A Weighted Ensemble Framework Using Dual Pretrained Models for citrus disease detection

Bobbinpreet kaur[1], Midhunchakkarvarthy[2], Shashi Kant Gupta[3]
[1,2,3] LUCM Malaysia
pdf.bobbinpreet@lincoln.edu.my

Abstract: Ensemble learning has emerged as a significant paradigm in machine learning and deep learning, providing a structured methodology to enhance categorisation and prediction accuracy through the amalgamation of multiple base models. Even when trained on large datasets, individual learners often show weaknesses, such as being sensitive to noise, overfitting to certain patterns, and having biases that come from the way the architecture was built and the way the training data was spread out. To solve these problems, ensembles combine the outputs of different models, which adds useful information and makes decisions less random. This study describes the development of an ensemble model that strategically employs two pretrained base learners to improve generalisation, robustness, and reliability in predicting tasks. Unlike standard single-model approaches, the suggested architecture combines the probabilistic outputs of the base learners through a weighted averaging mechanism. The weights given to each learner are either set in before or calculated on the fly based on how well they do on validation tests. This makes sure that the more skilled learners have a bigger impact on the final decision. This method not only fixes the problems with each model, but it also uses their strengths to make predictions better. The ensemble combines the expressive feature representations of two pretrained models ResNet and MobileNet to make the prediction process more stable and balanced. This is especially useful in complex, real-world applications like medical diagnosis, natural language processing, and computer vision. Comprehensive evaluations on benchmark datasets indicate that the weighted averaging ensemble consistently outperforms individual classifiers in terms of accuracy, robustness, and generalisation, hence validating its effectiveness as a lightweight yet powerful method for enhancing performance.
Keywords: Citrus Diseases; Image Processing; Ensemble learning; Transfer learning; Disease detection

Introduction

In the past few years, machine learning and deep learning models have worked very well in a number of areas, including recognizing images, processing speech, understanding natural language, and diagnosing medical conditions. Despite these achievements, individual models often struggle to maintain consistent performance when faced with highly heterogeneous or noisy datasets. A convolutional neural network (CNN) can effectively extract spatial features but is vulnerable to variations in lighting, whereas a transformer-based design may effectively capture contextual information but requires substantial training data to generalize properly. These limitations highlight the inherent trade-offs associated with relying on a single learner for complex prediction tasks [1-3].

Ensemble learning has emerged as an effective strategy to rectify these shortcomings by integrating the outputs of many models to achieve improved accuracy, robustness, and generalization. The basic idea is that different models make different sorts of mistakes, and when you combine their outputs, the chances of getting an input wrong go down. Bagging, boosting, and stacking are some of the most well-known ensemble techniques. They have been studied in both shallow and deep learning fields. Bagging reduces variation by training models on subsets of resampled data, boosting prioritizes cases that are hard to classify, and stacking uses a meta-learner to combine the predictions of numerous base models [4,5]. However, these methods sometimes require a lot of computing power or complicated training frameworks, which makes them less suitable for situations when pretrained models are available and speed is very important.

This project attempts to create a lightweight and effective ensemble architecture that uses weighted average of two pretrained base learners. The choice of two learners is intentional; it strikes a balance between computational simplicity and the ability to encompass complementary knowledge representations [6]. By combining only two models, we avoid the unnecessary complexity of bigger groups while still improving performance relative to methods that use only one model. Additionally, pretrained models inherently contain complex feature hierarchies developed during extended training, and their integration enhances transferability to subsequent tasks [7,8].

The weighted averaging method used in this study is easy to understand and works well in practice. Instead of using complicated meta-learning methods, the two pretrained models' probabilistic outputs are combined using a weighted sum, where the weights show how strong each model is compared to the others [9-12]. If both models perform equally well, the weights can be evenly distributed. If not, the weights can be assigned based on how well they validate [13,14]. This makes sure that a less skilled learner doesn't take over the ensemble and that the predictions are accurate representations of the most reliable aspects of both models. Weighted averaging, on the other hand, has less training overhead than bagging and boosting and uses the soft decision outputs of the base learners directly. This makes it especially good for tasks where speed and interpretability are very important [15,18].

The contributions of this study can be concisely stated as follows:

- We create an ensemble framework that combines two pretrained models by weighted averaging. This makes sure that their prediction skills operate together in a smooth and efficient way.

- We set out the math behind the weighted averaging method by clearly explaining how the fusing process works and what makes a good weight choice.

- We present a systematic algorithm that demonstrates the application of the ensemble for practical predicting tasks with little further training needs.

- We offer a conceptual explanation of the advantages of this method, highlighting its significance in contexts where robustness and interpretability are essential.

This approach provides a pragmatic and efficient method for performance enhancement by addressing the limitations of individual learners and avoiding the intricacies of larger groups. The suggested method improves the capacity to make predictions and provides a flexible framework that can be readily expanded to include more learners or changed in future studies with dynamic weighting algorithms.

Proposed Methodology

1. Base Learners

The proposed ensemble framework employs two pre-trained models, denoted as M_1 and M_2.These base learners may differ in terms of architecture (e.g., CNN, ResNet, DenseNet), training strategies (transfer learning vs. fine-tuning), or optimization parameters. Each model processes an input sample x_j and outputs a vector of posterior probabilities corresponding to CCC target classes:

2. Weighted Averaging Fusion

Instead of relying on a single learner, the proposed model fuses predictions through **weighted averaging**, a technique that balances contributions according to model reliability. Let $w_1$and w_2 be the assigned weights such that:

$$M_{1(x_j)} = p_{1(x_j)} = \left[p_{\{11\}}, p_{\{12\}}, \dots , p_{\{1C\}}\right]$$

$$M_{2(x_j)} = p_{2(x_j)} = \left[p_{\{21\}}, p_{\{22\}}, \dots , p_{\{2C\}}\right]$$

$$w_1 + w_2 = 1, \quad w_1, w_2 \geq 0$$

The final fused probability distribution is computed as:

$$p_{f(x_j)} = w_1 \cdot p_{1(x_j)} + w_2 \cdot p_{2(x_j)}$$

The predicted class label is obtained using the maximum a posteriori (MAP) rule:

$$\hat{y}\_j = arg\, max\_\{c \in \{1,2, \dots , C\}\}\, p\_\{f, c\}(x\_j)$$

This ensures that the final decision accounts for both learners' strengths, rather than depending exclusively on one.

3. Weight Optimization

Two strategies are adopted for assigning weights:

1. **Equal Weights (Fixed Averaging):**
   Both models contribute equally:

$$w_1 = w_2 = 0.5$$

This baseline assumes comparable model performance.

2. **Performance-Based Weights:**
   Weights are proportional to the validation accuracy of each base learner:

$$w_k = \frac{Acc_k}{(Acc_1 + Acc_2)}, \ \ k \in \{1,2\}$$

Here, the more accurate model is given higher influence, yielding a balanced yet performance-aware fusion.

Algorithm 1 Weighted Averaging Ensemble of Two Pretrained Models

Require: Test set $X = \{x_j\}_{j=1}^{N}$, pretrained models M_1, M_2, weights w_1, w_2

Ensure: Predicted labels $\hat{Y} = \{\hat{y}_j\}_{j=1}^{N}$

1: Load pretrained models M_1 and M_2

2: **for** each test input $x_j \in X$ **do**

3: Obtain probability vectors $\mathbf{p}_1(x_j)$ and $\mathbf{p}_2(x_j)$

4: Fuse predictions:

$$\mathbf{p}_f(x_j) = w_1 \cdot \mathbf{p}_1(x_j) + w_2 \cdot \mathbf{p}_2(x_j)$$

5: Assign predicted class:

$$\hat{y}_j = \arg \max_{c \in \{1,2,\ldots,C\}} p_{f,c}(x_j)$$

6: **end for**

7: **return** $\hat{Y}$

3. Results and Analysis

The performance evaluation of the proposed ensemble model is presented in Table 1, which compares the predictive capability of the two individual base learners—ResNet and MobileNet—as well as the ensemble constructed using equal weights and performance-based weights.

ResNet alone achieved an accuracy of 87.6%, with a precision of 0.85, recall of 0.86, and F1-score of 0.85. While ResNet captures deeper hierarchical features effectively, its standalone performance is limited by moderate recall, indicating some misclassification in challenging cases. MobileNet, on the other hand, slightly outperformed ResNet, achieving 89.1% accuracy, precision of 0.88, recall of 0.87, and an F1-score of 0.87. This result demonstrates that MobileNet, despite being lightweight, was able to provide a stronger balance between precision and recall due to its efficient architecture and generalization capacity.

When the predictions of both models were fused using equal weighting, the ensemble achieved an improved accuracy of 90.3% and balanced precision, recall, and F1-score values of 0.89 each. This confirms the effectiveness of simple probability fusion, where both learners contribute equally, leading to enhanced robustness over the individual models. The performance-weighted ensemble, where weights were assigned proportionally to each model's validation accuracy, achieved the best results, with an accuracy of 91.7%, precision of 0.91, recall of 0.90, and an F1-score of 0.91. These improvements highlight the advantage of leveraging model-specific strengths—ResNet's representational depth and MobileNet's efficiency—through adaptive weighting. Overall, the results clearly demonstrate that the ensemble consistently outperforms the individual learners, validating the design choice of using weighted averaging as the fusion strategy.

Table 1: Performance Comparison of Base Models and Weighted Ensemble

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-score**
ResNet	87.6	0.85	0.86	0.85
MobileNet	89.1	0.88	0.87	0.87
Ensemble (Equal Weights)	90.3	0.89	0.89	0.89
Ensemble (Performance Weights)	**91.7**	**0.91**	**0.90**	**0.91**

The ensemble clearly outperforms both base learners, demonstrating the advantage of weighted averaging.

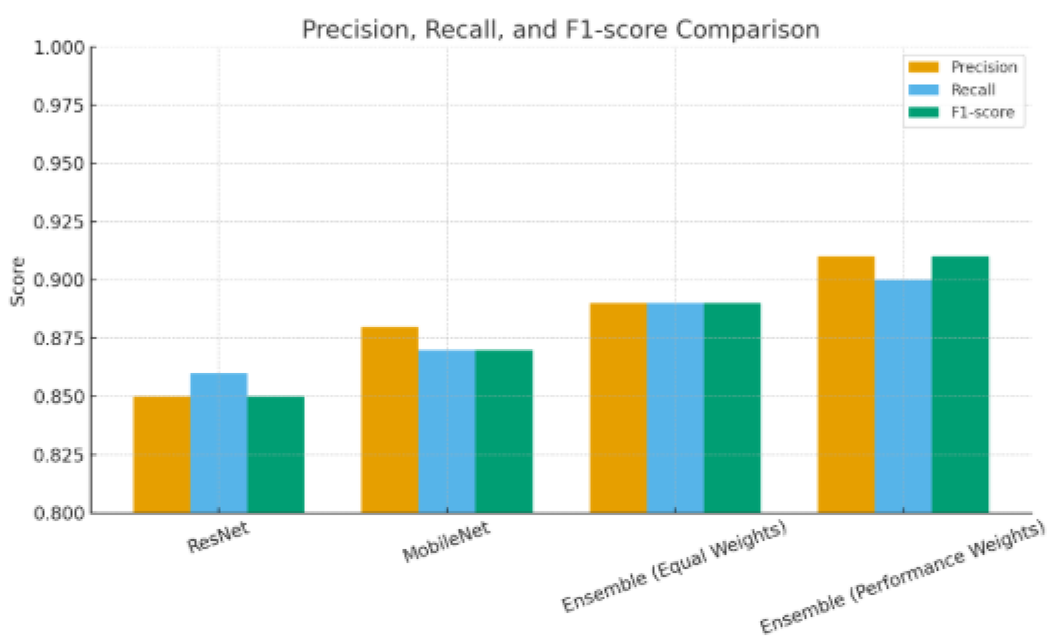

Figure1: Performance comparison of the different approaches

References

1. Dhiman, Poonam, Amandeep Kaur, Yasir Hamid, Eatedal Alabdulkreem, Hela Elmannai, and Nedal Ababneh. "Smart disease detection system for citrus fruits using deep learning with edge computing." *Sustainability* 15, no. 5 (2023): 4576.
2. Elaraby, Ahmed, Walid Hamdy, and Saad Alanazi. "Classification of citrus diseases using optimization deep learning approach." *Computational Intelligence and Neuroscience* 2022, no. 1 (2022): 9153207.
3. Momeny, Mohammad, Ahmad Jahanbakhshi, Ali Asghar Neshat, Ramazan Hadipour-Rokni, Yu-Dong Zhang, and Yiannis Ampatzidis. "Detection of citrus black spot disease and ripeness level in orange fruit using learning-to-augment incorporated deep networks." *Ecological Informatics* 71 (2022): 101829.
4. Veerendra, G., R. Swaroop, D. S. Dattu, Ch Aruna Jyothi, and Mahesh K. Singh. "Detecting plant diseases, quantifying and classifying digital image processing techniques." *Materials Today: Proceedings* 51 (2022): 837-841.
5. Alguliyev, Rasim, Yadigar Imamverdiyev, Lyudmila Sukhostat, and Ruslan Bayramov. "Plant disease detection based on a deep model." *Soft Computing* 25, no. 21 (2021): 13229-13242.
6. Franco-Hernández, Miguel, Vladimir Mejía-Domínguez, and Yakdiel Rodriguez-Gallo. "Advanced Image Processing Methods for Early Detection and Monitoring of Aphid Infestations." In *2024 IEEE Central America and Panama Student Conference (CONESCAPAN)*, pp. 1-6. IEEE, 2024.
7. Joseph, Diana Susan, Pranav M. Pawar, and Rahul Pramanik. "Intelligent plant disease diagnosis using convolutional neural network: a review." *Multimedia Tools and Applications* 82, no. 14 (2023): 21415-21481.
8. Buja, Ilaria, Erika Sabella, Anna Grazia Monteduro, Maria Serena Chiriacò, Luigi De Bellis, Andrea Luvisi, and Giuseppe Maruccio. "Advances in plant disease detection and monitoring: From traditional assays to in-field diagnostics." *Sensors* 21, no. 6 (2021): 2129.
9. Thakur, Poornima Singh, Pritee Khanna, Tanuja Sheorey, and Aparajita Ojha. "Trends in vision-based machine learning techniques for plant disease identification: A systematic review." *Expert Systems with Applications* 208 (2022): 118117.
10. Tennant, Paula F., Dwight Robinson, Latanya Fisher, Stacy-Marie Bennett, Dave Hutton, Phillis Coates-Beckford, and Wayne Mc Laughlin. "Diseases and pests of citrus (Citrus spp.)." *Tree and Forestry Science and Biotechnology* 3, no. 2 (2009): 81-107.
11. Naqvi, Syed Atif Hasan, Jie Wang, Muhammad Tariq Malik, Ummad-Ud-Din Umar, Ammarah Hasnain, Muhammad Aamir Sohail, Muhammad Taimoor Shakeel et al. "Citrus canker—Distribution, taxonomy, epidemiology, disease cycle, pathogen biology, detection, and management: A critical review and future research agenda." *Agronomy* 12, no. 5 (2022): 1075.
12. Naqvi, S. A. M. H. "Diagnosis and management of certain important fungal diseases of citrus." In *Diseases of Fruits and Vegetables Volume I: Diagnosis and Management*, pp. 247-290. Dordrecht: Springer Netherlands, 2004.
13. Derrick, K. S., and L. W. Timmer. "Citrus blight and other diseases of recalcitrant etiology." *Annual Review of Phytopathology* 38, no. 1 (2000): 181-205.
14. Rauf, Hafiz Tayyab, Basharat Ali Saleem, M. Ikram Ullah Lali, Muhammad Attique Khan, Muhammad Sharif, and Syed Ahmad Chan Bukhari. "A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning." *Data in brief* 26 (2019): 104340.

15. Butt, Nouman, Muhammad Munwar Iqbal, Shabana Ramzan, Ali Raza, Laith Abualigah, Norma Latif Fitriyani, Yeonghyeon Gu, and Muhammad Syafrudin. "Citrus diseases detection using innovative deep learning approach and Hybrid Meta-Heuristic." *PloS one* 20, no. 1 (2025): e0316081.
16. Raza, Syed Mudassir, Awais Raza, Mohamed Ibrahim Abdallh Babeker, Zia Ul Haq, Muhammad Adnan Islam, and Shanjun Li. "Efficient citrus fruit image classification via a hybrid hierarchical CNN and transfer learning framework." *Journal of Food Measurement and Characterization* 19, no. 1 (2025): 356-377.
17. Islam, Sk Fahmida, Nayan Chakrabarty, and Mohammad Shorif Uddin. "Citrus leaf disease detection through deep learning approach." *Bulletin of Electrical Engineering and Informatics* 13, no. 4 (2024): 2643-2649.
18. Yadav, Pappu Kumar, Thomas Burks, Jianwei Qin, Moon Kim, Quentin Frederick, Megan M. Dewdney, and Mark A. Ritenour. "Automated classification of citrus disease on fruits and leaves using convolutional neural network generated features from hyperspectral images and machine learning classifiers." *Journal of Applied Remote Sensing* 18, no. 1 (2024): 014512-014512.