Hybrid CNN-SVM Models for Seedling Classification

Wan-Xuan Ang1, a), Lee-Ying Chong1, 2, b), Siew-Chin Chong1, 2, c),   
Pey-Yun Goh1, 2, d)

1Faculty of Information Science and Technology

*2Centre for Advanced Analytics, COE of Artificial Intelligence  
Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia*

b) Corresponding author: lychong@mmu.edu.my

a) wanxuan0207@gmail.com

c) chong.siew.chin@mmu.edu.my

d) pygoh@mmu.edu.my

**Abstract.** Modern agriculture requires precise plant identification for effective crop management and weed control. This study evaluates deep learning methods for automated seedling classification using the Aarhus University dataset (5,539 images of 12 species). We compare the performance of four state-of-the-art CNN architectures (EfficientNetB3, DenseNet201, ResNet50, VGG16) combined with SVM classifiers, focusing on their ability to distinguish between crop seedlings and weeds—a key challenge in early-stage field management. Our experiments demonstrate that hybrid CNN-SVM models achieve exceptional accuracy, with EfficientNetB3 and DenseNet201 both reaching 95.89% classification accuracy, outperforming other configurations. While the system excelled at general classification, it faced challenges distinguishing visually similar species like Black-grass and Common Wheat. These results demonstrate the potential of computer vision for agricultural applications while highlighting areas for improvement in fine-grained plant recognition. The findings highlight the potential of computer vision in agriculture, particularly for weed control and crop monitoring. However, challenges remain in differentiating closely related species, suggesting the need for improved feature extraction or data augmentation techniques. Further work could explore larger datasets, advanced architectures, or multi-modal approaches to enhance accuracy in real-world field conditions.

# Introduction

Plant seedling classification is crucial for precision agriculture yet remains challenging due to visual similarities between species at early growth stages [1]. We tested deep learning methods on species recognition on the Aarhus University dataset of 5.539 seedling images belonging to 12 species. We compare hybrid architectures that involve four pretrained CNNs (EfficientNetB3, DenseNet201, VGG16, ResNet50) with SVM classifiers, based on the previous studies that have established the effectiveness of CNNs over the conventional feature-based techniques. In a systematic fashion, our approach extracts hierarchical features at various depths of the network, optimizes the hyperparameters of SVM using grid search, and tests on difficult-to-distinguish species pairs. The findings indicate that EfficientNetB3 and DenseNet201 combined with RBF-kernel SVMs attain the highest accuracy of 95.89% compared to other settings, thus illustrating certain weaknesses in distinguishing between similar grass species. The results propel computer vision methods in agricultural tasks besides offering useful guidelines to realistic implementation [2]. The work provides a contribution to the continued research efforts in the area of automated weed detection and crop monitoring systems by providing evidence of effectiveness of deep feature extraction with optimized classifiers.

# LITERATURE REVIEW

To classify the weed and corn seedlings at an early stage, Lanlan Wu et al. [3] used a Support Vector Machine (SVM) after extracting texture features of the grayscale-converted images depending on the RGB components. With feature extraction and dimensionality reduction using Gray-Level Co-occurrence Matrix (GLCM) and Principal Component Analysis (PCA), the SVM got 100 percent accuracy, whereas a backpropagation (BP) neural network got only 80 percent accuracy. Until Rumpf et al. [4] sequential SVM technique enhanced the classification of weed species on Cirsium arvense and Galium aparine. They achieved 97.74% accuracy with a linear SVM in the first step and 80% accuracy with a non-linear SVM using a radial basis function (RBF) kernel in subsequent steps, utilizing an image series covering different growth stages. Giselsson et al. [5] introduced novel shape features (Legendre Polynomial Feature Set) to classify cornflower and nightshade seedlings at the BBCH 12 growth stage. Their RBF-SVM model achieved 97.5% accuracy on a dataset comprising 139 cornflowers and 63 nightshade images. André Dantas de Medeiros et al. [6] combined FT-NIR spectroscopy and X-ray imaging to assess Urochloa brizantha seed quality. Their random forest model achieved 85% accuracy in germination prediction and 62% accuracy in vigor classification. Wenjian Liu et al. [7] evaluated Partial Least Squares Discriminant Analysis (PLSDA), SVM, and Deep Neural Network (DNN) models for classifying seed vigor in five tree species (348 NIR spectra samples). DNN delivered the best performance, achieving 100% accuracy in calibration and 96% in validation, surpassing PLSDA (88% accuracy) and SVM (89% accuracy).

Jiachun Liu et al. [8] developed a 10-layer CNN for leaf classification using the Flavia dataset, achieving 87.92% accuracy after augmentation and preprocessing, outperforming traditional SVM (80%) by using the Flavia dataset. Villaruz et al. [9] developed a deep learning approach to classify Philippine indigenous plant seedlings into five species using 2,500 images captured in natural environments, with ResNet50 achieving the highest accuracy of 98.93%, followed by GoogLeNet at 94.43% and AlexNet at 91.67%. Chavan et al. [10] proposed AgroAVNET, a hybrid CNN model combining AlexNet and VGGNET features, to classify weeds and crops from the public plant seedlings dataset, achieving an overall accuracy of 96.7%. Hmidi Alaeddine and Malek Jihene [11] enhanced AlexNet into DbneAlexNet for plant disease classification, achieving 99.48% accuracy on the Plant Village dataset via 3×3 convolutions, eLU activation, and batch normalization. Recently, Pallavi et al. [12] proposed a VGG16-based transfer learning model for classifying medicinal plant leaves across 35 species, achieving 93% accuracy using augmented image resize, outperforming ResNet50 (87.54%) and DenseNet121 (88.53%).

# METHODOLOGY

Figure 1 illustrates the Proposed Methodology for Plant Seedling Classification, which follows a structured and modular pipeline designed to ensure accurate and efficient classification of plant seedlings into their respective categories. The process begins with the Input Data, which typically consists of raw images of plant seedlings. These images undergo a Resizing step to standardize the input dimensions, ensuring compatibility with the subsequent deep learning model. Next is the Segmentation phase wherein background information that is not necessary is eliminated or reduced and the model is able to concentrate on the actual features of the seedlings and learning is more effective.

After the preprocessing step, the images are fed into a CNN-based Pretrained Model, which does Deep Feature Extraction. This step utilizes the convolutional layers of the pretrained model to automatically detect complex visual patterns, textures, and structures inherent to different plant species. Support Vector Machine (SVM) Classifier is then applied on these deep features that are more informative and discriminative than the raw pixel data. The final classification of the seedlings using the extracted features is done using the SVM, which is a potent supervised learning algorithm.

## Figure 1. Proposed methodology for plant seedling classification

Image Preprocessing

Lastly, the performance of the model is measured using Evaluation Metrics, accuracy, precision, recall, and F1-score, which give a thorough evaluation of the model in terms of its distinction between the classes of seedlings. The whole pipeline illustrated in Figure 1 shows a unified mechanism that incorporates the advantages of deep learning in feature extraction and traditional machine learning in classification to provide a feature-rich solution to the problem of plant seedling identification.

## Image Preprocessing

During image preprocessing raw plant seedling images were taken through a series of processing to remove the background and bring clarity on the seedling parts. This was done by using segmentation to extract the foreground (plant) and background. The cleaned-up images (background erased) were further stored in a compressed ZIP file to facilitate their storage and future utilization in training and analysis of models. This preprocessing provided cleaner input data with less noisy data and better classification.

## CNN Pretrained Extraction

Each model in this paper will use a standardized and unified feature extraction pipeline to provide consistency and efficiency across the architectures. All images are resized to 256 256 pixels, turned into an array and normalized with the model-specific preprocess\_input() functions that are matched to their model-specific ImageNet-pretrained weights. After preprocessing the images, convolutional backbones (EfficientNetB3, DenseNet201, VGG16, and ResNet50) are used to extract deep feature representations of these preprocessed images. This is done consistently across features by the extract\_features() function, which returns flattened feature vectors that can be used in subsequent tasks (e.g. classification or retrieval). This consistent pipeline not only streamlines the feature extraction process but also facilitates fair comparative analysis across different model architectures.

### EfficientNetB3

EfficientNetB3 sets a new state-of-the-art performance but is also computational efficient due to compound scaling (uniform scaling of network width, depth, and resolution). The design includes mobile inverted bottleneck convolution (MBConv) blocks that have squeeze-and-excitation (SE) modules and provide feature representation with parameter reduction. To extract features, middle layers (block4a\_activation and block6a\_activation) are chosen to obtain multi-scale hierarchical features. These features are pooled using GlobalAveragePooling2D () and flattened into a small feature vector, which can be used in transfer learning applications, such as classification or retrieval.

### DenseNet201

DenseNet201 is based on the idea of dense connectivity to maximize feature reuse and gradient flow between layers, in which every layer is directly fed by all previous layers, and feeds its feature maps to all following layers, resulting in an efficient information highway. For feature extraction, strategically selected layers (conv2\_block6\_concat for low-level textures, conv3\_block12\_concat for mid-level patterns, and conv5\_block32\_concat for high-level semantics) are concatenated after pooling, yielding rich multi-scale descriptors from its 1,920-channel final blocks.

### VGG16

VGG16 is a classical convolutional neural network architecture renowned for its simplicity and effectiveness, consisting of 16 weight layers organized into five sequential blocks of 2-3 convolutional layers followed by max-pooling layers that progressively reduce spatial dimensions. For feature extraction, intermediate layers block4\_conv3 (capturing mid-level part-based features at 28×28×512 resolution) and block5\_conv3 (encoding high-level semantic information at 14×14×512 resolution) are selected, as they provide complementary visual representations that balance spatial detail and abstract understanding. These features are transformed into compact descriptors through GlobalAveragePooling2D and concatenated into a 1024-dimensional feature vector.

### ResNet50

ResNet50 is a 50-layer deep convolutional neural network that revolutionized deep learning through its innovative use of skip connections, which mitigate vanishing gradients by enabling unimpeded gradient flow via identity mappings. For feature extraction, intermediate outputs like conv3\_block4\_out (14×14×1024) and conv4\_block6\_out (7×7×2048) are typically selected—the former captures mid-level part-based features, while the latter encodes high-level semantic information. These features are pooled and concatenated into discriminative 3,072-dimensional vectors that combine localized part-based details with holistic representations.

## SVM Classifier

The Support Vector Machine (SVM) classifier is a supervised learning model that aims to find the optimal hyperplane that best separates data points of different classes in a high-dimensional feature space. In this study, the SVM classifier is applied after feature extraction, utilizing the deep feature vectors generated by pretrained CNN backbones such as ResNet50, EfficientNetB3, VGG16, and DenseNet201. These rich, high-level features are well-suited for SVM, which excels in handling complex, non-linearly separable data by kernel functions. The classifier works by maximizing the margin between support vectors of different classes, ensuring robust generalization. Figure 2 shows the architecture of SVM in which the input feature vector x is first passed through an input layer, then into a high-dimensional space (implicitly or explicitly, depending on the kernel) and finally it is employed to define a decision function in terms of support vectors and a bias term b giving the final output. This architecture offers a flexible and effective image classification system that can be scaled and used effectively particularly with the strong representation features of deep learning.

**A diagram of a machine learning

AI-generated content may be incorrect.**

**Figure 2.** Architecture of support vector machines

# Evaluation Metric

A set of metrics was used to assess the model effectiveness aimed at handling the possible class imbalance and deliver more subtle information beyond accuracy. Whereas accuracy in Equation 1 was a measure of overall correctness, precision in Equation 2 was a measure of the model not making false positives, and recall in Equation 3 was a measure of the sensitivity of the model to true positives. As a main evaluation measure, the F1-score in Equation 4 was used, as it balances both issues that are particularly important in the case of imbalanced seedling classes. The classification report showed detailed class-wise performance measures including precision, recall and F1-score along with macro and micro averages, which could be further used to analyze the model behavior on all 12 plant species specifically. Such a multi-metric approach provided a solid evaluation, focusing on generalizability in terms of accuracy and the ability to detect minority-classes in terms of F1-score.

|  |  |
| --- | --- |
|  | (1) |

|  |  |
| --- | --- |
|  | (2) |

|  |  |
| --- | --- |
|  | (3) |

|  |  |
| --- | --- |
|  | (4) |

# RESULT AND DISCUSSION

## Dataset Used

The dataset published by the Aarhus University Signal Processing group in collaboration with the University of Southern Denmark [13], is well suited to the task of plant seedling classification. It comprises about 5,539 different plant images belonging to 12 different species and taken at different stages of growth. The data is separated into training and testing data. This dataset supports the development of classifiers that can determine a plant's species based on its photograph.

# Experimental Setup

This study investigates the effectiveness of combining deep feature extraction from pretrained convolutional neural networks (CNNs) with a Support Vector Machine (SVM) classifier to perform multi-class classification of plant seedlings. The experimental pipeline consists of image preprocessing, feature extraction using four different pretrained models, feature standardization, and classification using SVM with hyperparameter optimization. All experiments were conducted using Python and TensorFlow in a Google Colab environment.

We employed four pretrained CNN models (DenseNet201, EfficientNetB3, VGG16, and ResNet50) as fixed feature extractors. All models were initialized with ImageNet weights and their top classification layers were removed (include\_top=False). Intermediate layers were selected based on their semantic depth (shallow, mid, and deep), as detailed in Table 1. Selected layers for feature extraction from each pretrained model, and their outputs were passed through Global Average Pooling to reduce spatial dimensions. These pooled features were then concatenated to form a single feature vector per image, which was subsequently flattened. The feature vectors were standardized using StandardScaler to normalize the feature space. The dataset was split into 80% training and 20% validation sets, maintaining class stratification for balanced representation. Classification was performed using SVM, with hyperparameters optimized via Grid Search (5-fold cross-validation). The parameter grid included C: [0.1, 1, 10], kernel: ['linear', 'rbf'], and gamma: ['scale', 'auto'].

## TABLE 1. Selected layers for feature extraction from each pretrained

|  |  |
| --- | --- |
| **Pretrained Model** | **Selected Layers** |
| EfficientNetB3 | block4a\_activation, block6a\_activation |
| DenseNet201 | conv2\_block6\_concat, conv3\_block12\_concat, conv5\_block32\_concat |
| VGG16 | block4\_conv3, block5\_conv3 |
| ResNet50 | conv3\_block4\_out, conv4\_block6\_out |

# Experimental Result

All four pretrained CNN models integrated with SVM classifiers demonstrated strong classification performance on the Plant Seedling dataset, with detailed class-wise metrics presented in Table 2. EfficientNetB3+SVM and DenseNet201+SVM achieved the highest overall accuracy of 95.89% using RBF kernels (C = 10, gamma = 'scale'). These models exhibited excellent generalization, achieving perfect precision and recall (1.00/1.00) for Charlock, Small-flowered Cranesbill, and Maize, and near-perfect scores for other classes like Common Chickweed and Scentless Mayweed. However, both models showed lower performance for visually similar classes such as Black-grass (recall: 0.74) and Common wheat, though DenseNet201 slightly outperformed EfficientNetB3 in recall for the latter (0.93 vs. 0.91).

ResNet50+SVM followed closely with 95.05% accuracy, showing competitive class-wise results, including perfect scores for Charlock and Small-flowered Cranesbill. Despite slightly reduced precision and recall for Black-grass (0.73/0.68) and Common wheat (0.95/0.91), it remained consistent across most categories. VGG16+SVM, using a linear kernel (C = 0.1), recorded the lowest accuracy at 93.26% and showed weaker performance for Black-grass (0.58/0.68) and Loose Silky-bent (0.87/0.83), though it still achieved perfect scores in a few classes. Overall, deeper CNN models paired with RBF-based SVMs offered more robust and consistent class-wise results than the shallower VGG16, especially for complex, multi-class classification tasks involving subtle visual differences.

## TABLE 2. Precision and recall scores using CNN features and SVM

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Type** | **EfficientNetB3+SVM** | | **DenseNet201+SVM** | | **ResNet50+SVM** | | **VGG16+SVM** | |
| **precision** | **recall** | **precision** | **recall** | **precision** | **recall** | **precision** | **recall** |
| Black-grass | 0.96 | 0.74 | 0.75 | 0.74 | 0.73 | 0.68 | 0.58 | 0.68 |
| Charlock | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Cleavers | 1.00 | 0.97 | 1.00 | 0.98 | 1.00 | 0.95 | 1.00 | 0.95 |
| Common Chickweed | 0.97 | 0.98 | 0.98 | 0.99 | 0.98 | 0.98 | 0.97 | 0.98 |
| Common wheat | 0.95 | 0.91 | 0.95 | 0.93 | 0.95 | 0.91 | 0.95 | 0.84 |
| Fat Hen | 0.98 | 1.00 | 0.98 | 0.97 | 0.96 | 0.98 | 0.97 | 0.98 |
| Loose Silky-bent | 0.90 | 0.92 | 0.89 | 0.91 | 0.88 | 0.91 | 0.87 | 0.83 |
| Maize | 0.98 | 1.00 | 0.98 | 1.00 | 0.98 | 1.00 | 0.98 | 0.98 |
| Scentless Mayweed | 0.96 | 0.98 | 0.98 | 0.99 | 0.96 | 0.98 | 0.95 | 0.96 |
| Shepherds Purse | 1.00 | 0.93 | 1.00 | 0.96 | 1.00 | 0.93 | 0.98 | 0.96 |
| Small-flowered Cranesbill | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Sugar beet | 0.99 | 0.99 | 0.96 | 0.97 | 0.95 | 0.97 | 0.94 | 0.96 |

# Discussion

Table 3 presents the classification results of various CNN feature extractors combined with SVM. The EfficientNetB3+SVM and DenseNet201+SVM models achieved the highest accuracy of 95.89% and identical macro F1-scores of 0.95, excelling in handling distinct plant species but showing minor confusion between similar grass types like Black-grass and Common wheat. ResNet50+SVM, with an accuracy of 95.05%, also showed strong performance but had slightly lower precision for Black-grass, indicating shared difficulties across all models with similar grassy species. VGG16+SVM, at 93.26% accuracy and a macro F1-score of 0.93, performed well on simpler classes but struggled more with challenging species like Black-grass and Loose Silky-bent, likely due to its shallower architecture and less efficient feature representation compared to the more advanced models. Thus, while EfficientNetB3+SVM and DenseNet201+SVM were the top performers, ResNet50+SVM remained competitive, and VGG16+SVM was less effective for closely resembling classes.

## TABLE 3. Performance comparison for four pretrained CNN models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy (%)** | **Precision** | **Recall** | **F1-score** |
| EfficientNetB3 + SVM | 95.89 | 0.96 | 0.95 | 0.95 |
| DenseNet201 + SVM | 95.89 | 0.96 | 0.95 | 0.95 |
| ResNet50 + SVM | 95.05 | 0.95 | 0.94 | 0.95 |
| VGG16 + SVM | 93.26 | 0.93 | 0.93 | 0.93 |

Figure 3 compares the classification accuracies of state-of-the-art models and the proposed CNN-SVM hybrids on the Plant Seedling dataset. Prior models like SVM by Jakovlev et al. (2021) [14] and DenseNet121 by Ofori & El-Gayar (2020) [15] achieved 92.4% and 92.21%, respectively, while Mostafa et al. (2022) [16] reported the lowest at 81.90%. In contrast, the proposed hybrid models performed better, with EfficientNetB3+SVM and DenseNet201+SVM reaching 95.89%. These results show that CNN-SVM hybrids outperform existing solutions by combining strong feature extraction with robust classification.

A graph of a graph with numbers and a bar

AI-generated content may be incorrect.

**Figure 3.** Comparison of classification accuracy of state-of-the-art models and the proposed approaches

# CONCLUSION

The experimental results demonstrate that hybrid models combining deep feature extraction with SVM classifiers achieve high accuracy in plant seedling classification, with EfficientNetB3+SVM and DenseNet201+SVM both reaching 95.89% accuracy. These models outperform ResNet50+SVM (95.05%) and VGG16+SVM (93.26%), highlighting the advantage of advanced CNN architectures in capturing discriminative features. However, challenges remain in distinguishing visually similar species like Black-grass and Common wheat, suggesting potential improvements through data augmentation or attention mechanisms. Overall, this study confirms the effectiveness of deep learning-based feature extraction for automated plant species identification, contributing to precision agriculture and ecological research.

# References

1. H. A. Elnemr, “Convolutional Neural Network Architecture for Plant Seedling Classification,” International Journal of Advanced Computer Science and Applications **10**(8), 319–325 (2019).
2. M. Padhiary, D. Saha, R. Kumar, L. N. Sethi, and A. Kumar, “Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicle for farm automation,” *Smart Agricultural Technology* **8**, 100483 (2024).
3. L. Wu and Y. Wen, “Weed/corn seedling recognition by support vector machine using texture features,” African Journal of Agricultural Research **4**(9), 840–846 (2009).
4. T. Rumpf, C. Römer, M. Weis, M. Sökefeld, R. Gerhards, and L. Plümer, ‘Sequential support vector machine classification for small-grain weed species discrimination with special regard to Cirsium arvense and Galium aparine’, Computers and Electronics in Agriculture**80**, 89–96 (2012).
5. T. M. Giselsson, H. S. Midtiby, and R. N. Jørgensen, “Seedling Discrimination with Shape Features Derived from a Distance Transform,” *Sensors* **13**(5), 5585–5602 (2013).
6. A. D. de Medeiros et al*.*, “Machine Learning for Seed Quality Classification: An Advanced Approach Using Merger Data from FT-NIR Spectroscopy and X-ray Imaging,” *Sensors* **20**(15), 4319 (2020).
7. W. Liu, J. Liu, J. Jiang, and Y. Li, “Comparison of partial least squares-discriminant analysis, support vector machines and deep neural networks for spectrometric classification of seed vigour in a broad range of tree species,” *J Near Infrared Spectrosc*, **29**(1), 33–41 (2021).
8. J. Liu, S. Yang, Y. Cheng, and Z. Song, “Plant Leaf Classification Based on Deep Learning,” in *2018 Chinese Automation Congress (CAC)*, 3165–3169 (2018).
9. J. A. Villaruz, J. A. A. Salido, D. M. Barrios, and R. L. Felizardo, “Philippine Indigenous Plant Seedlings Classification Using Deep Learning,” in *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 1–5 (2018).
10. T. R. Chavan and A. V Nandedkar, “AgroAVNET for crops and weeds classification: A step forward in automatic farming,” *Comput Electron Agric*, **154**, 361–372 (2018).
11. H. Alaeddine and M. Jihene, “Deep Batch-normalized eLU AlexNet For Plant Diseases Classification,” in *2021 18th International Multi-Conference on Systems, Signals & Devices (SSD)*, 17–22 (2021).
12. P. K. N, D. Hegde, R. Parimelazhagan, S. Sanjay, and S. P. Acharya, “Enhanced Medicinal Plant Leaf Classification via Transfer Learning: A VGG16-Driven Approach for Precise Feature Extraction,” in *2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, 1–6 (2024).
13. T. M. Giselsson, R. N. Jørgensen, P. K. Jensen, M. Dyrmann, and H. S. Midtiby, “A Public Image Database for Benchmark of Plant Seedling Classification Algorithms,” Computer Vision and Pattern Recognition in arXiv, 2017.
14. D. Jakovlev, I. Kamaletdinova, and G. Shevlyakov, “Feature-Based Plant Seedlings Classification,” in *Proceedings of International Scientific Conference on Telecommunications, Computing and Control*, T. and K. S. Voinov Nikita and Schreck, Ed., Singapore: Springer Singapore, 165–175 (2021).
15. M. Ofori and O. El-Gayar, “Towards Deep Learning for Weed Detection: Deep Convolutional Neural Network Architectures for Plant Seedling Classification,” *AMCIS 2020 Proceedings*, 2020.
16. S. Mostafa, D. Mondal, M. A. Beck, C. P. Bidinosti, C. J. Henry, and I. Stavness, ‘Leveraging Guided Backpropagation to Select Convolutional Neural Networks for Plant Classification’, *Front Artif Intell*, **5**, 2022.