การประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 48
The 48th Electrical Engineering Conference (EECON-48)
วันที่ 19-21 พฤศจิกายน 2568 ณ โรงแรมฟูราม่า จังหวัดเชียงใหม่

Attention-Based Fusion of Deep Learning Model for Grape Leaf Disease Classification

Tanawat Palamee[1], Buntueng Yana[1*], Sitthichat Meekhwan[1], Krisadakorn Khongsep[1]

Department of Electrical Engineering, School of Engineering, University of Phayao
Email: beetanawat7@gmail.com[1], mr.buntueng@gmail.com[1*], mr.sitthichat.n@gmail.com[1], kritsadakorn.mark@gmail.com[1]

Abstract

Grape leaves are susceptible to diseases that can significantly reduce crop yield and quality, leading to economic losses. While manual disease identification is common, it is often inaccurate and time-consuming. To address this, automatic recognition of grape leaf diseases through deep learning has become essential.

This study proposes a fusion model that combines the strengths of the VGG-16 and MobileNetV3 architectures to improve the accuracy and efficiency of grape leaf disease classification. The research also investigates the potential benefits of integrating an attention mechanism into the fusion model. The models were trained and evaluated using the Plant Village dataset, which contains 4,062 images of grape leaves categorized into four classes: Black Rot, Esca, Leaf Blight, and Healthy. To address the imbalance in the number of images per class, data augmentation techniques were applied.

The performance of the proposed models was compared against base models, including VGG-16 and MobileNetV3, using metrics such as accuracy, F1-score, and AUC. The results indicate that the fusion model with an attention mechanism achieved the highest performance with an accuracy of 99.55%, an F1-score of 99.55%, and an AUC of 99.99%. This outperformed the baseline fusion model, VGG-16, and MobileNetV3. The study demonstrates that combining different model architectures and incorporating attention mechanisms can lead to more accurate and efficient solutions for real-world agricultural applications.

Keywords: Grape leaf disease, Fusion network, Attention-Based model

1. Introduction

Agriculture is currently one of the most important food products for human nutrition [1], As the global population continues to grow, mainly in developing countries, the demand for food will rise significantly. Grapes are an economically important crop worldwide, particularly Central Europe and Southwest Asia. However, grape leaves are easily affected diseases caused by environmental factors, bacteria, and viruses [2]. These diseases include Black Rot, Esca, and Leaf Blight. They can greatly decrease grape yield and quality, leading to economic losses for farmers and disruptions in trade. From historical times up to today. There are multiple techniques for identifying diseases in grape leaves, manual identification. Mostly farmers use personal experience or expert visual inspection, which is often incorrect and time-consuming. Moreover, Inaccurate diagnoses often result in inappropriate pesticide use, which negatively impacts both the vineyard environment and fruit quality. Hence automatic recognition of grape leaf diseases is essential for improving diagnostic accuracy.

Today, machine learning techniques help achieve this. Effective quality control is also important to increase income from agricultural services [3]. Deep learning is a machine learning approach that relies on artificial neural networks and large datasets to deliver high prediction accuracy and quickly diagnosed plant diseases. There are common algorithms used for detection plant disease such as VGG, Resnet, Inception, and EfficientNet [4].

This study proposes a fusion model to improve the accuracy of grape leaf disease classification by merging strengths of both model VGG-16's deep future extraction and MobileNetV3's lightweight efficiency. The fusion model aims to achieve a balance between performance and processing cost. Furthermore, the study explores the future benefits of combining Attention mechanism into the fusion model. This addition is expected to enhance the model's ability.

2. Literature review

Many researchers have adopted CNN to automate disease detection tasks using leaf images. Deep learning, particularly CNN, has emerged as a predominant method for plant disease classification owing to its robust feature extraction capabilities. Within the agricultural sector.
Kunduracioglu el al(2024)[4] the research focuses on classifying grape leaves and identify diseases using 14 CNN models and 17 vision transformer models. The CNN-based models achieved an impressive accuracy of 99.03% and F1-scores of 97.80% and 97.62%, respectively. The models have strong potential for farmers and agricultural professionals by providing decision-making and improving production efficiency. This study emphasizes the opportunity of deep learning to transform plant health by evaluating agriculture. But limitations in real-world agricultural environments. Similarly, Mandal et al. [5] developed and compared several CNN models such as DenseNet121, ResNet50, VGG16, EfficientNetB7 etc. for automatic grapevine disease detection. While EfficientNetB7 achieved the highest accuracy at 99.6%, However, challenges of this work like limited disease scope and GPU dependency.

*Corresponding author

Many farmers lack access to GPthe ceiling devices, limiting practical deployment and larger models, for example EfficientNetB7 although it has excellent accuracy, but performance hit a ceiling after 150MB. To improve accuracy and computational efficiency, fusion-based approaches have been proposed. Tyagi et al. [6] explored the deep learning, particularly CNN in automating the identification of medicinal plant species and detecting diseases from leaf images. Among the tested models, InceptionV3 outperformed VGG-16 and VGG-19, attaining the most precision of 70.12%. While the research effectively highlights the advantages of pre-trained architecture and data augmentation techniques, it also points out limitations such as moderate accuracy, a relatively small dataset, limited training epochs, and computational inefficiency, especially in VGG models. The research also emphasized how fusion-based designs can provide improved performance over single-model architectures Yang et al. [7] introduced WaveLiteNet, a lightweight deep learning model designed to detect and classify five types of tea leaf disease such as Anthracnose and Cercospora Leaf spot. The model merges 2D Discrete Wavelet Transform with MobileNetV3 to improve feature extensions obtained 98.70% accuracy during testing. showing that using small models with frequency features can make them fast and accurate. This inspires the current study, which proposes a fusion model combining VGG-16 and MobileNetV3, aiming to leverage VGG-16 and MobileNetV3 to improve grape leaf disease classification accuracy while reducing computational cost. Moreover, the future directions and specialized techniques, Xie et al. [8] proposed real-time detector for grape leaf disease by applying CNN algorithms classification, it is suitable for mobile platform deployment. Zhang, et al. [9] presented YOLOv5-CA, coordinate attention into the YOLOv5 framework. for detecting Grape Downy Mildew disease. It obtains high precision 85.59% and ran at 58.82 FPS, demonstrating both high accuracy and speed an enhanced model for detecting. However, its dependence on RGB images, a single disease focus, and a limited dataset restrict broader application. Jin et al. [10] introduces GrapeGAN, an unsupervised image to enhancement method designed to improve grape leaf disease validation. Enhances images led to VGG-16 and InceptionV1 achieving up to 96.13% accuracy. However, dataset size was small requiring cross-validation and model complexity posed limitations for broader deployment. In addition to fusion and attention-based models. Together, these studies demonstrate that while CNN remains effective for grape leaf disease classification, combining architectures and integrating lightweight strategies such as MobileNet, attention mechanisms, wavelet transforms, and GAN-based image This upgrade might lead to easier and more scalable solutions. for real-world use. Therefore, lightweight models combined with fusion strategies and attention mechanisms are promising directions, which motivates this study.

2.1 Publicly available datasets

From review, there are four public available datasets: Grapes Net, Grapevine , Grape Leaf Disease, and Plant Village. This study reviews four publicly available datasets for grape leaf image analysis in Table 1.

Table 1 Summary Dataset

Dataset	Description
Grapes Net	Grapes Net dataset has 11,000+ grape collection images in real-world Indian vineyards using RGB and RGB-D (Intel RealSense D435i) cameras. It includes multiple lighting setups, angles, occlusions, and distances featuring both single and multi-cluster setups. Each image is provided with reference data.
Grapevine	Grapevine dataset includes 500 grape leaf images, with 100 images each from five grapes such as: ak, ala idris, buzgulu, etc. The images are classified into separate folders for each type, the dataset well-suited for grape type classification using leaf features[11]. This research uses the dataset to evaluate the success of grape leaf images in diagnosing grape types.
Niphad Grape Leaf Disease	The Niphad Grape Leaf Disease dataset have 2,726 RGB images of grape leaves collected from vineyards in Niphad, Maharashtra, India. Separate into four classes: Downy Mildew, Bacterial Rot, Powdery Mildew, and Healthy, supporting research in image-based grape leaf disease classification.
Plant Village	The Plant Village dataset, the largest public resource of plant disease images, was used in this study to detect diseases in grape leaves with high accuracy The Plant Village dataset was introduced by Hughes and Salathé The dataset includes more than 50,000 images of plant leaves with different diseases. This study analyzed approximately 4,000 grape leaf images, categorized into four classes: Black rot, Esca, Leaf blight, and healthy. The leaf images are stored in one folder

The PlantVillage dataset was selected for this research due to its significant scale and widespread acceptance within the academic community. Comprising over 4,000 images of grape disease. This dataset provides the substantial data volume required to train robust deep learning models. Its frequent use is a benchmark in numerous computer vision and agricultural technology. The dataset's suitability is crucial, as it ensures our results can be contextually evaluated against a broad of existing work in plant disease detection. Figure 1 provides an example of diseases from this data set.

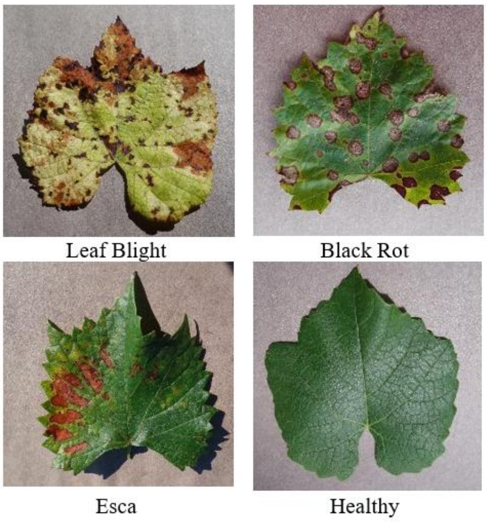

Fig. 1 Exemplar of Disease in Plant Village dataset

Pictures don't depict diseases that affect grape leaves. People who work on farms and experts who look at items and identify them by hand often make mistakes and take a long time. People could use pesticides erroneously because of these faulty diagnoses, which is terrible for the crops and the environment. Because they all have the same symptoms, it might be hard to discern Black Rot, Esca, and Leaf Blight from. It gets worse because the weather, bacteria, and viruses may all change how the leaves look. This is too complicated, and humans can't see everything, therefore we need more advanced technologies like deep learning to automatically collect and understand a lot of visual information to make diagnoses more accurate.

2.2 Algorithms

From reviews multiple deep learning architectures widely used for image classification and plant disease detection. The selected algorithms VGG-16, MobileNetV3, ResNet-18, Inception-V4, and DenseNet-121. Their performance and efficiency are compared to identify the most suitable model. We record a related model in leaf classification in Table 2.

This study selected several widely used deep learning architectures to compare their performance on plant disease detection. The VGG-16 model is CNN noted for its simplicity and effectiveness, which comes from using multiple 3x3 convolution layers stacked together; however, it is a large model with many parameters and is slow to train. In contrast, ResNet-18 is skip connections to facilitate the training of its 18 layers, which helps avoid the vanishing gradient problem and allows for faster training. The Inception-V4 model, another CNN, is recognized for being highly efficient and accurate by using special Inception modules that run multiple filter sizes in parallel, though its structure is complex. Finally, DenseNet-121 is a Densely Connected CNN where every layer is connected to all preceding layers. This design encourages feature reuse with fewer parameters but is known to be memory-intensive.

Table 2 Common model in leaf disease classification

Model	Description
VGG-16	VGG-16 is a CNN-based model known for its simple yet effective architecture. Its key idea is using small 3x3 convolution layers stacked on top of each other to create a deep network with a total of 16 weight layers. While it is effective for tasks like deep feature extraction, its primary disadvantages are its large model size, high number of parameters, and slow training speed, which can lead to computational inefficiency.
MobileNetV3	Based on the provided document, MobileNetV3 is characterized by its "lightweight efficiency". It is considered a small, fast model, and is referred to as a "lightweight strategy" for deep learning.
Resnet-18	ResNet-18 is a Residual Network that has a total of 18 layers. Its fundamental concept is the use of "residuals" to aid in the training of deep networks. This design helps avoid the vanishing gradient problem, which allows for faster training. Its main downside is that it is considered slightly more complex than VGG.
Inception-V4	Inception-V4 is an Inception-based CNN. Its core concept is the use of special "Inception modules" that combine multiple filter sizes in parallel. This design makes the model very efficient and accurate. Its primary disadvantage is its complex structure.
Densenet121	DenseNet-121 is a Densely Connected CNN. Its key architectural idea is that each layer connects to all previous layers within the network. This design offers the advantages of encouraging feature reuse while using fewer parameters overall. However, its primary drawback is that it is very memory-intensive to run.

3. Methodology

Fig 2 displayed full research. The initial step in research is moving images to Image Preprocessing so models may use them. The dataset's four classes Black Rot, Esca, Leaf Blight, and Healthy started with different numbers of pictures. Flipping, rotating, and scaling photographs ensured that each class had as many as Esca. This ensured model fairness. The research models in two ways after preprocessing. The suggested VGG-16 and MobileNetV3 model must be trained and evaluated first. The second method compares VGG16 and MobileNetV3. After that, the proposed and base models undergo a Performance Evaluation to assess Accuracy, F1-score, and AUC. Result and Analysis concludes the process. This step evaluates results in choosing the best model.

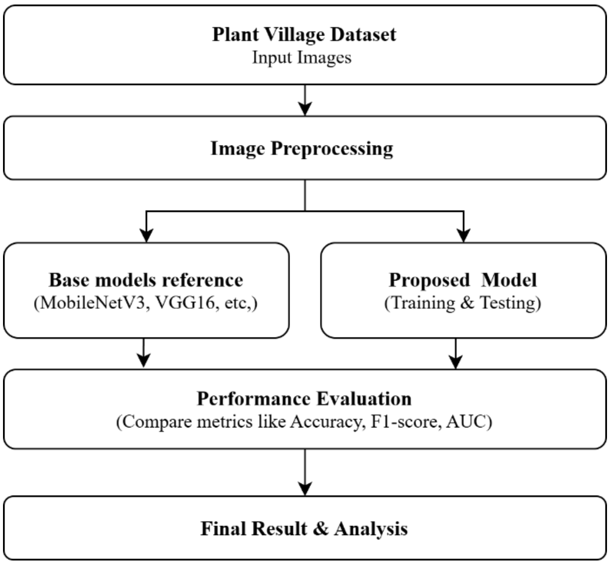

Fig. 2 Research Methodology framework

The initial stage involves preprocessing the Plant Village dataset images and incorporating them to ensure a balanced and equitable training environment. It then presents a fusion model both with and without an attention mechanism. This model has been trained and evaluated using VGG-16 and MobileNetV3. The study's most commendable aspect is its comprehensive performance assessment, employing Accuracy, F1-score, AUC, and Average Testing Time to provide a complete overview. This method identifies the optimal classification model for grape leaf diseases that is both computationally efficient and precise.

3.1 Preprocessing process

The Plant Village dataset have all grape leaf 4,062 image divided into 4 classes disease: Black Rot, Esca,. Leaf blight, and .Healthy. However, each class has a different number of images. Like Black rot 1,180 image, healthy 423 image etc. To avoid bias in model training and Poor accuracy for minority classes therefore balance each class has an equal number of images. and using the number of Esca images for main because Esca is the greatest number of images 1383 by using image augmentation techniques such as flipping, rotation, and scaling.

3.2 Proposed Model

The proposed model combines the best parts of VGG-16 with MobileNetV3. The design starts by running the VGG-16 and MobileNetV3 models on the same input image at the same time to find various features. We wish to combine the deep feature extraction of VGG-16 with the lightweight efficiency of MobileNetV3. After concatenation, the features from both pathways are combined into a feature vector that is more complex. Then, the model employs an attention mechanism to pick out the most essential parts of the unified vector that help it categorize illnesses. Finally, the updated features are looked at by fully connected layers of 512, 512, and 256 neurons to make the classification result.

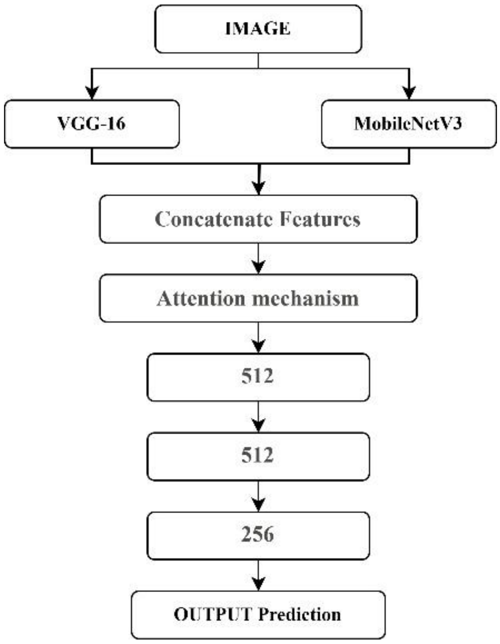

Fig. 3 Proposed model

To isolate and understand the specific contribution of the attention mechanism, the study also evaluated a modified version of the proposed network. This second model, referred to as the "simple fusion" model, follows the same initial architecture of concatenating features from VGG-16 and MobileNetV3 but omits the attention mechanism layer entirely. In this simplified structure, the combined feature vector is passed directly to the dense classification layers. Both the model with the attention mechanism and the simple fusion model were trained and tested, allowing for a direct comparison to quantify the performance improvement gained by integrating the attention component. For evaluation, we employed 10-fold cross-validation. In each fold, 80% of the dataset was allocated for training, 10% for validation, and 10% for independent testing. All reported metrics (Accuracy, F1-score, AUC, and Testing Time) are the averages across folds, ensuring robustness and fairness in performance assessment.

Table 3 Hyperparameters

Hyperparameters	Parameter
Batch size	32
epochs	20
Learning late	1e-4
k-fold	10

Table 3 Hyperparameters shows the hyperparameters that were used during testing. For consistency, these parameters are used on all of the models examined in this study.

Attention Mechanism

To enhance the fused feature representation, we employed a self-attention mechanism inspired by Transformer-style attention. The concatenated feature vector $F \in R^{72,128}$ is passed into the attention module, which computes attention weights to emphasize informative features and suppress redundant ones. We adopted the scaled dot-product attention formulation introduced in Vaswani et al [12], which defines:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$

where Q, K, and V are linear projections of the concatenated feature vector F, and d_k is the dimension of the key vectors. The attention-weighted output is then multiplied elementwise with FFF, yielding a refined representation before the classification layers.

3.3 Evaluation Matric

To evaluate the performance of each model, three key metrics were used: Accuracy, F1-score, and average testing time. Based on the confusion matrix, where each row shows predicted classes and each column shows actual classes, these metrics are derived from the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (2)$$

$$Precision = \frac{TP}{TP + FP} \qquad (3)$$

$$Recall = \frac{TP}{TP + FN} \qquad (4)$$

$$F1 - Score = \frac{2 \times TP}{2 \times TP + FP + FN} \qquad (5)$$

These metrics give a complete picture of each model's performance. Accuracy is a simple measure of correctness, but the F1-score balances Precision and Recall, ensuring the model is accurate and consistent in detecting positive cases. Additionally, average testing time is essential for assessing model feasibility. This integrated evaluation guarantees that the final model balances diagnostic accuracy, reliability, and computational efficiency, which is crucial for real-world implementation.

4. Result

Based on the study's results, the fusion model with an attention mechanism demonstrated the best performance for grape leaf disease classification, achieving the highest accuracy and F1-score. The fusion model incorporating an attention mechanism outperformed all other models, including the simple fusion model and the individual VGG-16 and MobileNetV3 base models. The key performance metrics for the evaluated models are summarized below:

Table 4 Performance of models

Model	Accuracy	F1-score	AUC
VGG-16	98.99	98.98	99.98
MobileNetV3	98.70	98.70	99.97
Simple-Fusion	99.22	99.22	99.99
Proposed Model	99.55	99.55	99.99

Results Table 4 demonstrate that the self-attention fusion achieves the best performance (Accuracy = 99.55%, F1 = 99.55%, AUC = 99.99%). While VGG-16 and MobileNetV3 alone were competitive, the fusion models consistently outperformed single architectures. Training history graphs illustrate the performance of four different models over 20 epochs in terms of accuracy and loss.

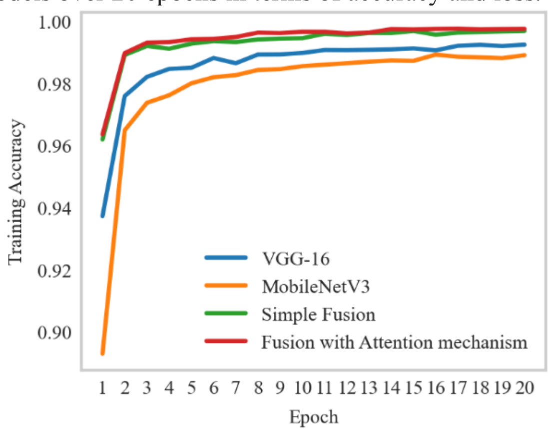

Fig. 2 Historical data of training accuracy

According to the training accuracy graph, all models improved over the 20-epoch training period, with the most significant gains occurring in the first few epochs. Throughout the training process, the fusion with attention mechanism model consistently achieved the highest accuracy. The simple fusion model performed at a level slightly below the attention-based model. Both the VGG-16 and MobileNetV3 models demonstrated lower training accuracy when compared to the two fusion models.The training loss graph demonstrates a rapid decrease in error for all models within the initial epochs, followed by a more gradual decline.

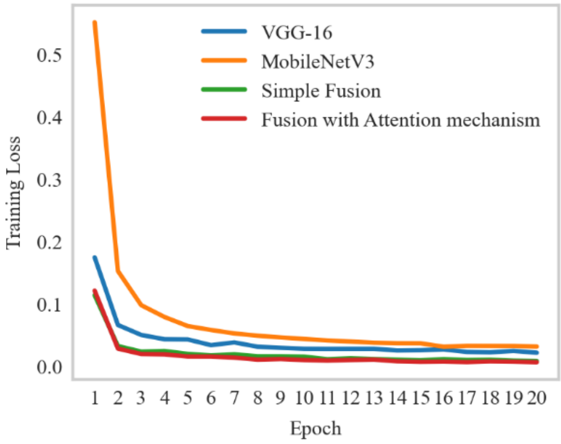

Fig. 3 Training loss of all models

As depicted in the training loss graph, the fusion with attention mechanism model consistently maintained the lowest training loss, indicating it was the most effective at minimizing errors during the training phase. The loss for the simple fusion model was slightly higher than that of the attention-based model. While MobileNetV3 started with a significantly high training loss, it improved rapidly. VGG-16 began with a lower loss than MobileNetV3; however, it was ultimately outperformed by both fusion models, which achieved lower final loss values.

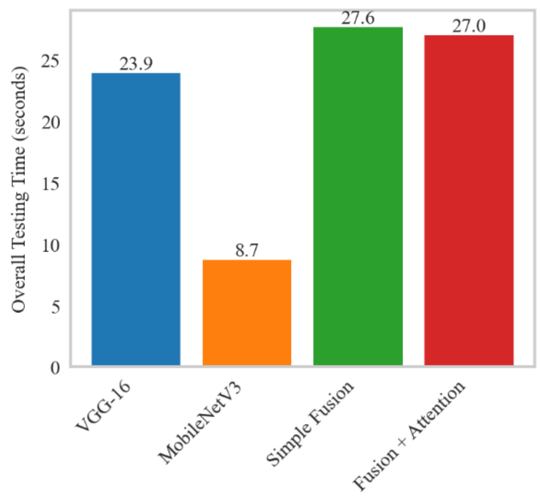

Fig. 4 Testing time of learning models

This bar chart illustrates the overall testing time in seconds for four different deep learning models: VGG-16, MobileNetV3, simple fusion, and fusion. with attention mechanism. The specific testing times recorded were 23.9 seconds for VGG-16, 8.7 seconds for MobileNetV3, 27.6 seconds for Simple Fusion, and 27.0 seconds for Fusion with Attention mechanism. Based on this data, MobileNetV3 is clearly the most computationally efficient model with the shortest testing time, while the Simple Fusion model is the least efficient, requiring the longest time to complete the testing process.

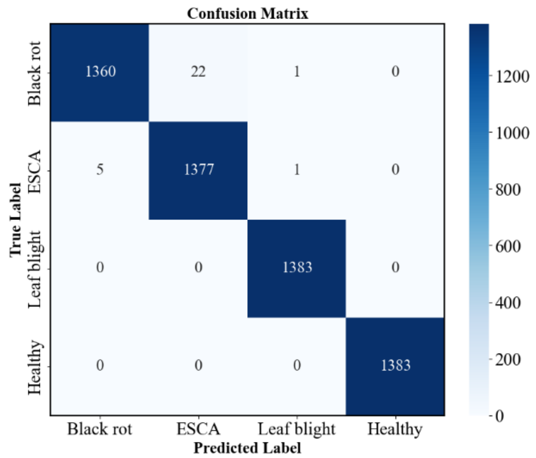

Fig. 5 Confusion Matrix of proposed model

This confusion matrix evaluates the classification performance of the proposed fusion with attention mechanism model across four distinct classes: Black rot, ESCA, Leaf blight, and Healthy. The matrix reveals that the model correctly classified 1360 instances of Black rot, 1377 of ESCA, 1383 of Leaf blight, and 1383 of Healthy. Misclassifications were minimal, with the most frequent error being the prediction of 22 Black rot samples like ESCA. Based on this data, the model demonstrates exceptionally high accuracy, achieving perfect classification for the Leaf blight and Healthy classes and showing only slight confusion between the two disease types, Black rot and ESCA.

5. Conclusions

This study successfully developed and evaluated a fusion model to improve the classification of grape leaf diseases. The results conclusively show that the proposed fusion model, which combines the VGG-16 and MobileNetV3 architecture and incorporates an attention mechanism, is the most effective. This model achieved the highest performance with an accuracy of 99.55%, an F1-score of 99.55%, and an AUC of 99.99% While the fusion models delivered superior accuracy, a trade-off with computational efficiency was observed. The Fusion with Attention mechanism model, despite its high accuracy, had a testing time of 27.0 seconds. In contrast, the standalone MobileNetV3 model was significantly faster, with a testing time of only 8.7 seconds, though its accuracy was lower at 98.70%

In conclusion, this research demonstrates that fusing different deep learning architectures and integrating an attention mechanism can lead to more accurate and robust solutions for real-world agricultural challenges like grape leaf disease detection. For future work, we aim to optimize the model for practical deployment on edge devices such as Raspberry Pi or mobile phones, enabling real-time disease detection in vineyards. Additionally, we plan to expand comparisons to other fusion strategies, including ResNet + MobileNet and CNN-Transformer hybrids, and validate the approach on real-world vineyard datasets beyond Plant Village.

6. References

[1] Sandhu, H., Bottom-Up Transformation of Agriculture and Food Systems. Sustainability, 2021. **13**: p. 2171.
[2] Soni, T. Deep Learning for Accurate Grape Disease Classification with ResNet50. in 2024 5th International Conference on Smart Electronics and Communication (ICOSEC). 2024.
[3] Basha, S.J., et al. Comparative Analysis of Time Series Forecasting Models to Predict Amount of Rainfall in Telangana. in 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS). 2022.
[4] Kunduracioglu, I. and I. Pacal, Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases. Journal of Plant Diseases and Protection, 2024. **131**(3): p. 1061-1080.
[5] Mandal, R. and M. Kheir, Vineyard vigilance: Harnessing deep learning for grapevine disease detection. Journal of Emerging Investigators, 2024.
[6] Tyagi, K., Implementing Inception v3, VGG-16 and VGG-19 Architectures of CNN for Medicinal Plant leaves Identification and Disease Detection. Journal of Electrical Systems, 2024. **20**: p. 2493-2501.
[7] Yang, J., et al., Lightweight wavelet-CNN tea leaf disease detection. PLOS ONE, 2025. **20**(5): p. e0323322.
[8] Xie, X., et al., A Deep-Learning-Based Real-Time Detector for Grape Leaf Diseases Using Improved Convolutional Neural Networks. Frontiers in Plant Science, 2020. **Volume 11 - 2020**.
[9] Zhang, Z., et al., Deep Learning Based Automatic Grape Downy Mildew Detection. Frontiers in Plant Science, 2022. **Volume 13 - 2022**.
[10] Jin, H., et al., GrapeGAN: Unsupervised image enhancement for improved grape leaf disease recognition. Computers and Electronics in Agriculture, 2022. **198**.
[11] Koklu, M., et al., A CNN-SVM study based on selected deep features for grapevine leaves classification. Measurement, 2022. **188**: p. 110425.
[12] Vaswani, A., et al., Attention Is All You Need. in 31st Conference on Neural Information Processing Systems (NIPS 2017), 2017.

Tanawat Palamee is currently pursuing the B.S. degree in Electrical Engineering at the University of Phayao, Thailand. His research interests machine learning, and robotic agriculture.

Buntueng Yana (Member, IEEE) received B.Eng. and M.Eng. degrees in Electrical Engineering from Chiang Mai University, Thailand, in 2006 and 2009, respectively, and a Ph.D. in Information Science and Technology from Osaka University, Japan, in 2019. He is currently a lecturer with the Department of Electrical Engineering at the University of Phayao, Thailand, where he began his academic career. His research interests include robotics, automatic control, power generation, and the application of machine learning to medical applications.

Sitthichat meekhwan received the bachelor's degree in electrical engineering from the University of Phayao (UP), Phayao, Thailand, in 2019. He is currently studying for a master's degree at the Faculty of Engineering. His current research interests in Machine learning, deep learning, and robotic

Krisadakorn Khongsep He is currently studying for a bachelor's degree in the Faculty of Engineering, Department of Electrical Engineering, University of Phayao. Current research interests: machine learning.