Artificial intelligence-based solutions for coffee leaf disease classification

Tri Cong Pham[1*], Van Duy Nguyen[2,3], Chi Hieu Le[4], Michael Packianather[5] and Van-Dung Hoang[6]

[1]School of Computer Science and Engineering, Thuy Loi University, Hanoi, Vietnam
[2]Institute of Biotechnology and Environment, Nha Trang University, Khanh Hoa, Vietnam
[3]School of Engineering, Newcastle University, Newcastle upon Tyne, United Kingdom
[4]Faculty of Engineering and Science, University of Greenwich, Kent ME4 4TB, UK
[5]School of Engineering, Cardiff University, Cardiff CF24 3AA, UK
[6]Faculty of Information Technology, HCMC University of Technology and Education, Ho Chi Minh City, Vietnam
*Corresponding author, E-mail:phtcong@tlu.edu.vn

Abstract. Coffee is one of the most widely consumed beverages and the quantity and quality of coffee beans depend significantly on the health and condition of coffee plants, particularly their leaves. The automation of coffee leaf disease classification using AI is an essential need, providing not only economic benefits but also contributing to environmental conservation and creating better conditions for sustainable coffee cultivation. Through the application of AI, early disease detection is facilitated, thereby reducing pest and disease control costs, minimizing crop losses, increasing coffee productivity and product quality, and promoting environmental preservation. Many studies have proposed AI algorithms for coffee disease classification. However, numerous algorithms employ classical algorithms, while some utilize deep learning, the current state-of-the-art in computer vision. The challenge lies in the fact that when using deep learning, a substantial amount of data is required for training. The design of deep learning architectures to enhance model accuracy while still working with a small training dataset remains an area of ongoing research. In this study, we propose deep learning-based method for coffee leaf disease classification. We propose the combination of different deep convolutional neural networks to further improve overall classification performance. Early and late fusion have been conducted to evaluate the effectiveness of the pre-trained model. Our experimental results demonstrate that the ensemble method outperforms single-model approaches, achieving high accuracy and precision in BRACOL coffee disease leaf.

1. Introduction

Coffee is one of the most consumable goods in the whole world. Overtime, the consumption demand is increasingly for both aspects: quantity and quality. In order to have the better quality of coffee, it is necessary to take care the development of coffee plants. However, due to factors such as: environment, air condition, land, bacteria, and viruses, can cause devastating losses for coffee growers. For the farmers, when they face plant diseases, the traditional approaches for these diseases are based heavily on the heuristic experiences. By the observations, they can know the problem and finding the corresponding solution. However, this approach has a limitation such as: time delays, misdiagnosis, and the ability when scaling at the huge plant area. Manually inspecting for diseases is a labor-intensive and time-consuming process, often characterized by significant delays. This leads to several ramifications resulting from late or incorrect disease detection, including: 1) Increased costs associated with disease treatment in terms of medication and labor when the disease has already spread extensively; 2) Reduced crop development potential, subsequently impacting coffee bean quantity and quality; 3) Adverse environmental effects due to the heightened use of plant protection chemicals. To reduce these effects, it is crucial to quickly diagnosis these diseases then finding the extract treatment for each plant. Therefore, we need a smart and robustness solution for this problem.

With tremendous performance of the deep learning model in many domains, the use of the latest architecture has been adapted for specific task from both computer vision and nature language processing. One of most attention domains besides the medical is the agriculture. In fact, there are many tasks in the agriculture are in charge of deep learning and obtain the promising results such as: counting fruit [1, 2], yield prediction [3, 4], detecting and classifying the diseases [5, 6]. The potential impact of deep learning in agriculture are clearly increased by both the accuracy and the processing time. It can help farmers not only improve the quality of their product but also reduce the expense for the development of the product through optimizing resource allocation and reducing unnecessary pesticide usage. However, to train deep learning networks, a large amount of data is required. Therefore, the challenge when applying this technique still is collecting data, labeling and expert, interpretability. It is then necessary to take advantage from other domain with huge knowledge to transfer to target domain to solve the data limitation. In addition, optimizing the design of deep learning architectures to enhance network efficiency while still using the same small dataset is also a potential approach for the problem of coffee leaf disease classification.

In this study, instead of using single deep learning models that rely on a single architecture, we investigate the ability of ensemble model in which model is combined of multiple deep learning model to make more accurate and robust predictions. The proposal of ensemble relies on the different characteristics of each model, making model more robustness, resulting in improved generalization, reduced over fitting, and enhanced performance across a wide range of applications. From these observations, in this paper, we propose the use of the latest deep learning model to automatic classify coffee leaf disease.

2. Related works

Artificial intelligence has been employed for crop disease classification in agriculture in numerous studies [7, 8, 9, 10, 11, 12, 13, 14, 15]. These research endeavors utilize disease-infected images of crop leaves to classify diseases across various types of crops, including tomato [9, 10], soybean [11], apple [12], cucumber [13], banana [14], and wheat [15]. Initially, the predominant approaches relied on traditional machine learning methods using handcrafted features. In [7], the authors applied k-means clustering followed by fuzzy logic to determine disease based on pixel color information from the leaves. In [8], additional leaf attributes, such as texture and shape, were employed for classification. However, traditional approaches suffered from high computational demands and sensitivity to variations in conditions when dealing with new samples. Consequently, the majority of research endeavors transitioned to deep learning approaches, where feature extraction proves to be more robust. For instance, in [9], a deep learning-based approach was introduced for the simultaneous detection and classification of nine tomato diseases. The study conducted extensive experiments involving various Convolutional Neural Networks (CNNs) to perform disease area segmentation followed by classification. In [11], a two-step process was proposed for soybean leaf disease classification. Initially, an image processing technique was employed to extract leaves from complex backgrounds, after which different CNNs were utilized for the classification of these extracted leaves. In general, contemporary research predominantly relies on the application of CNNs, which have demonstrated strong performance in the context of classification tasks. One of the most extensive datasets for this particular task was introduced in the research conducted by [16]. This dataset comprises 50,000 carefully curated images, belonging 12 categories of healthy leaves and 14 categories of diseases from various plants, including apple, blueberry, grape, and others.

In the context of coffee leaf disease classification, the predominant methodologies primarily rely on the utilization of Convolutional Neural Networks (CNNs) [17, 18, 19, 20]. In the study by [17], two distinct approaches were introduced to address this issue, one based on texture attributes and the other on CNNs. For the texture attribute approach, the authors employed the Gray Level Co-Occurrence Matrix (GLCM) and Local Binary Pattern (LBP) as two distinct methods for leaf feature extraction. These features are fed into a basic feed-forward network for recognition. Regarding CNNs, they made slight modifications to the architecture of AlexNet [21] by utilizing 128x128-pixel inputs and incorporating three convolutional layers. The results obtained using CNNs demonstrate a significant improvement when compared to the texture attribute-based approaches. In [18], to address the constraint of limited training data, they introduced transfer learning, which involves using a pre-trained model from a large source domain and fine-tuning it with a dataset from the target domain. Initially, they trained an embedded network to ensure its capability to adequately represent input images from the extensive source domain. Subsequently, this network was utilized as a feature extractor for classification purposes within the target domain. Recently, in the study conducted by [20], two tasks were undertaken: classification and estimation of the severity of coffee leaf diseases. This was achieved by employing a single CNN architecture with two distinct outputs, each corresponding to one of the aforementioned tasks.

3. Proposed Method

In this research, we propose an ensemble deep learning approach for coffee leaf disease classification, comprising two steps: feature extraction and classification, as depicted in Figure 1. In the feature extraction step, instead of using features from a single CNN, we employ deep features from multiple top-performing CNNs and Vision Transformer with the same dimensions. Subsequently, we aggregate these feature vectors into a single vector of equal length. In the classification step, we utilize fully connected layers (FC) in conjunction with a softmax function in the final layer to perform classification.

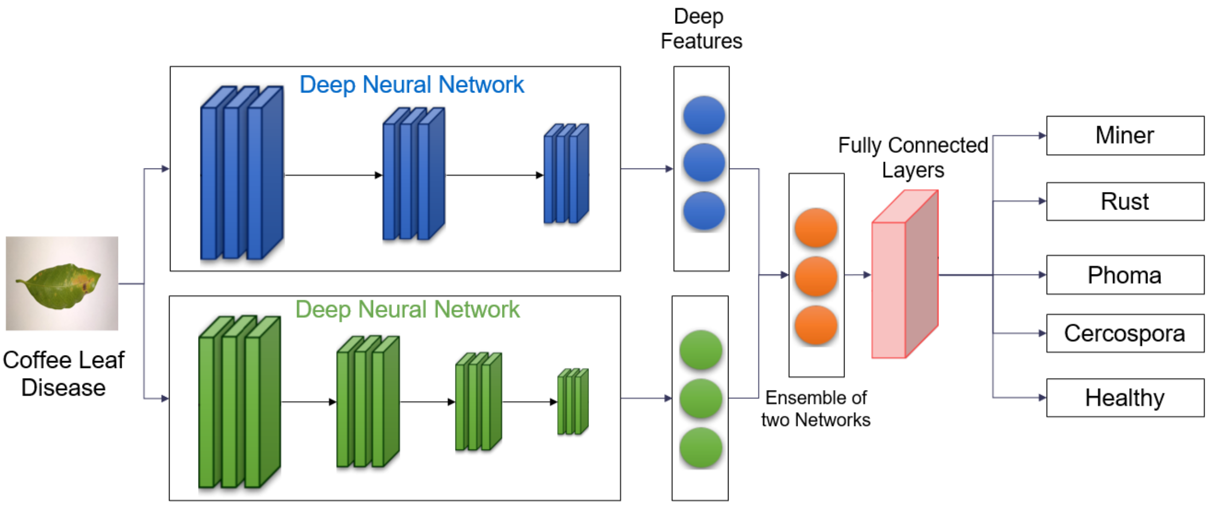

Figure 1. Proposed ensemble deep learning for Coffee leaf disease classification.

Ensemble feature of deep neural networks: Ensemble methods refer to techniques in which features are combined from different modalities or domains to create a superior feature representation. Among all computer vision classification algorithms, the feature extraction step is the most crucial, as the inability to extract meaningful features results in misclassifications into incorrect labels. Conversely, if meaningful features are successfully extracted, providing valuable information, they contribute to enhancing the algorithm’s accuracy. With recent advancements, deep convolutional neural networks

(CNNs) [21, 22, 23, 24] have demonstrated their capabilities in various computer vision tasks, including object recognition and classification. The success of these architectures is largely attributed to the availability of extensive datasets. However, in certain domains where data collection is considerably limited, the utilization of pre-trained models becomes imperative. Each of these architectures employs a distinct approach to extract and represent features from input data. Therefore, it becomes more intriguing to accumulate these advantageous representations to acquire superior features. In the context of this study, we limit our consideration to the ensemble of two deep neural networks.

Feature fusion and fully connected layers: there are two distinct methods for combining these features: early fusion and late fusion. Early fusion involves the combination of raw or low-level features from different sources at an initial stage. Conversely, late fusion involves combining higher-level representations at a later stage, often during the decision-making phase. Late fusion allows for modality-specific processing and offers greater flexibility in handling diverse data sources; however, it may potentially overlook certain interactions between modalities. The selection between early and late fusion depends on the specific task, the characteristics of the data, and the desired level of integration and interpretability. For both networks, we exclude the final fully connected layers. However, since the output dimensions of each network differ, in the case of early fusion, we introduce an additional fully connected layer to ensure they share the same output dimension, denoted as $d_{ensemble}$. In contrast, for late fusion, these features are directly fed into the FC layer, which produces 5 outputs.

4. Experiments and Results

4.1. Materials

In this study, to evaluate our proposed algorithm, we utilized the publicly available BRACOL dataset [25]. This dataset is designed for evaluating two tasks: disease classification and severity. However, our focus is solely on the disease classification task. In this task, leaves are categorized into five classes: Rust, Miner, Cercospora, Phoma, and Healthy. To facilitate comparisons with other studies, we followed the same experimental setup as described in [25], where we utilized 2,147 cropped symptom images, with each leaf containing only one disease.

Table 1. Number of sample data for train/validation/test dataset.

Disease	Train	Validation	Test
Rust	694	148	149
Miner	414	89	90
Cercospora	265	56	57
Phoma	353	76	75
Healthy	179	39	38

4.2. Experiment setting

To enhance the diversity of the training data, we employed various data augmentation strategies, including Random Horizontal Flip, Random Vertical Flip, Random Rotation, and Color Jitter. The Stochastic Gradient Descent (SGD) optimizer was utilized with an initial learning rate of 0.01 and a weight decay of 0.0005. All models used in our experiments were initialized with pre-trained weights obtained from each respective network. The experiments were conducted using an NVIDIA A40 GPU with 48GB of memory.

4.3. Performance Evaluation

Ablation Study: In this study, we conducted a comprehensive evaluation of ensemble features from renowned CNNs (such as MobileNet [26], DenseNet [27], EfficientNet [28], GoogleNet [29], ResNet [23], VGG [22]) and Vision Transformer [24] using two feature combination methods: early fusion and late fusion. The results of the seven ensemble combinations are presented in Table 2 (early fusion) and Table 3 (late fusion).

In Table 2, which presents the results of the early fusion strategy of seven ensemble methods, the most favorable outcomes are observed when both EfficientNet and MobileNet are used in combination. This combination achieves an accuracy, precision, and recall of 97.80%, 97.45%, and 97.92%, respectively. Additionally, it is worth noting that combining EfficientNet with the recent deep learning model Vision Transformer yielded promising results, even with minimal fine-tuning on a relatively small dataset.

Table 2. Ablation study early fusion on BRACOL dataset.

Architecture	Accuracy	Precision	Recall
MobileNet [26] + DenseNet [27]	96.09	95.40	96.28
EfficientNet [28] + Vision Transformer [24]	97.56	97.16	97.48
MobileNet [26] + EfficientNet [28]	**97.80**	**97.45**	**97.92**
GoogleNet [29] + EfficientNet [28]	95.84	95.44	95.40
Resnet [23] + EfficientNet [28]	97.07	96.72	96.99
VGG [22] + EfficientNet [28]	96.82	96.42	96.64
MobileNet [26] + Vision Transformer [24]	97.31	96.96	97.26

With the late fusion strategy, as presented in Table 3, the combination of MobileNet and Vision Transformer achieves the best performance, with accuracy, precision, and recall scores of 97.80%, 97.54%, and 97.70%, respectively. These results are comparable to the best outcome obtained using the early fusion strategy, with a minor difference in precision and recall (with precision being slightly higher by 0.09% but recall decreasing by 0.22% compared to the early fusion strategy).

Table 3. Ablation study late fusion on BRACOL dataset.

Architecture	Accuracy	Precision	Recall
MobileNet [26] + DenseNet [27]	96.58	96.00	96.55
EfficientNet [28] + Vision Transformer [24]	97.31	96.94	97.21
MobileNet [26] + EfficientNet [28]	96.09	95.62	96.23
GoogleNet [29] + EfficientNet [28]	96.33	95.95	96.36
Resnet [23] + EfficientNet [28]	97.07	96.83	97.12
VGG [22] + EfficientNet [28]	96.33	96.37	95.46
MobileNet [26] + Vision Transformer [24]	**97.80**	**97.54**	**97.70**

Comparison with other approaches in the literature: Table 4 provides a comprehensive comparison between the proposed method and state-of-the-art approaches on the BRACOL dataset. It is evident that the results obtained with handcrafted features are relatively low, demonstrating their limited capacity compared to deep learning-based methods. The use of CNNs, specifically Resnet50, highlights the effectiveness of deep learning in capturing intricate patterns in image data. Furthermore, thanks to the fusion strategy, the proposed method achieves an improvement of approximately 1% compared to using a single deep neural network.

Table 4. Comparison with the state-of-the-art methods on BRACOL dataset.

Method	Accuracy	Precision	Recall
Co-occurrence Matrix (GLCM) [20]	56.50	56.50	56.50
Local Binary Patterns (LBP) [20]	84.75	84.75	84.75
CNNs (Resnet50) [30]	97.07	96.85	96.99
Our proposed method	**97.80**	**97.45**	**97.92**

The confusion matrices of the two best-performing combinations of the two fusion methods are illustrated on the left (MobileNet + EfficientNet using early fusion) and on the right (MobileNet + Vision Transformer using late fusion) of Figure 2, respectively. Overall, the recall for each disease is consistently high, exceeding 95% in both cases. With these results, the algorithms can be confidently applied in real-world applications. Additionally, both models yield nearly identical results, with the first exception of Rust being misclassified as Cercospora in 3 and 2 samples for early fusion and late fusion, respectively. The second exception is the misclassification of Cercospora as Miner in 1 and 2 samples for early fusion and late fusion, respectively. Apart from these instances, all other metrics are the same.

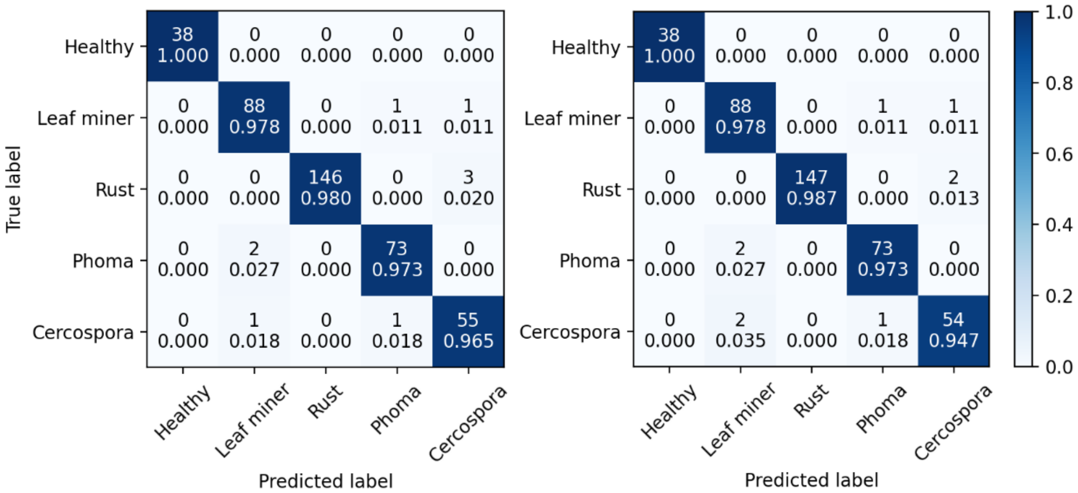

Figure 2. Confusion matrix on BRACOL dataset of 1) Early fusion of MobileNet and EfficientNet on the left; and 2) Late fusion of MobileNet and Vision Transformer on the right.

5. Conclusion and future works

In this work, we proposed the use of deep learning technique for the coffee disease leaf classification. In details, we use the ensemble model which based on two different deep neural networks for accumulating representation. The combination of distinct features from various backbones leads to an enhanced classification effectiveness of the ensemble algorithm. The proposed model improves the accuracy in comparison with the single architecture. In order to further improve the accuracy, the semi-supervised or unsupervised can be considered to take the advantages from the huge leaf dataset. Moreover, we also need to think about the interpretable insights into why a particular classification was made. This can help farmers understand the decision process and build trust in AI-based systems.

Acknowledgments

This work was supported by the Vingroup Innovation Foundation (VINIF) annual research support program with the research grant number VINIF.2021.DA00047.

References

[1] Rahnemoonfar M, Sheppard C 2017 Deep count: fruit counting based on deep simulated learning *Sensors* **17**

[2] Gao F *et al.* 2022 A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard *Comput. Electron. Agric.* **197**

[3] Khaki S, Wang L 2019 Crop yield prediction using deep neural networks *Front. Plant Sci.* **10**

[4] Khaki S, Wang L, Archontoulis S V 2020 A cnn-rnn framework for crop yield prediction *Front. Plant Sci.* **10**

[5] Saleem M H, Potgieter J, Arif K M 2019 Plant disease detection and classification by deep learning *Plants* **8**

[6] Li L, Zhang S, Wang B 2021 Plant disease detection and classification by deep learning—a review *IEEE Access*. **9** 56683–56698
[7] Sannakki S S *et al.* 2011 Leaf disease grading by machine vision and fuzzy logic *Int J.* **2** 1709–1716
[8] Siricharoen P *et al.* 2016 A lightweight mobile system for crop disease diagnosis *Image Analysis and Recognition: 13th International Conference*
[9] Fuentes A *et al.* 2017 A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition *Sensors* **17**
[10] Xiao J R *et al.* 2020 Detection of strawberry diseases using a convolutional neural network *Plants* **10**
[11] Karlekar A, Seal A 2020 Soynet: Soybean leaf diseases classification. Computers and Electronics *Agriculture* **172**
[12] Liu B *et al.* 2017 Identification of apple leaf diseases based on deep convolutional neural networks *Symmetry* **10**
[13] Ma J *et al.* 2018 A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network *Comput. Electron. Agric*. **154** 18–24
[14] Amara J, Bouaziz B, Algergawy A 2017 A deep learning-based approach for banana leaf diseases classification *Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband*
[15] Johannes A *et al.* 2017 Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case *Comput. Electron. Agric*. **138** 200–209
[16] Hughes D *et al.* 2015 An open access repository of images on plant health to enable the development of mobile disease diagnostics *arXiv preprint arXiv:1511.08060*
[17] Sorte L X B *et al.* 2019 Coffee leaf disease recognition based on deep learning and texture attributes *Procedia Comput. Sci*. **159** 135–144
[18] Afifi A, Alhumam A, Abdelwahab A 2020 Convolutional neural network for automatic identification of plant diseases with limited data *Plants* **10**
[19] [19] Sousa I C *et al.* 2022 Genomic prediction of leaf rust resistance to arabica coffee using machine learning algorithms *Scientia Agricola* **78** e20200021
[20] Lisboa E, Lima G, Queiroz F 2021 Coffee leaf diseases identification and severity classification using deep learning *Anais Estendidos do XXXIV Conference on Graphics, Patterns and Images*
[21] Krizhevsky A, Sutskever I, Hinton G E 2012 Imagenet classification with deep convolutional neural networks *Adv. Neural Inf. Process.* **25**
[22] Simonyan K, Zisserman A 2014 Very deep convolutional networks for large-scale image recognition *arXiv preprint arXiv:1409.1556*
[23] He K *et al.* 2016 Deep residual learning for image recognition *Proceedings of the IEEE conference on computer vision and pattern recognition*
[24] Dosovitskiy A *et al.* 2020 An image is worth 16x16 words: Transformers for image recognition at scale *arXiv preprint arXiv:2010.11929*
[25] Krohling R A, Esgario J, Ventura J A 2019 Bracol–a brazilian arabica coffee leaf images dataset to identification and quantification of coffee diseases and pests *Mendeley Data* **1**
[26] Howard A G *et al.* 2017 Mobilenets: Efficient convolutional neural networks for mobile vision applications *arXiv preprint arXiv:1704.04861*
[27] Huang G *et al.* 2017 Densely connected convolutional networks *Proceedings of the IEEE conference on computer vision and pattern recognition*
[28] Tan M, Le Q 2019 Efficientnet: Rethinking model scaling for convolutional neural networks *International conference on machine learning*
[29] Szegedy C *et al.* 2015 Going deeper with convolutions *Proceedings of the IEEE conference on computer vision and pattern recognition*
[30] Esgario J G M, Krohling R A, Ventura J A 2020 Deep learning for classification and severity estimation of coffee leaf biotic stress *Comput. Electron. Agric*. **169**