A Hybrid CAE–MOCNN Architecture for Efficient Multiple-output Plant Species and Disease Classification

Dr. Parul Sharma
Department of CSE
Model Institute of Engineering and Technology
Kot Bhalwal, Jammu, Jammu and Kashmir, India
parul.cse@mietjammu.in

Dr. Mehak Mengi
Department of CSE
Model Institute of Engineering and Technology
Kot Bhalwal, Jammu, Jammu and Kashmir, India
mehak.cse@mietjammu.in

Abstract

Sustainable agriculture and efficient plant management majorly depend on the precise identification and classification of the plant species and their diseases. The CNN models are now in the phase where multiple outputs or tasks could be carried out at a time. However, such models pose high complexities by introducing large numbers of trainable parameters. To overcome this issue, a hybrid CNN–CAE architecture is proposed, where compact feature representations learned by a convolutional autoencoder are passed to a multi-output CNN consisting of shared layers and attention-based classifiers for the simultaneous prediction of plant species and disease types. Experiments conducted on the PlantVillage dataset demonstrate that the proposed model achieves 91.75% accuracy in species classification and 89.64% in disease identification, all while reducing trainable parameters by over 50%. These findings highlight the model's efficiency and effectiveness.

1.1 Introduction

Accurate recognition of plant species and their associated diseases is essential for supporting sustainable agricultural practices and improving crop productivity. Convolutional Neural Networks (CNNs) have been pioneering in image-based applications due to their strong ability to learn complex visual patterns and perform effective feature extraction and classification [1][2]. These models are now developed for better efficiency in the classification of the related tasks. The Multi-Output Convolutional Neural Networks (MOCNNs) are such models which uses a shared layer before branching into multiple outputs branches with their own set of convolutional and other layers for predictions such as such as plant species and disease categories [3].

The increasing complexity of these models comes with computational challenges. To overcome this issue, the research has been conducted in which a hybrid CNN–CAE architecture with integrated attention mechanisms has been utilized for enhancing classification training efficiency at the cost of slight reduction in classification results [4][5]. In this framework, the CAE encoder first compresses the input data into a lower-dimensional representation, which is then processed by a shared CNN classifier to produce multiple outputs [6][7].

2. Research Methodology

The experimentation was performed on publicly available PlantVillage dataset [8]. Its subset has been utilized for the multiple output tasks of plant species and disease classification. The images inside the dataset has the resolution of 256 × 256 pixels. For each class, a fixed number of samples was selected and divided into training, validation, and testing subsets comprising 320, 80, and 20 images per class, respectively.

In total, 15 classes were selected on the criteria of two hierarchical labels: plant species (6 classes) and disease categories (7 classes).The selected classes were: Apple_Healthy, Apple_BlackRot, Apple_Scab, Pepper_Healthy, Pepper_BacterialSpot, Peach_Healthy, Peach_BacterialSpot, Potato_Healthy, Potato_EarlyBlight, Potato_LateBlight, Tomato_LateBlight, Tomato_Healthy, Grapes_Esca, Grapes_BlackRot, Grapes_Healthy.

2.1 Hybrid CAE–MOCNN Model Architecture

The research has been conducted in two phases. The first phase the CAE is trained for the image compression while retaining necessary features for classification. In the second stage, the compressed yet feature-rich images are given to MOCNN for multiple output classifications.

Stage 1: Convolutional Autoencoder (CAE)

The CAE architecture consists of two primary components: Encoder and Decoder. In encoder, three convolutional blocks with max-pooling, are utilised to continue downsampling the input image resolution while preserving essential features . Whereas the decoder utilizes three convolutional blocks with upscaling to reconstruct the original image. The overall architecture is shown in the Figure 2.1.

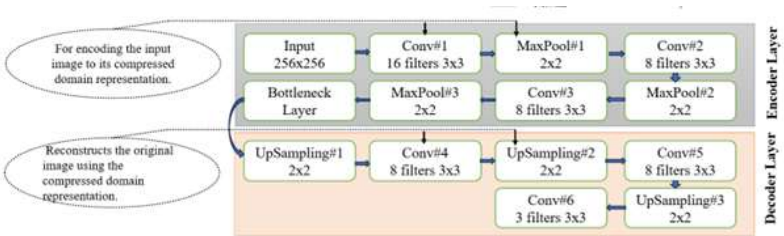

Figure 2.1: The CAE architecture

Stage 2: Multiple Output Classification (MOCNN)

Following the compression of images by the CAE encoder, the resulting low-dimensional representations were used as input to a CNN model structured for Multiple output Classification. The CNN architecture is composed of shared convolutional layers and task-specific branches for species and disease classification. *Shared convolutional layers***:** Two initial convolutional layers with 32 and 64 filters, respectively, followed by batch normalisation and max pooling layers. *Task-specific branches:* The network includes separate branches for each classification task: Species Classification Branch and Disease Classification Branch [9].

3 Training Procedure

The MOCNN was trained initially with optimisation algorithm of Stochastic Gradient Descent, with an initial learning rate set to 0.001 and momentum to 0.9, and a weight decay of 1e−6 to help control overfitting, while a batch size of 10 was used during training. In the second phase, a convolutional autoencoder (CAE) with a total of six layers was trained, where the encoder was composed of three convolutional layers followed by three max-pooling layers, and the decoder followed a similar structure using three convolutional layers along with three upscaling layers. 100 epochs were used to tran CAE with ADAM optimizer.The batch size of 32 was set, and the normalized root mean square error

(NRMSE) was adopted as the reconstruction loss function. For evaluating the models, the performance metrics has been utilized as can be seen in Table 1 and 2. For effective loss calculation while retaining necessary features, NRMSE (Normalized Root Mean Squared Error) is employed. The equation for NRMSE is given below:

$$\text{NRMSE}\ (X^0, X^R) = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X^0 - X^R)}}{Pmax - Pmin}$$

Where, X_0 and X^R are the pixel value of the original and regenerated leaf images for all N number of image pixels. P_{max} and P_{min} are the intensities with maximum and minimumin the input image.

Results and Discussion

The experiment was designed to reduce the complexity of the multi-output CNN (MOCNN) that predicts both species and disease from a single leaf image. Complexity was quantified by the trainable-parameter count and was targeted primarily by narrowing convolutional filters and reducing units in dense layers. The results obtained are presented below:

4.1 Model Efficiency: Trainable parameters and Time

MOCNN: The trainable parameters are majorly contributed by the convolution and dense layers. The MOCNN introduces 19,584 in shared layers, 1,07,523 from species and 5,01,404 from disease layers. Total accumulated to 6,28,511. The model took approximately 180 min to train.

Hybrid CAE–MOCNN*:* The CAE itself contributes 4,163. The accumulated parameters of CAE and MOCNN after using compressed images introduce 3,12,907 trainable parameters. The hybrid model took approximately 70 minutes to train.

4.1 Metric results before Integration of CAE

The results were compiled before and after the integration of CAE to the MOCNN. The Table 1 represent the results obtained before and Table 2 shows the results obtained after the integration.

Table 1: Accuracy, precision, recall and F1 score of MOCNN for classification of species and disease before integration of CAE

Metrics	Species	Disease
Training Accuracy	95.9%	95.00%
Testing Accuracy	94.25%	93.79%
Precision	96.91%	95.65%
Recall	96.06%	95.19%
F1 Score	96.48%	95.41%

4.2 Metric results after Integration of CAE

Table 2: Accuracy, precision, recall and F1 score of MOCNN for classification of species and disease after integration of CAE

Metrics	Species	Disease
Training Accuracy	92.78%	93.36%

Testing Accuracy	91.75%	89.64%
Precision	95.08%	94.14%
Recall	93.75%	92.54%
F1 Score	94.41%	93.33%

4.3 Impact of CAE Integration on Model Performance

Incorporating a convolutional autoencoder (CAE) into the MOCNN framework led to a considerable decrease in model complexity, with only a slight decrease in classification performance across accuracy, precision, recall, and F1 metrics. The parameter count fell from 628,511 to 312,907 (a 50.21% reduction), while training time was shortened from 180 to 70 minutes (61% improvement). The resulting efficiency–performance balance was considered a favourable trade-off, particularly in scenarios with limited computational resources.

5. Conclusion

A hybrid CAE–MOCNN architecture was proposed to classify plant species and associated diseases while significantly reducing computational complexity. By incorporating a convolutional autoencoder (CAE) for dimensionality reduction into the MOCNN framework, the number of trainable parameters was reduced by approximately 50.21%—from 628,511 to 312,907. This reduction came at the cost of a modest decline in test accuracy: of approximately 2.5 percentage points for species classification and 4.15 points for disease identification. However, training time was substantially improved, dropping from 180 to 70 minutes. These results indicate that integrating a CAE into CNN-based architectures can achieve an effective trade-off between computational efficiency and predictive performance, making it a viable approach for deployment in resource-constrained environments. Future work could explore enhancements such as adaptive attention mechanisms and lightweight CNN modules to further improve accuracy while maintaining low model complexity.

REFERENCES

[1] Sharma, Parul, and Pawanesh Abrol. 2022. "Multi-component image analysis for citrus disease detection using convolutional neural networks." *Crop Protection* 193: 107181.

[2] Sun, Xuewei, Guohou Li, Peixin Qu, Xiwang Xie, Xipeng Pan, and Weidong Zhang. 2022. "Research on plant disease identification based on CNN." *Cognitive Robotics* 2: 155-163.

[3] Bedi, Punam, and Pushkar Gole. 2021. "Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network." *Artificial Intelligence in Agriculture* 5 : 90-101.

[4] Chen, Riyao, Haixia Qi, Yu Liang, and Mingchao Yang. 2022. "Identification of plant leaf diseases by deep learning based on channel attention and channel pruning." *Frontiers in plant science* 13: 1023515.

[5] Fu, Rui, Xuewei Wang, Shiyu Wang, and Hao Sun. 2025. "PMJDM: a multi-task joint detection model for plant disease identification." *Frontiers in Plant Science* 16: 1599671.

[6] Khan, Abdullah, Muhammad Zaheer Sajid, Nauman Ali Khan, Ayman Youssef, and Qaisar Abbas. 2025. "CAD-Skin: a hybrid convolutional Neural Network–autoencoder framework for precise detection and classification of skin lesions and cancer." *Bioengineering* 12, no. 4: 326.

[7] Srikanth, T., and M. Radhika Mani. 2025. "Hybrid twin attention based convolutional stacked sparse autoencoder for classification of defected weld images." *Computers and Electrical Engineering* 124: 110328.

[8] Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. 2016. "Using deep learning for image-based plant disease detection." *Frontiers in plant science* 7: 215232.

[9] Yang, B. 2024. "A novel plant type, leaf disease and severity identification framework using CNN and transformer with multi-label method. *Sci. Rep*. 14.