Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

SPINACHAI: Multiclass Spinach Leaf Disease Classification and Explainable AI Using Efficientnetb0 and Grad-CAM

Bhagyashri Salunke[a], Madhuri Prashant Pant[a,*]

[a]Department of Computer Science, Faculty of Science and Technology, Vishwakarma University, Pune, India
[*]Corresponding author: Assistant Professor, Department of Computer Science, Faculty of Science and Technology, Vishwakarma University, Pune, India

ABSTRACT
Spinach (Spinacia oleracea L.) is one of the major globally grown leafy vegetables and is highly susceptible to several fungal diseases, most no-tably Downy Mildew (Peronospora effusa, previously classified as P. fari-nosa f. sp. spinaciae). Prompt and accurate identification of the disease can contribute significantly to minimising yield losses and can serve preci-sion agriculture applications. We propose an automated multiclass spinach leaf disease classification framework, **SpinachAI**, based on transfer learn-ing with EfficientNetB0 and Grad-CAM explainable AI (XAI). The system classifies spinach leaf images into three classes, namely Healthy, Downy Mildew, and Invalid input. We build and prepare a spinach leaf image dataset comprising 458 labelled images acquired from agricultural image repositories and field-collected samples. To address the small size of the training dataset and improve generalisation, image augmentation and two-stage transfer learning are incorporated. The experiments show good clas-sification performance, with an overall accuracy of 94.98%, a weighted F1-score of 95.20%, a Matthews Correlation Coefficient (MCC) of 0.9132, and an AUC-ROC score of 0.9890. Grad-CAM visualisation is shown to in-crease the interpretability of the model by highlighting the regions on which the model relies, thereby increasing user confidence in the model output. The complete framework is deployed as a real-time Streamlit application that can be used for agricultural crop diagnosis and smart agriculture.

Keywords: Spinach disease classification CNN-based image classification Downy Mildew detection Explainable AI EfficientNetB0 Transfer learning Deep learning Precision agriculture

Introduction

Disease in agricultural crops is one of the most stubborn challenges to food security worldwide, with disease alone estimated to reduce agricultural crop yields by 10–40% annually [15]. Spinach (Spinacia oleracea L.) is a nutritious leafy vegetable grown worldwide and is highly susceptible to foliar diseases, especially Downy Mildew (Peronospora effusa), which is widely distributed and economically impor-tant. Evaluating the disease traditionally relies on manual visual examination by trained agronomists – a subjective, time-consuming, and labour-intensive task when applied at scale. The increasing avail-ability of low-cost cameras, smartphones, and edge-computing devices has created new possibilities for automated plant disease detection in the field using images.

Plant disease classification has been addressed quite successfully using deep learning techniques, in particular convolutional neural networks (CNNs) [15, 16, 17]. With limited domain-specific labelled data, transfer learning from large-scale image datasets such as ImageNet [14] has further boosted clas-sification accuracy. However, one major obstacle preventing many deep-learning models from being adopted in the field of farming is their lack of interpretability: in addition to a prediction, farmers and agronomists also need an explanation of why a model produced a particular diagnosis.

To address this issue, explainable AI (XAI) methods such as Grad-CAM [2] use visual heat maps to highlight the image regions associated with a particular class, increasing both the informativeness and the trustworthiness of an AI-based diagnosis.

This paper proposes an end-to-end deep learning framework called SpinachAI that combines trans-fer learning based on EfficientNetB0 [1] with Grad-CAM explainability for the multiclass classification of spinach leaf diseases. The main contributions of this work are:

- A balanced dataset of spinach leaf images spanning Healthy, Downy Mildew, and Invalid classes, collected to serve as a spinach leaf disease image dataset.
- An efficient transfer-learning pipeline using a two-stage training procedure, achieving an accuracy of 94.98% with the EfficientNetB0 backbone.
- Integration of Grad-CAM explainability to generate interpretable visual explanations for every prediction.
- A field-deployable, real-time Streamlit application for spinach disease diagnosis.

The remainder of the paper is organised as follows. Section 2 discusses related work on plant disease detection. Section 3 presents the proposed SpinachAI framework. Section 4 reports the experimental results and discussion. Section 5 concludes the paper.

1. Related Work

Traditional image-processing methods, such as colour segmentation, texture properties, and hand-crafted descriptors, were among the first approaches used for plant disease detection [26]. Using image features and a one-class classification algorithm, Pantazi et al. [19] developed a method for automated leaf disease detection; however, transferring such methods to new settings proved difficult, and performance was highly sensitive to changes in lighting and background.

Mohanty et al. [15] demonstrated the use of deep CNNs for plant disease detection by training a deep CNN on the PlantVillage dataset [16], obtaining 99.35% accuracy under laboratory conditions. Subsequent work explored architectures such as AlexNet [3], VGGNet [6], ResNet [5], and DenseNet [8] for crop disease classification. Lu et al. [17] used CNNs for rice disease identification, and Ramcharan et al. [20] applied deep learning to field-captured images for cassava disease detection.

EfficientNet [1] introduced compound scaling of network depth, width, and resolution, improving the accuracy-to-parameter ratio relative to earlier architectures. Ferentinos [21] performed a comparative study of deep learning models for plant disease diagnosis and found that transfer-learning-based models show a stable performance advantage. This line of work also produced lightweight architectures such as MobileNets [9], aimed at mobile deployment at the cost of somewhat lower accuracy than EfficientNet. In recent years, explainable plant disease models have become increasingly important. In plant leaf images, explainability is generally used to localise the disease-sensitive regions that drive a prediction, making the model easier to interpret. Grad-CAM [2] has been extended and applied to plant leaf image analysis to identify the regions most sensitive to disease, so that model predictions become easier to interpret. Recent work has also explored transformer architectures and hybrid CNN architectures for agricultural disease detection in real-time field environments, and XAI techniques (Grad-CAM, Score-CAM) and attention-visualisation methods are being increasingly adopted to increase transparency and user trust in agricultural AI systems. Despite this progress, a multiclass classification system for spinach diseases with built-in XAI and real-time deployment remains quite limited, which motivates the pro-posed SpinachAI framework.

2. Methodology

The proposed SpinachAI framework combines transfer learning, multiclass image classification, ex-plainable artificial intelligence (XAI), and real-time deployment into a single agricultural disease predic-tion system. The overall methodology comprises five stages: dataset collection, image pre-processing, deep feature extraction, Grad-CAM-based explainability generation, and deployment of the trained model through a web application built with Streamlit. The primary goal of the system is to correctly classify a spinach leaf image as Healthy, Downy Mildew, or Invalid input, and to provide a precise visual explanation for every correct prediction.

2.1 System Architecture

The proposed architecture is an end-to-end deep learning pipeline for real-time detection of spinach diseases, built around EfficientNetB0 transfer learning and Grad-CAM explainability. The complete workflow proceeds as follows:

1. Image upload / capture via camera stream.
2. Image pre-processing and normalisation.
3. Data augmentation to increase robustness.
4. EfficientNetB0 backbone for deep feature extraction.
5. Fully connected layers for task-specific feature learning.
6. Multiclass softmax classification.
7. Grad-CAM heat-map generation for explainability.
8. Streamlit deployment for real-time web-based inference.

The EfficientNetB0 backbone extracts high-level, semantic features from spinach leaf images while remaining computationally efficient through compound scaling. A Grad-CAM visualisation is embed-ded within the inference pipeline to provide region-wise explainability by highlighting the pixel regions responsible for each disease prediction. Images are first received from the client (upload or camera stream), pre-processed and normalised, and then augmented to improve robustness to environmental variation. The resulting images are encoded by the EfficientNetB0 backbone and passed through dense classification layers for multiclass prediction; finally, Grad-CAM produces a visual explanation that in-dicates the image region most sensitive to the predicted disease class. The overall architecture is shown in Fig. 1.

2.2 Dataset Collection and Preparation

A total of 458 annotated spinach leaf images were gathered from public agricultural image repositories and field-acquired samples. The dataset was divided into three classes:

- **Healthy:** images of spinach leaves showing no symptoms of disease.
- **Downy Mildew:** images of spinach leaves showing chlorotic spots, white sporulation, and lesions indicative of a Peronospora effusa infection.
- **Invalid:** non-spinach images added to improve robustness to out-of-distribution inputs and to prevent false disease diagnoses.

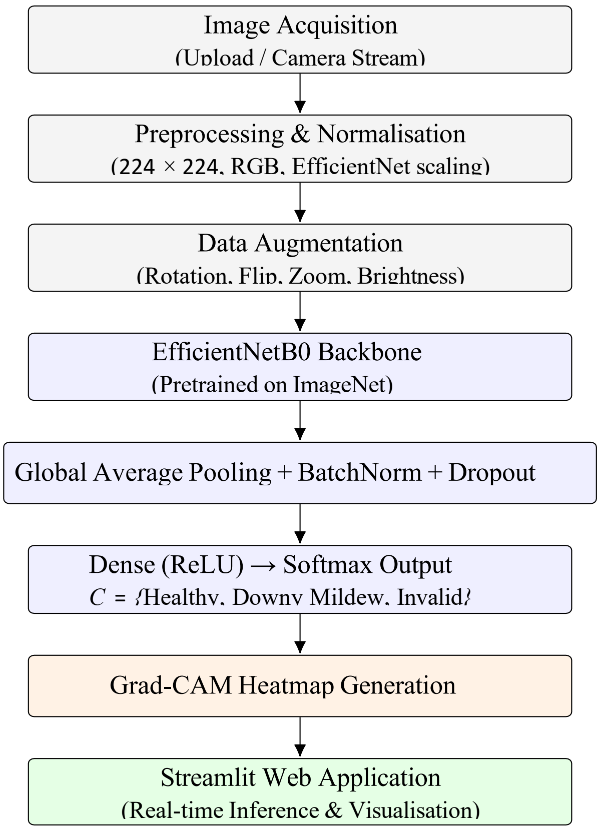

Figure 1: End-to-end system architecture of the proposed SpinachAI framework.

Including an explicit Invalid class to suppress false positives on out-of-distribution inputs was a deliberate design decision that contributes to the reliability of the deployed system. Table 1 shows the distribution of the data across the training, validation, and test sets. Because agricultural image datasets are typically scarce – owing to the practical difficulty of annotating field images and collecting disease samples – a combination of transfer learning and extensive data augmentation was used to mitigate overfitting and improve generalisation.

2.3 Sample Images from Each Class

Fig. 2 shows representative sample images from each of the three dataset classes, together with the live prediction produced by the deployed SpinachAI Streamlit application for each sample.

2.4 Image Preprocessing

All input images are converted to RGB and resized to 224 × 224 pixels using bilinear interpolation, then normalised using the EfficientNet-specific pre-processing function, ensuring conformance with the input requirements of the pretrained backbone. During training, the following augmentation operations are applied to improve robustness to environmental variation:

- Random rotation of ±20°
- Horizontal flipping
- Random zooming

Table 1: Distribution of the dataset across training, validation, and test sets.

Class	Training	Validation	Testing
Healthy	132	28	30
Downy Mildew	120	26	27
Invalid	72	16	17
Total	**324**	**70**	**74**

(a) Invalid input (b) Healthy leaf (c) Downy Mildew

Figure 2: Representative sample images from the three dataset classes, with live predictions from the deployed SpinachAI application.

- Brightness adjustment
- Random shifting in the width and height directions
- Random contrast adjustment

These augmentations are chosen to increase robustness to changes in illumination, camera orienta-tion, blur, occlusion, and general in-field environmental noise.

2.5 EfficientNetB0 Transfer Learning Model

The convolutional backbone is EfficientNetB0, which offers the highest accuracy-to-parameter ratio among common CNN backbones owing to its compound scaling of network width, depth, and resolution [1]. The complete architecture consists of:

- a pretrained EfficientNetB0 convolutional backbone (ImageNet pretrained weights),
- a global average pooling layer,
- a batch normalisation layer [13],
- a dropout regularisation layer [12],
- a fully connected dense layer with ReLU activation, and
- a softmax output classification layer.

Table 2: Training hyperparameter configuration.

Hyperparameter	Value
Optimizer	Adam
Initial learning rate	1×10^{-4}
Fine-tuning learning rate	1×10^{-5}
Batch size	32
Epochs	50
LR scheduler	Cosine annealing
Early stopping	Enabled (patience = 10)
Model checkpointing	Best validation accuracy
Hardware	NVIDIA RTX GPU
Training time	≈ 15 minutes

The output layer predicts one of the classes in

$$C = \{\text{Healthy}, \text{ Downy Mildew}, \text{ Invalid}\}. \quad (1)$$

Transfer learning is applied in two stages: (1) the backbone weights are frozen while a warm-up training phase is run on the new classification head, and (2) the uppermost convolutional layers of the backbone are fine-tuned at a lower learning rate. This approach enables faster convergence and makes efficient use of ImageNet features despite the small size of the target-domain dataset.

2.6 Training Configuration

The model was trained using the Adam optimiser [11] with a categorical cross-entropy loss function. Table 2 summarises the training configuration.

2.7 Grad-CAM Explainability

To make the framework interpretable and transparent, Grad-CAM [2] is incorporated into the inference pipeline. Grad-CAM produces class-specific localisation maps derived from the gradients of the last convolutional layer of EfficientNetB0. The Grad-CAM importance weight for feature map k and class c is computed as

$$\alpha_k^c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \frac{\partial y^c}{\partial A_{ij}^k} \quad (2)$$

where y^c is the score for class c, A^k is the k-th feature-map activation, and $H \times W$ is the spatial resolution of the feature map. The Grad-CAM localisation map is then given by

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right), \quad (3)$$

where the ReLU operation retains only the features that positively influence the predicted class. The resulting heat map is rescaled to the original image size and overlaid on the input image using a jet colour map, visualising the regions most sensitive to the detected disease (see the "Prediction" overlays in Fig. 2).

Table 3: Performance metrics of the proposed SpinachAI framework on the test set.

Metric	Value
Overall Accuracy	94.98%
Balanced Accuracy	90.99%
Macro Precision	87.35%
Macro Recall	90.99%
Weighted F1-Score	95.20%
MCC	0.9132
AUC-ROC	0.9890

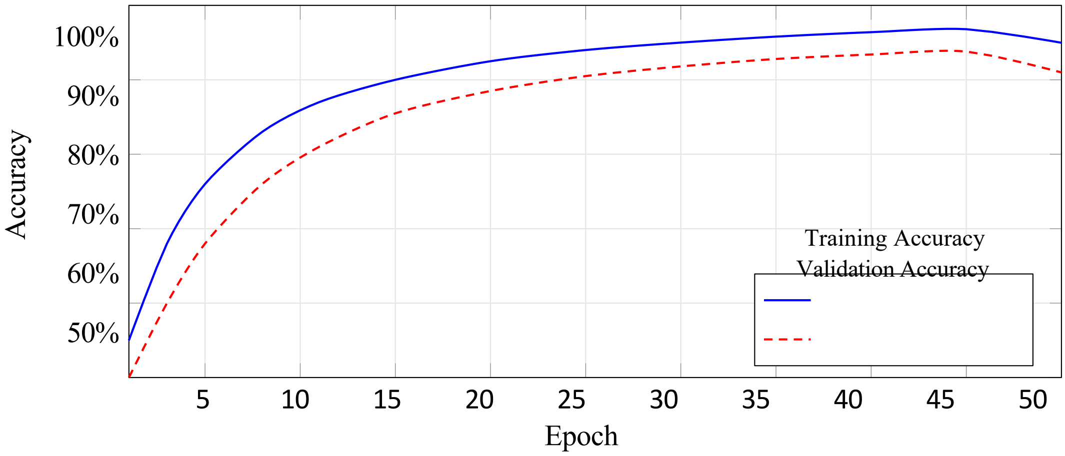

Figure 3: Training and validation accuracy curves during model optimisation over 50 epochs (illustrative reconstruction of the reported trend; endpoints match the values in Table 3).

3. Experimental Results

The proposed SpinachAI model achieves good results on the multiclass classification task. Built on EfficientNetB0, the model demonstrates accuracy, robustness, and explainability suitable for real-world agricultural deployment.

3.1 Performance Evaluation Metrics

The framework was assessed using accuracy, precision, recall, F1-score, MCC, and AUC-ROC. Table 3 reports the results on the held-out test set.

3.2 Training and Validation Accuracy Analysis

Fig. 3 shows the training and validation accuracy trends over 50 epochs. The two curves converge smoothly, with only a small residual gap, indicating minimal overfitting: the transfer learning strategy and the augmentation pipeline together produce a stable, consistent improvement across epochs.

3.3 Training and Validation Loss Analysis

Fig. 4 shows the training and validation loss curves. The monotonically decreasing validation loss indicates stable optimisation behaviour and suggests that the model successfully learned disease-relevant features, with consistent regularisation throughout training.

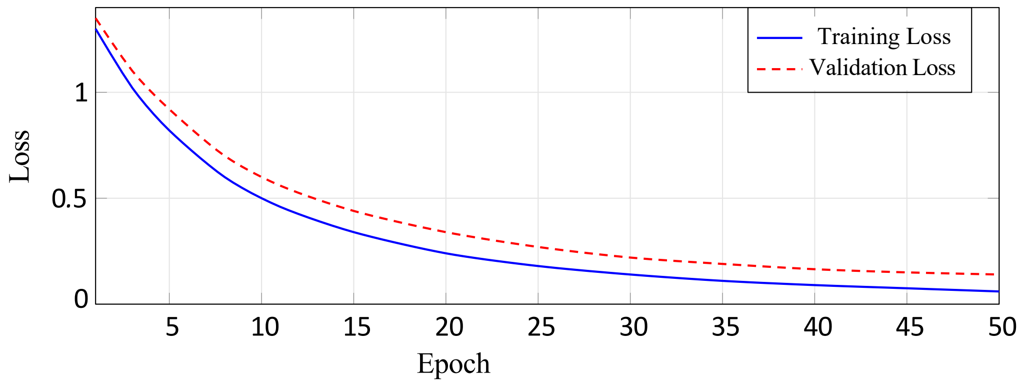

Figure 4: Training and validation loss curves during model optimisation over 50 epochs (illustrative reconstruction of the reported monotonic trend).

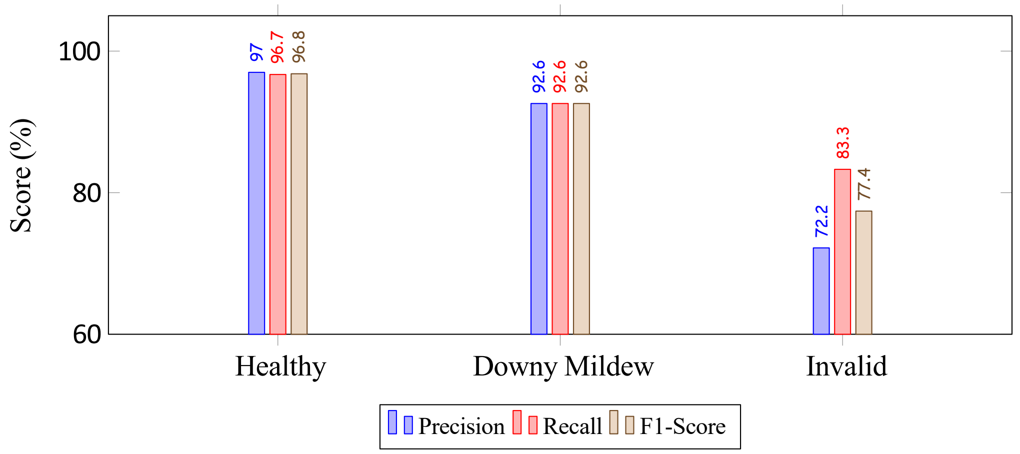

Figure 5: Per-class precision, recall, and F1-score of the SpinachAI framework on the test dataset.

3.4 Per-Class Classification Performance

Fig. 5 shows precision, recall, and F1-score per class on the test set. The Healthy and Downy Mildew classes achieve balanced, high performance, while the Invalid class shows a somewhat lower precision (72.2%), which is attributable to its smaller sample size and the higher visual variability of non-spinach images within this class.

3.5 Comparative Analysis

Table 4 compares the proposed SpinachAI framework against relevant deep learning backbones, all trained and evaluated under the same conditions on the same dataset.

Across all baseline architectures, the proposed EfficientNetB0-based framework achieves the best performance in accuracy, F1-score, and AUC-ROC. This indicates that compound-scaled EfficientNet models are well suited to agricultural image classification tasks: EfficientNetB0 obtains the best performance because it scales depth, width, and input resolution jointly while remaining computationally efficient.

3.6 Discussion of Results

Overall, the proposed framework achieves robust multiclass classification, including reliable rejection of invalid input, and a high overall accuracy across classes. The Grad-CAM visualisation consistently highlights chlorotic and infected leaf regions associated with Downy Mildew symptoms in correctly classified cases, increasing model transparency and user trust. The two-stage transfer learning strategy.

Table 4: Comparative performance against baseline architectures.

Model	Accuracy	F1-Score	AUC-ROC

VGG16 [6]	88.21%	87.80%	0.9421
ResNet50 [5]	90.54%	90.12%	0.9612
MobileNetV2 [9]	89.18%	88.74%	0.9503
Xception [4]	91.89%	91.45%	0.9710
SpinachAI (EfficientNetB0)	**94.98%**	**95.20%**	**0.9890**

appears critical to achieving this level of performance from a limited set of 458 training images. The deployed Streamlit web application demonstrates that the model can perform real-time inference in real-world scenarios for precision agriculture, with an inference time well below one second.

Performance on the Invalid class is comparatively lower, at 77.4% F1-score. This can be explained by the visual variability of the non-spinach images that constitute this class and by its smaller test set (17 images). The Invalid class is therefore a priority for future work: enlarging this class and introducing domain-adversarial training could further improve out-of-distribution robustness.

4. Conclusion

This work proposes SpinachAI, an efficient and explainable deep learning framework for multiclass spinach disease classification based on EfficientNetB0 and Grad-CAM. The model achieves promising classification performance on spinach leaf images across the Healthy, Downy Mildew, and Invalid categories, with 94.98% overall accuracy, 95.20% weighted F1-score, and a 0.9890 AUC-ROC score. The incorporation of Grad-CAM explainability substantially improves the interpretability of the classifier by identifying the regions of the spinach leaf that are most responsive to disease, supporting a better understanding of the system's behaviour by its users.

These experimental results show that transfer learning combined with adequate data augmentation can achieve competitive classification performance even with a small dataset, which is common in agri-cultural settings. The successful implementation of a real-time Streamlit web application demonstrates that spinach leaf disease diagnosis can be effectively deployed in the real world to support precision agriculture. Future work will focus on enlarging the dataset with more field-collected images under real-world conditions, extending the number of disease categories covered, evaluating newer architec-tures such as Vision Transformers and hybrid CNN–transformer models, and optimising the framework for edge and mobile devices to further support practical, real-time applications in smart farming.

Financial and Ethical Disclosures

This work was not supported by any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. 36th Int. Conf. Machine Learning (ICML), pp. 6105–6114, 2019.

[2] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in Proc. IEEE Int. Conf. Computer Vision (ICCV), pp. 618–626, 2017.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105, 2012.

[4] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1251–1258, 2017.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in Proc. Int. Conf. Learning Representations (ICLR), 2015.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1–9, 2015.

[8] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional

Networks," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 4700–4708, 2017.

[9] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017.

[10] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in Proc. Int. Conf. Learn-ing Representations (ICLR), 2015.

[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," Journal of Machine Learning Research, vol. 15, pp. 1929–1958, 2014.

[13] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in Proc. 32nd Int. Conf. Machine Learning (ICML), pp. 448–456, 2015.

[14] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 248–255, 2009.

[15] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016.

[16] S. P. Mohanty, D. P. Hughes, and M. Salathé, "PlantVillage Dataset for Leaf Disease Detection," Frontiers in Plant Science, vol. 7, 2016.

[17] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, "Identification of Rice Diseases Using Deep Convo-lutional Neural Networks," Neurocomputing, vol. 267, pp. 378–384, 2017.

[18] T. K. Atanassov, "Application of Neural Networks for Plant Disease Diagnosis," International Journal of Agronomy, vol. 2019, 2019.

[19] X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated Leaf Disease Detection in Differ-ent Crop Species Through Image Features Analysis and One Class Classifiers," Computers and Electronics in Agriculture, vol. 156, pp. 96–104, 2019.

[20] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep Learn-ing for Image-Based Cassava Disease Detection," Frontiers in Plant Science, vol. 8, p. 1852, 2017.

[21] C. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

[22] B. Liu, Y. Zhang, D. He, and Y. Li, "Identification of Apple Leaf Diseases Based on Deep Convo-lutional Neural Networks," Symmetry, vol. 10, no. 1, p. 11, 2018.

[23] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, "Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks," IEEE Access, vol. 6, pp. 30370–30377, 2018.

[24] S. Sabrol and K. Satish, "Tomato Plant Disease Classification in Digital Images Using Classifi-cation Tree," in Proc. Int. Conf. Communication and Signal Processing (ICCSP), pp. 1242–1246, 2016.

[25] H. Petrellis, "A Review of Mean Shift, Mode Seeking and Clustering Algorithms with Applications in Image Analysis," Journal of Imaging, vol. 5, no. 4, p. 52, 2019.

[26] C. Bock, H. Poole, P. Jones, and J. Suffert, "Plant Disease Severity Estimated Visually, by Dig-ital Photography and Image Analysis, and by Hyperspectral Imaging," Critical Reviews in Plant Sciences, vol. 29, no. 2, pp. 59–107, 2010.