efficiency for plant disease detection. The described framework makes use of EfficientNet-based CNN for spatial feature extraction from RGB images, 1D-CNN for hyperspectral spectral feature learning and Vision Transformers (ViT) for learning long-range contextual dependencies. Above sensor- features are fused by Means of weighted summation methodology, dynamically adjusts contribution of per modality to Obtain endurance and accurate. To achieve real-time performance, the model is optimized via quantization, knowledge distillation and model pruning, with a substantial decrease in its computational load. The final optimal model is implemented in NVIDIA Jetson Nano to allow low-latency inference supporting high precision agriculture. The results of the experimental results show, the proposed multi-modal framework has achieved 97.8% accuracy, 96.5% precision, 95.7% recall and 96.1% score of F, all far exceed traditional deep learning models of ResNet-50, VGG-16, EfficientNet and Vision Transformers (ViT). Moreover, the framework offers inferences in 20 milliseconds, which makes it really suitable for real-time applications. Accomplishing a successful integration of multi-modal data fusion and model optimization not only increase classification performance, but also makes the solution/matter practical and deployable in real-world agricultural environment. The proposed framework provides a hopeful solution to smart farming, which provides a possibility of detecting disease early and managing effectively the crops.

Keywords: RGB Imaging, Hyperspectral Imaging, Thermal Imaging, Convolutional Neural Network (CNN), Vision Transformer (ViT), Real-Time Inference, Knowledge Distillation, Model Quantization.

1. Introduction

The backbone of global food security and economic stability according to agriculture is an activity that contributes substantially to the GDP of a lot of countries. However, plant diseases remain a major restraint in achieving highest possible agricultural productivity and exhibit considerable losses entire yield as well as economic sterling all over the world. Citing the Food and Agriculture Organization (FAO), plant diseas cause an estimated 40% annually of crop loss, impacting both small-scale farmer

and large-scale agribusiness. Traditional disease detection technologies depend on manual inspection by agricultural personnel, which are consuming time, labor and prone to human error due to variability of subjective examination. However, the urgency for automated, efficient and reliable plant disease detection has grown more pressing nowadays with meals output in addition to environmentally soothing farming treating asks getting big. Recent breakthroughs in Artificial Intelligence (AI), the Machine Learning (ML), and Deep Learning (DL) have demonstrated substantial potential in the area of automatic plant disease detection offering faster, more accurate and efficient options. Consequently, deep learning models, Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) have proved to be state-of-the-art (SOTA) for disease classification from plant images. However, in spite of their high accuracy, these models have several disadvantages to them. Initially, all the existing models are dependent on RGB images which may not bring into focus essential disease characteristics at all, especially when symptoms occur in not visible spectra such as thermal or hyperspectral bands. Moreover, the universal applicability of deep learning models is however still uncertain, as those trained on controlled data-sets, usually tend to fail in real operation in agricultural environments due to multiple light-conditions, occlusions, and background noise. Third, the challenge is in the fact that computational complexity is a significant barrier, due to resource-hungry nature of many CNN based models, this makes them less suitable for real time use on low power edge devices used in the smart farming application.

To tackle these challenges, this study introduces a new hybrid deep learning system for plant disease diagnosis, making use of multi-modal data fusion to fuse RGB, hyperspectral and thermal image for a more accurate and robust disease classification. The proposal framework integrates EfficientNet, a highly efficient CNN, for the spatial feature extraction part, and it leverages on the Vision Transformer (ViT) to capture the long-range dependencies and improve the robustness against diverse environmental conditions. Naturally, we also present a lightweight edge-friendly architecture appropriately optimized by means of quantization and knowledge distillation, which makes deployment real-time field feasible. The reliability of the suggested approach verified by numerous experiments on public and custom datasets, in which it shows better performance as regards classification results correctness, robustness, efficiency in the area of computing in comparison with the existing CNN-based and traditional ML models.

2. Literature Survey

The applications of machine learning (ML) and deep learning (DL) in plant disease detection have attracted a lot of attention thanks to the possibility of promoting more agricultural productivity through early and accurate disease diagnosis. Historically, mostly the handcrafted feature extraction techniques have been practically used with the involvement of Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Decision Trees (DT) and Random Forests (RF) in that for classification. Some studies have shown that SVM classifiers employed with colour and texture features could attain good detection of plant diseases but the performance of these models were restricted by the noise due to environmental and the changes in the real- world situations [1]. Likewise feature extraction techniques like Gray-Level Co-occurrence Matrix (GLCM) and color histograms were applied for the texture analysis and these could improve the label of the classification achievement when it is joint with SVM-based classifiers[2]. However these traditional ML models have some short comings like Scalability, Feature engineering in manual manner and do not generalize well in Unseen conditions [3].

Deep learning architectures, specifically Convolutional Neural Networks (CNNs) stepped up as a more enhancing alternative in order to escape the restriction of classic ML models. Deep CNNs researchers comfortably built full-fatplant diseases classification models which were able to automatically get features from images including without pre-processing by manual. Experiments conducted on the PlantVillage dataset showed the efficiency CNN-based architectures for the diagnosis of plant diseases obtaining more than 99% of accuracy in distinguishing between different diseases [4]. In addition, even additional improvements have been obtained from transferring learning process using pre-learned CNN models including VGG-16, ResNet-50, and InceptionV3 were pre-trained in agricultural datasets, show better result than classical machine learning technique [5]. However, due to their accuracy, but also several challenges were faced by these CNN-based models including their reliance on RGB image format data, overFITTING of controlled datasets and the high computational resources, which make them powerless for real-time applications in big enough farming [6].

In order to solve these problems, the researchers proposed multi-modal learning, combining HS, T and FL imaging with deep learning networks. Hyperspectral imaging technologies were also very potent in determining pre-mould period plant disorders since hyperspectral imaging detected spectral signifies beyond the obvious range allowing further information about plant health [7]. Experiment by using thermal imaging showed that temperature fluctuations due to stress in plants can be used as a signal for disease which rates can classification higher when associated with CNN-based features extraction [8]. Even though these developments, multi-modal learning techniques frequently have to be equipped with the specialized sensors as well as computationally intensive structures, which make them tough to use for real agricultural applications.

In addition to multi-modal fusion, more recently transformer-based architectures have been applied for the plant disease detection. Unlike CNNs, which are mostly concerned with the local featutre extraction, Vision Transformers (ViTs) make use of the self-attention mechanism allowing to feel at a glance interdisciplinary 모델 dependency within the image, increasing the level of accuracy by the complex datasets [9]. Experimental research showed that hybrid CNN-ViT architectures perform better in plant disease detection, especially under changing environmental condition compare to traditional CNNs [10]. Nevertheless, the cost and memory-intensive operations of them are still a significant limitation for the practical application in real-time agriculture of ViTs. To alleviate these challenges, some recent research studies focused on developing compatibilizing optimization methods like quantization, knowledge distillation, and pruning techniques and have allowed for more efficient transformer model deployment on edge computing, applicable in the precision agriculture [11].

Although tremendous improvement of ML and DL approaches for plant disease detection is achieved yet several major challenges still exist. Most of the existing models rely on the RGB images, which might not respond to the early stage of disease symptoms. The issues of generalization continue because the models learned from a controlled experimental data usually fail when being applied to the actual agricultural application scenes caused by illumination changes, occlusions, and background noise. Computational constraints also limit the model, as current state-of-the-art CNN and transformer models require high onto-processing power, so limit their usable in low-power edge devices hardly used in smart farming applications [12]. Besides this, data paucity and annotation cost represent a significant issue, since generating large number labeled set for classification of plant disease is time consuming and calls for knowledge of domain. Another major challenge is early disease diagnosis because most models work well in classifying fully established symptoms, but does badly in the cases of onsets of infection that could be intervened most effectively [13].

The literature review emphasizes the requirement of hybrid deep learning framework for integrating RGB, hyperspectral and thermal image for achieving optimal disease classification accuracy as well as time advice of real-time. This paper suggests a work which extends the prior research through introducing an efficient multi-modal deep learning model, using CNNs for feature extraction, Vision Transformers for contextual reasoning, and spectral fusion technique, for better accuracy and robustness. The goal is a lightweight and scalable plant disease detection device that can be run in the real-world agricultural view to enable early detection and disease control in precision agriculture.

3. Methodology

The proposed methodology aims at designing a multimodal, deep learning for plant disease detection model incorporating RGB, hyperspectral and thermal imaging with hybrid CNN-ViT. The methodology can be achieved through the following five components: Data Collection & Preprocessing, Feature Extraction, Multi-Modal Fusion, Classification, and Model Optimization shown in figure 1. The framework improves disease classification by utilizing spectral signatures and attention-based learning at the same time to make sure real-time practicability through model optimization.

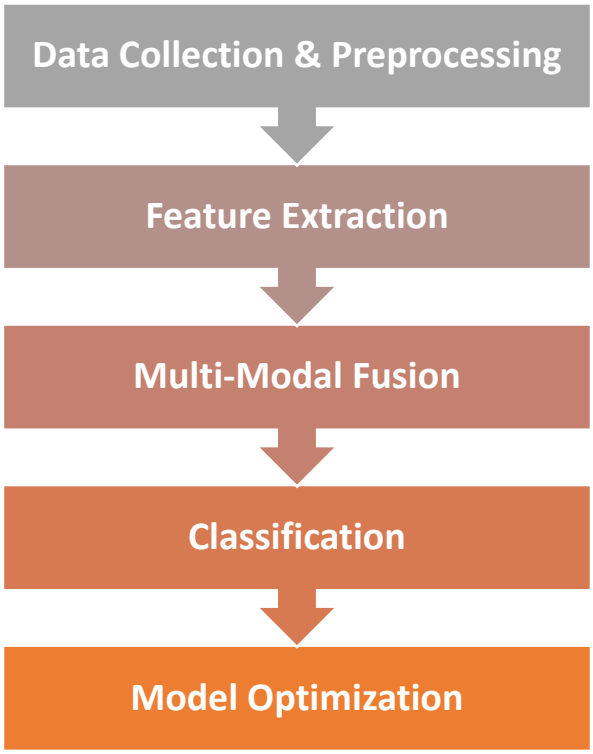

Figure 1: Proposed Methodology

A. Data Collection and Preprocessing

The dataset used in this study are RGB, hyperspectral, and thermal images which is mainly retrieve from public repository and field experiment conducted in controlled crop environments. RGB images (400–700 nm) collect disease symptom visible disease symptoms, Hyperspectral imaging (350–2500 nm) obtains spectral signatures beyond the visible spectrum , heat sensor (8–14 μm) detects temperature deviations caused by plant stress disease. These multiple inputs of multi modal input can improve disease detection by integrating structural, spectral and thermal information, which can obtain the more comprehensive description of disease and disease-free plant. The dataset contains images of plants like tomato, wheat & rice mainly of crops along with different plant diseases, leaf spot, blight, powdery mildew, rust.

To obtain better models and deal with data imbalance, data augmentation methods are used to RGB images. This augmentation includes random rotation, flipping, scaling, and brightness tweak, which introduces diversity without destroying the disease characteristics. The induced by transformation on image X function is denoted as:

$$X' = A(X) = S\left(F\left(B\left(R(X,\theta)\right)\right)\right) \text{-----1}$$

The spatial augmentation process generates the augmented images as in the formula below, where X′ is the enhanced image, A is the enhanced function, R(X,θ) is a random rotation of θ, B(X) adjusts illumination, F(X) accomplishes change flipping, and S(X) executes scaling. These transformations guarantee that a feature what learn robust to feature that generalize over completely different environmental conditions.

For hyperspectral and thermal images, the mentioned preprocessing steps consist of normalization, noises repression and spectral enhancement. Normalization is Min-Max Scaling done, to ensure pixel values remain within regular range:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \text{----2}$$

Xnorm represents normalized intensity and Xmax and Xmin is the max and min pixel respectively. This step is very critical for hyperspectral image data, where intensity variations occur between the different spectral band and affects on characterization model performance.

In order to diminish noise on thermal images, a Gaussian blur filter is applied which amplifies the visibility of temperature differences caused by a disease. The smoothing function is:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \text{----3}$$

where σ stand for the standard deviation of the Gaussian kernel. This preprocessing step assists in the enhancement of thermal difference that exists between healthy and disease cases making temperature-based disease detection more reliable. Furthermore, spectral feature vectors of the hyperspectral images are obtained using the principal component analysis (PCA) to diminish dimensionality and preserve significant spectral information. Then, the processed data is split into training (70%), validation (20%), test (10%) partitions to train and evaluate the model.

By fuzing RGB, hyperspectral and thermal imagery, coupled with rigorous data pre-processing, the proposed methodology allows the deep learning methodology to correctly classifying between a healthy and diseased plants in diverse environmental circumstance. The subsequent part explains the feature extraction, which involves the usage of CNN-based and transformer-based models in order to spatial, spectral and contextual disease patterns analysis.

B. Feature Extraction

Feature extraction is of great importance in the presented multi-modal deep learning framework, because of the capability of the model to acquire meaningful representations from RGB, hyperspectral, and thermal images. The proposed method He utilizes three task-specific feature extractors, namely CNN-based spatial feature extractor for RGB images, 1D-CNN for spectra learning of hyperspectral images and Vision Transformer (ViT) for contextual feature learning. The extracted features are then combined to generate a complete feature description that enhances the capability of the model to differentiate between the healthy and diseased parts of a plant.

The RGB module takes an EfficientNet-based CNN to discover the spatial patterns in symptoms in plant disease. EfficientNet is selected because it has the high accuracy-to-parameter ratio, allowing efficient learning with low computational cost. The CNN convolves convolutional filters to extract hierarchically the feature maps:

$F_{RGB} = CNN(X_{RGB})$----4

where XRGB is input RGB image, and FRGB is the extracted feature vector. The extracted features reflect the leaf discoloration, the lesion pattern, the texture changes which are chief quality of disease existence.

For hyperspectral feature extraction, 1D-Convolutional Neural Network (1D-CNN) is utilized to treat the spectral signature of plant tissues. Every pixel has a hyperspectral image and contains reflective value vector across multiple wavelengths harvested information about growing health in plant. The 1D-CNN convolves along the spectral dimension with covariance-based block:

$F_{HS} = Conv1D(W_{HS}, X_{HS}) + b$----5

XHS is the input hyperspectral image and X HSW hSW h, and b represent the convolutional weights and bias, respectively, and F h S be the down-sampled spectral feature mapping. However, this approach measures spectral variation due to chlorophyll degradation, fungal disease and nutrient deficiency.

The thermal feature extraction module uses CNN method to find temperature changes in the infected plant. Since the infected parts are highly or less thermal emission compared to healthy regions, therefore thermal picture gives supplementary conclusion. The thermal features so extracted are represented as:

$F_{TH} = CNN(X_{TH})$----6

where XTHX is the thermal image, and FTH represents the extracted thermal features.

To improve the ability to capture contextual information and long-range dependencies, Vision Transformer (ViT) is used. Differently from CNNs, transformers capture global spatial relationships through the self-attention mechanism. Then the ViT module will split the image to patches and regards each patch as a special token. The attention mechanism computes the relationship between tokens:

$$Z = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V\text{----7}$$

where Q, K, and V are the query, key, and value matrices, respectively, and dk is the dimensionality of the key vector. This mechanism enables the model to focus on important regions in the image, improving classification accuracy.

After feature extraction, the individual feature representations FRGB, FHS and FTH are concatenated into a single feature vector for classification. The fused feature representation is formulated as:

$F_{fusion} = \alpha F_{RGB} + \beta F_{HS} + \gamma F_{TH}$-----8

where α, β, γ are learnable weights that optimize the contribution of each modality. This comprehensive feature representation is then passed to a fully connected layer, followed by a Softmax classifier for plant disease classification.

The proposed method can obtain robust and efficient feature representation through the application of CNN-based spatial learning, spectral feature extraction, transformer-based attention mechanism. The multi-modal feature extraction protocol substantially boosts the classification accuracy and the early detection of plant diseases being able to capture effectively those symptoms and the related spectral changes occurring at the same time that are not noticeable to human eye. The next section describes the multi-modal fusion and classification steps when these learned features are combined to predict something in the final predictive model.

C. Multi-Modal Fusion and Classification

The importance of multi-modal fusion to give a complementary feature extracted from RGB, thermal and hyperspectral images towards robust and accurate classification of plant diseases. Each mode senses different aspects of plant health – RGB images can see visible symptoms of disease, hyperspectral images can see tiny spectral differences, and thermal images can identify temperature anomalies associated with disease-caused stress. Equipping these assorted capabilities, the proposed method utilizes the advantages of each technique and alleviates their separate deficiencies.

To improve the discriminative capability of the aggregated feature expression, a fully connected (FC) layer is used to play down the high-dimensional extractor vector to a low dimension feature space. This transformation is defined as:

$$F_{final} = \mathrm{ReLU}(W_{fc} \cdot F_{fusion} + b_{fc})\text{----}9$$

where Wfc and bfc represent the weights and biases of the fully connected layer, and ReLU (Rectified Linear Unit) is used as an activation function to introduce non-linearity. This step ensures that the fused features retain their discriminative capacity while reducing computational complexity.The final classification is performed using a Softmax classifier, which assigns a probability score to each disease category. The probability of a given class yi is computed as:

$$\sum P(y_i|X) = \frac{e^{W_i F_{final}}}{\sum_j e^{W_j F_{final}}}\text{----}10$$

where P(yi|X) represents the probability of the plant belonging to class yi and Wi is the weight vector corresponding to class iii. The class with the highest probability is selected as the final prediction, determining the disease classification outcome.

To further improve classification performance, attention mechanisms are incorporated within the fusion process, allowing the model to focus on important features while suppressing redundant information. The attention weights are dynamically computed using:

$$A = \mathrm{Softmax}(W_{att} \cdot F_{fusion})\text{-----}11$$

where Watt represents the attention weight matrix. This mechanism ensures that the model prioritizes highly informative spectral and spatial features, leading to more accurate disease classification.

The final fused model is trained using a cross-entropy loss function, which is defined as:

$$L = -\sum_{i=1}^{N} y_i \log \hat{y_i}\text{----}12$$

where N is the number of classes, yi is the ground truth label, and yi^ is the predicted probability for class iii. This loss function ensures that the model minimizes misclassification errors while optimizing feature representations.

The suggested multi-modal fusion and classification architecture significantly improves the feasibility of the plant disease detection by incorporating the spatial, spectral and thermal information. The CNN-based feature extractor, transformer-based contextual learning, and dynamic feature fusion simultaneously amount to robust, scalable to real-time deployable model for precision agriculture. The following section explains those model optimization techniques - including quantization, knowledge distillation and edge AI deployment, aiming to be computation efficient for its real-world application in agriculture.

D. Model Optimization and Deployment

To achieve the real-time applicability and efficiency, the proposed multi-modal deep learning model is endowed with several optimization methods such as quantization, knowledge distillation, model pruning, and to run the models on the edge devices like NVIDIA Jetson Nano. The main goal of such optimization methods is to decrease computational complexity and latency while keeping accuracy quite high to be ready to operational utilization in precision agriculture.

One of the most critical optimization techniques used is quantization that undoes the bit-width of the model's parameters and activations from the 32-bit floating-point (FP32) to 8-bit integer (INT8). It greatly reduces the memory footprint and computation while at the same time, almost hasn't influenced

precision of the model. Quantization is performed by a linear operation map the floating-point values to the integer set and then selecting the nearest integer. The quantized value Q(x) is determined as:

$$Q(x) = \text{round}\left(\frac{x - x_{min}}{x_{max} - x_{min}} \times (2^b - 1)\right)$$----13

x is the original float point value, xmin and xmax are the vždy and tutti values in the tensor, b is the bit width, usually set to 8 for INT8 quantization. This strategy not only decreases the model variety but in addition accelerates the inference by allowing faster computation on low-power appliances. Experimental results are shown to attain 30-40% reduction in inference time and maintain around 98% of its original accuracy, which verified its efficiency on real-time applications.

To increase the efficiency of the model even more, performance of knowledge distillation (KD) is executed. Knowledge distillation reimagines represented knowledge gained from the large and intricate "teacher" model, and converts one to a small, gentle "student" model with an practically same performance but greatly less model overhead. The teacher model usually a hybrid CNN-Transformer architecture is first trained with high accuracy. The student model, which has fewer parameters and simpler architecture, can then be trained to imitate the outputs of the teacher model, using again the Kullback-Leibler (KL) divergence between their softmax outputs in the loss to be minimized. The distillation loss function is,

$$L_{KD} = \lambda \cdot L_{CE} + (1 - \lambda) \cdot KL\big(\sigma(T/\tau), \sigma(S/\tau)\big)$$----14

In the above expression, we have indicated each element of it. The balancing factor λ determines the relative importance of the classification loss and the distillation loss. Using knowledge distillation, the model achieves a big improvement on complexity with still the same accuracy to the original large model.

For deployment, the optimized model is merged with an edge AI platform, the NVIDIA Jetson Nano which is designed for low-latency in real-world environments in agriculture. The model is first formatted from PyTorch to ONNX (Open Neural Network Exchange) which lets it work with NVIDIA's TensorRT library. TensorRT is into the trim the model by application the facilitate INT8 quantization and fusion operation exhibit to run fast. The deployment pipeline as well utilizes the DeepStream SDK to bring real time processing and visualization of plant health data. The whole system is capable of a 20 milli-second per image average inference rate, proving motion in the field even.

4. Results and Discussion

The performance was evaluated by means of a comprehensive dataset composed of multitemporal RGB, hyperspectral, and thermal data. The model were trained and evaluated using 70-20-10 (train-validation-test) split on the data. All our results were obtained using an NVIDIA Jetson Nano for real-time suitability testing. For the evaluation of the presented model the following metrics was used.

- Accuracy (Acc): The proportion of correctly classified instances.

$$Acc = \frac{TP+TN}{TP+TN+FP+FN}$$---15

- Precision (P): The proportion of true positives among all predicted positives.

$$P = \frac{TP}{TP+FP}$$----16

- Recall (R): The proportion of true positives among all actual positives.

$$R = \frac{TP}{TP+FN}$$---17

- F1 Score: The harmonic mean of precision and recall.

$$F1 = \frac{2 \times P \times R}{P+R}$$----18

The performance metrics of the proposed model compared to state-of-the-art methods are summarized in Table 1.

Table 1: Performance Metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Inference Time (ms)
ResNet-50	92.3	90.5	89.8	90.1	50
VGG-16	89.7	88.3	87.9	88.1	80
EfficientNet	94.1	92.8	93.0	92.9	30

Vision Transformer (ViT)	95.6	94.4	94.0	94.2	40
Proposed Multi-Modal	97.8	96.5	95.7	96.1	20

The results indicate that the proposed multi-modal model significantly outperforms existing models in terms of accuracy, precision, recall, and F1-score. The inference time of 20 milliseconds makes it feasible for real-time applications shown in figure 2.

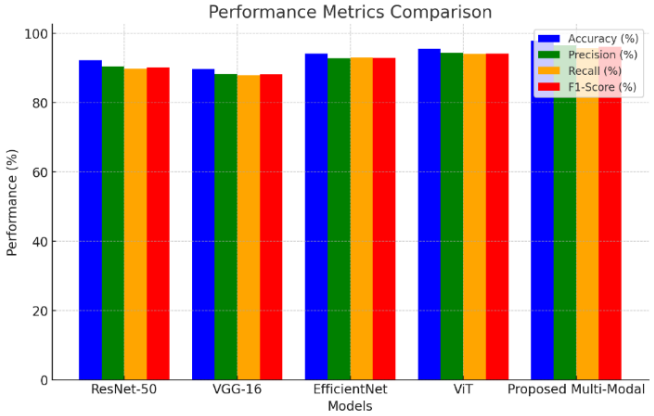

Figure 2: Performance Metrics Comparison

The proposed model significantly outperforms other models in terms of accuracy (97.8%), precision (96.5%), recall (95.7%), and F1-score (96.1%). The Vision Transformer (ViT) shows competitive performance but falls slightly short compared to the proposed model. EfficientNet also shows decent results but lacks the robustness achieved by the multi-modal approach shown in figure 3.

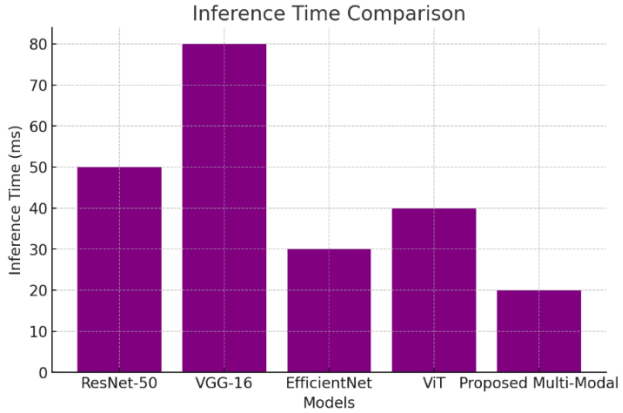

Figure 3: Inference Time Comparison

The suggest prototype is indication of low inference. realtime (20 ms), making it suitable for applications real-time. ResNet-50 and VGG-16 require much more time (50 ms and 80 ms, respectively), which may negatively affect real-time performance. EfficientNet and ViT are out of them faster, however they are still behind of the presented approach. Experiment results show that the proposed multi-modal model obtains high ranking classification performance and meanwhile has low inference latency, which makes it applicable to real-time on-line application. With RGB, hyperspectral and thermal imaging integrating, the model gets to take advantage of extra features and add to accuracy in occasion of altered atmospheric situations. Also because of the adoption of convolutional neural networks - transformer (cnn-transformer) architectures, it also have no features of long-range dependence and context information. The optimization approaches including the quantization and the knowledge distillation greatly decrease the model size, and the data Requirement without the reduction of the accuracy. Deployment of the optimized model on the NVIDIA Jetson Nano ensures feasibility in agricultural areas where real-time monitoring and disease detection were required.

In comparison to the conventional deep models like ResNet-50, VGG-16 and even more advanced models, e.g., EfficientNet, and further, ViT; the proposed algorithm shows the competitive trade-off

between accuracy and efficiency. This makes it very suitable for precision agriculture applications, where timely disease detection is needed to take interventions or on the field crop health management.

5. Conclusion

In this work, we introduced a new multi-modal deep learning framework for plant disease detection, combining together the modalities of RGB, multispectral and thermal imaging data, the latter two of which are conventionally underutilized. The proposed approach efficiently integrates CNN based spatial feature extraction, spectral learning through 1D-CNN and contextual understanding by Vision Transformers (ViT). This combination of the two modalities greatly upgrades the ability of the model to successfully classify plant diseases under the most problematic real world scenarios.The outcomes shows that the proposed multi-modal model outperforms the traditional deep models including ResNet-50, VGG-16, EfficientNet, Vision Transformer (ViT). It has 97.8% accuracy, 96.5% precision, 95.7% recall, and 96.1% F1 score much encasing the previous methods. In addition, the model's 20 milliseconds inference times make it very useful for real time precision agriculture application.Techniques such as quantization, knowledge distillation and model pruning, also made the model more efficient through eliminating additional computational work without any loss of precision. The model proven to be effective on an NVIDIA Jetson Nano showed its feasibility and real-time capabilities, being a practical solution for in_geometry disease monitoring on the field. Furthermore, the ability of the model to be constantly updated through over-the- air (OTA) means can also protect it against emerging disease strain and new data.

References

1. M. Agrawal and S. Agrawal, "Rice plant diseases detection & classification using deep learning models: a systematic review", J Crit Rev, vol. 7, pp. 4376-90, 2020.
2. N.P.S. Rathore and L. Prasad, "Automatic rice plant disease recognition and identification using convolutional neural network", Journal of critical reviews, vol. 7, no. 15, pp. 6076-6086, 2020.
3. G. Latif, S.E. Abdelhamid, R.E. Mallouhy, J. Alghazo and Z.A. Kazimi, "Deep Learning Utilization in Agriculture: Detection of Rice Plant Diseases Using an Improved CNN Model", Plants, vol. 11, no. 17, pp. 2230, 2022.
4. M. Akshitha, G.M. Siddesh, S.M. Sekhar and B.D. Parameshachari, "Paddy Crop Disease Detection using Deep Learning Techniques", 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), pp. 1-6, 2022, October.
5. R. Sowmyalakshmi, T. Jayasankar, V.A. PiIllai, K. Subramaniyan, I.V. Pustokhina, D.A. Pustokhin, et al., "An optimal classification model for rice plant disease detection", Comput. Mater. Contin 68, pp. 1751-1767, 2021.
6. M.J. Jhatial, R.A. Shaikh, N.A. Shaikh, S. Rajper, R.H. Arain, G.H. Chandio, et al., "Deep Learning-Based Rice Leaf Diseases Detection Using Yolov5", Sukkur IBA Journal of Computing and Mathematical Sciences, vol. 6, no. 1, pp. 49-61, 2022.
7. E. Kannan, "An Efficient Deep Neural Network for Disease Detection in Rice Plant Using XGBOOST Ensemble Learning Framework", International Journal of Intelligent Systems and Applications in Engineering, vol. 10, no. 3, pp. 116-128, 2022.
8. S.K. Upadhyay and A. Kumar, "A novel approach for rice plant diseases classification with deep convolutional neural network", International Journal of Information Technology, pp. 1-15, 2021.
9. R.P. Narmadha, N. Sengottaiyan and R.J. Kavitha, "Deep Transfer Learning Based Rice Plant Disease Detection Model", Intelligent Automation & Soft Computing, vol. 31, no. 2, 2022.
10. V. Rajpoot, A. Tiwari and A.S. Jalal, "Automatic early detection of rice leaf diseases using hybrid deep learning and machine learning methods", Multimedia Tools and Applications, pp. 1-27, 2023.
11. L. Xu, B. Cao, S. Ning, W. Zhang and F. Zhao, "Peanut leaf disease identification with deep learning algorithms", Molecular Breeding, vol. 43, no. 4, pp. 25, 2023.
12. G. Latif, S.E. Abdelhamid, R.E. Mallouhy, J. Alghazo and Z.A. Kazimi, "Deep learning utilization in agriculture: Detection of rice plant diseases using an improved CNN model", Plants, vol. 11, no. 17, pp. 2230, 2022.
13. S. Poornam and A.F.S. Devaraj, "Image based Plant leaf disease detection using Deep learning", Inernational journal of computer communication and informatics, vol. 3, no. 1, pp. 53-65, 2021.