*Article*

# Advanced Image Preprocessing and Integrated Modeling for UAV Plant Image Classification

Girma Tariku [1,*], Isabella Ghiglieno [2], Anna Simonetto [2], Fulvio Gentilin [3], Stefano Armiraglio [4], Gianni Gilioli [2] and Ivan Serina [1]

[1] Department of Information Engineering (DII), University of Brescia, Via Branze 38, 25123 Brescia, Italy; ivan.serina@unibs.it
[2] Agrofood Research Hub, Department of Civil, Environmental, Architectural Engineering and Mathematics, University of Brescia, Via Branze 43, 25123 Brescia, Italy; isabella.ghiglieno@unibs.it (I.G.); anna.simonetto@unibs.it (A.S.); gianni.gilioli@unibs.it (G.G.)
[3] RiD Lab, Department of Civil, Environmental, Architectural Engineering and Mathematics, University of Brescia, Via Branze 43, 25123 Brescia, Italy; fulvio.gentilin@unibs.it
[4] Museum of Natural Sciences, 25128 Brescia, Italy; sarmiraglio@comune.brescia.it
* Correspondence: g.tariku@unibs.it

**Abstract:** The automatic identification of plant species using unmanned aerial vehicles (UAVs) is a valuable tool for ecological research. However, challenges such as reduced spatial resolution due to high-altitude operations, image degradation from camera optics and sensor limitations, and information loss caused by terrain shadows hinder the accurate classification of plant species from UAV imagery. This study addresses these issues by proposing a novel image preprocessing pipeline and evaluating its impact on model performance. Our approach improves image quality through a multi-step pipeline that includes Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) for resolution enhancement, Contrast-Limited Adaptive Histogram Equalization (CLAHE) for contrast improvement, and white balance adjustments for accurate color representation. These preprocessing steps ensure high-quality input data, leading to better model performance. For feature extraction and classification, we employ a pre-trained VGG-16 deep convolutional neural network, followed by machine learning classifiers, including Support Vector Machine (SVM), random forest (RF), and Extreme Gradient Boosting (XGBoost). This hybrid approach, combining deep learning for feature extraction with machine learning for classification, not only enhances classification accuracy but also reduces computational resource requirements compared to relying solely on deep learning models. Notably, the VGG-16 + SVM model achieved an outstanding accuracy of 97.88% on a dataset preprocessed with ESRGAN and white balance adjustments, with a precision of 97.9%, a recall of 97.8%, and an F1 score of 0.978. Through a comprehensive comparative study, we demonstrate that the proposed framework, utilizing VGG-16 for feature extraction, SVM for classification, and preprocessed images with ESRGAN and white balance adjustments, achieves superior performance in plant species identification from UAV imagery.

**Keywords:** deep learning (VGG-16); image preprocessing; machine learning classifiers; plant species identification; unmanned aerial vehicles (UAVs)

## 1. Introduction

The rapid advancement of digital technologies is transforming environmental sciences, enabling more efficient and sustainable ecosystem management [1,2]. Plant image classification using techniques like deep learning offers a promising approach for rapid, precise, and cost-effective plant identification [3], biodiversity assessment, disease detection [4,5], and the evaluation of ecosystem services provisioning. This is particularly relevant in the context of unmanned aerial vehicles (UAVs), which have emerged as valuable tools for remote sensing and data acquisition, offering cost-effectiveness and flexibility [6].

UAVs are increasingly employed in various applications, including target detection for rescue and delivery operations [7], efficient package delivery [8], the cooperative transportation of goods [9], and the precise mapping of environmental features. However, their use in plant image classification faces several challenges. Limited flight duration [10] often necessitates high-altitude operations, leading to reduced spatial resolution [11,12] and difficulty in detecting fine-scale plant characteristics [13,14]. Additionally, camera optics and pixel size limitations can compromise image clarity and quality [15], while shadows cast by terrain features frequently obscure image details [16], hindering accurate analysis.

Despite the potential of UAV-based plant image classification, the existing research often lacks comprehensive evaluations of image preprocessing techniques and their combined impact on classification accuracy. Previous studies have highlighted the importance of image preprocessing for enhancing the quality of UAV plant images [17–20]. However, they often lack rigorous comparisons to determine the most effective techniques for specific applications. Additionally, there is a gap in research regarding the combination of different image preprocessing pipelines.

This study addresses these gaps by investigating the effectiveness of various image preprocessing techniques and their integration into different pipelines for plant species classification from RGB UAV images. We focus on a real-world case study, aiming to accurately detect seven plant species (*Agropyron repens*, *Ailanthus altissima*, *Arrhenatherum elatius*, *Artemisia verlotiorum*, *Populus nigra*, *Rubus caesius*, and *Ulmus minor*) from RGB UAV images.

Our research makes the following key contributions:

- Comprehensive Evaluation of Preprocessing Techniques and Pipelines: We thoroughly investigated the impact of various image preprocessing techniques on plant species classification accuracy from RGB UAV images. This includes identifying the most effective combinations of preprocessing techniques within different pipelines to optimize classification performance for seven target species.
- Hybrid Deep Learning and Machine Learning Approaches: We explore the performance of hybrid models combining a pre-trained VGG-16 for feature extraction with different classifiers (SVM, RF, XGBoost, and VGG-16 neural network layers), evaluating their effectiveness in plant species classification.

## 2. Related Work

Previous studies have highlighted the importance of image preprocessing techniques for enhancing the quality of UAV plant images. For example, SPAGRI-AI, a dataset designed to evaluate the effectiveness of super-resolution techniques like U-Net++, ESRGAN, and SwinIR on crop/weed detection using YOLOv5, was introduced [17]. Similarly, the effectiveness of ESRGAN in improving image resolution for accurate olive tree crown extraction with the U-2-Net model was demonstrated [18]. The role of Contrast-Limited Adaptive Histogram Equalization (CLAHE) in enhancing spatial details and minimizing color distortion in fused UAV images was highlighted by [19], addressing a crucial gap in research related to UAV-specific image fusion tasks. Additionally, improvements in the accuracy of Vegetation Index (VI) calculations under varying light conditions, making VI a more practical tool for small-scale agriculture, were demonstrated through white balance adjustments during UAV image preprocessing [20].

However, the existing studies often lack rigorous comparisons to definitively determine the most effective image preprocessing techniques for specific applications. Additionally, there is a gap in research regarding the combination of different image preprocessing pipelines. For example, Pandey and Jain developed an intelligent system for crop identification and classification using UAV images by employing a conjugated dense convolutional neural network [21]. While this study highlights the potential of advanced deep learning models in UAV-based plant classification, it does not extensively compare different preprocessing techniques. Similarly, Reedha et al. utilized a Transformer Neural Network for weed and crop classification of high-resolution UAV images, showcasing the advantages of modern neural architectures in remote sensing applications [22]. Nonetheless, these studies

primarily focus on the classification model itself rather than the preprocessing pipeline, leaving a gap in the exploration of how different preprocessing approaches might impact classification outcomes.

Methodologies for classifying plants from UAV RGB images typically focus on two main approaches: training large, labeled datasets from scratch using deep learning models, or training smaller datasets by employing pre-trained backbone models [23]. Utilizing pre-trained backbones offers advantages over training deep learning from scratch, including lower data requirements, faster model training, effective generalization to new tasks, and assistance in preventing overfitting issues [24]. Therefore, pre-trained backbone methods play a crucial role in enhancing the efficiency and robustness of plant classification models, particularly in scenarios where labeled data are scarce or computational resources are limited [25,26]. The combination of deep learning (DL) and machine learning (ML) models, based on a group learning approach, often yields higher precision and robustness than what can be achieved by individual models alone [27].

This study builds upon this existing research by performing the following:

- Conducting a comprehensive evaluation of various image preprocessing techniques and their combinations. This goes beyond previous studies that often focus on individual techniques or lack rigorous comparisons.
- Investigating the impact of preprocessing on the performance of hybrid deep learning and machine learning models. This explores the potential for synergistic performance gains by combining the strengths of both deep learning and traditional methods.

## 3. Materials and Methods

### 3.1. Objective and Overall Approach

This study aims to classify plant species using UAV RGB images of plant canopies from a dataset collected by [28]. The dataset consists of 1374 UAV images representing seven species: *Agropyron repens*, *Ailanthus altissima*, *Arrhenatherum elatius*, *Artemisia verlotiorum*, *Populus nigra*, *Rubus caesius*, and *Ulmus minor*. Addressing challenges such as low resolution, inconsistent lighting, and blurriness inherent in UAV imagery, we employ a systematic three-step approach consisting of image preprocessing, feature extraction, and classification, as illustrated in Figure 1. This approach addresses typical UAV-related challenges such as resolution, lighting inconsistencies, and noise, aiming to enhance the quality of ecological data collected through UAVs for research purposes.
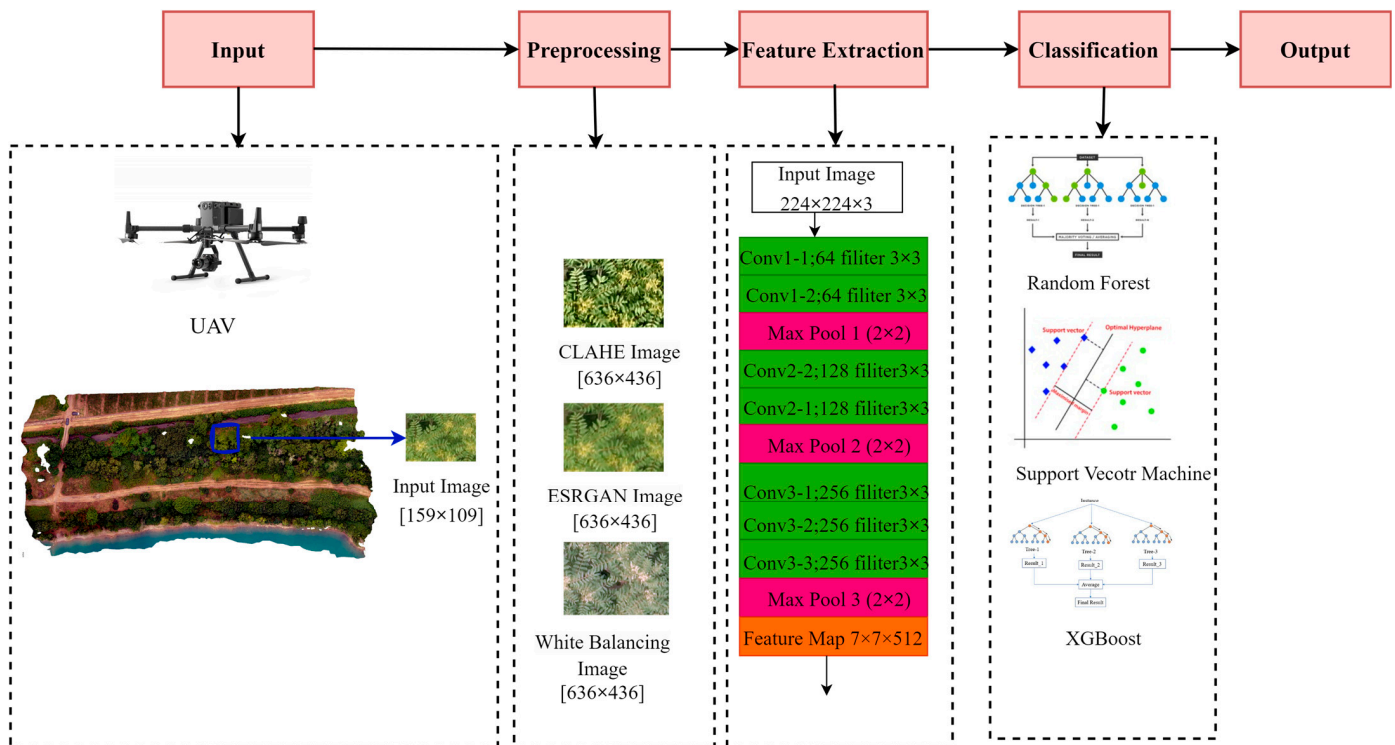
1. Image Preprocessing: We enhance image quality using a combination of techniques:
   - Enhanced Super-Resolution Generative Adversarial Network (ESRGAN): Improves image resolution.
   - Contrast-Limited Adaptive Histogram Equalization (CLAHE): Enhances contrast.
   - White Balancing: Corrects color imbalances.

   These techniques address the challenges posed by low-quality or inconsistently illuminated images, ensuring optimal data for the subsequent analysis.

2. Feature Extraction: We utilize a pre-trained VGG-16 model to extract relevant features from the preprocessed images. This deep convolutional neural network is known for its robustness and accuracy in image feature extraction. By using a well-established CNN like VGG-16, we ensure that the critical spatial features of plant species captured by UAVs are preserved and accurately represented. Before feeding the images into the VGG-16 model, we perform the necessary preprocessing steps such as resizing and normalizing to ensure consistent input dimensions and data distribution. The VGG-16 model then processes the images through convolutional layers, activation functions, batch normalization, and max pooling to identify patterns and reduce dimensions.

3. Classification: We evaluate the performance of four different classifiers:
   - Support Vector Machine (SVM)
   - Random forest (RF)
   - Extreme Gradient Boosting (XGBoost)

- VGG-16 Neural Network's Classification Layer

Each classifier is applied independently to the feature vectors extracted by the VGG-16 model without utilizing an ensemble framework. This comparison helps assess the robustness of different machine learning approaches in processing and interpreting UAV-collected ecological data. This allows for directly comparing their effectiveness in classifying plant species based on the extracted features. By independently assessing and comparing the performance of each classifier, we aim to identify the method offering the highest classification accuracy and gain a clear understanding of the strengths and limitations of each model in the context of plant species identification.



**Figure 1.** Overall working process of the proposed approach.

### 3.2. Image Preprocessing Techniques

This section details the specific image preprocessing techniques employed in this study to enhance the quality of the UAV plant images. These techniques address the challenges related to low resolution, contrast, and color imbalances, ultimately improving the accuracy of plant species identification.

3.2.1. Enhanced Super-Resolution Generative Adversarial Network (ESRGAN)

To improve the resolution of the images in our dataset, we utilized the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) [29]. ESRGAN is a deep learning algorithm specifically designed for image super-resolution, which involves creating high-resolution images from lower-resolution inputs. This is achieved through a generative adversarial network (GAN), where two neural networks compete as follows: one generates the higher-resolution image, and the other attempts to distinguish the generated image from a real high-resolution photo. This competitive process helps refine the system's ability to create images indistinguishable from real ones. By improving the resolution of the UAV images, ESRGAN enhances the visibility of intricate plant features that are essential for accurate species classification in ecological research.

To ensure the generated high-resolution images are both accurate and realistic, the ESRGAN model is trained using a combination of loss functions. Mean squared error (MSE)

measures the pixel-wise difference between the predicted high-resolution image and the ground-truth image, ensuring accuracy. On the other hand, perceptual loss evaluates the perceptual similarity between the model-generated image and the ground-truth image, guaranteeing the image appears natural to the human eye.

### 3.2.2. Contrast-Limited Adaptive Histogram Equalization (CLAHE)

To address potential low contrast issues in the images, we implemented Contrast-Limited Adaptive Histogram Equalization (CLAHE) [30] using OpenCV in Python. CLAHE enhances image contrast by redistributing pixel values based on the specific characteristics of each image histogram. This method extends the traditional histogram equalization, which can amplify noise in regions with few pixels. CLAHE ensures that contrast improvements are localized, effectively enhancing visual details while mitigating the over-amplification of noise. This method is particularly valuable in UAV-based ecological research, where lighting conditions can vary dramatically across a large landscape.

CLAHE works by first converting the image from RGB (Red, Green, Blue) to LAB color space. LAB separates the image into lightness (L), green/red (A), and blue/yellow (B) channels. Since contrast is directly controlled by lightness, CLAHE is applied specifically to the L channel. Unlike the traditional histogram equalization, which operates on the entire image, CLAHE works on small, predefined regions called tiles. This localized approach helps prevent noise amplification, a common issue with global equalization methods.

Furthermore, CLAHE offers two key parameters for controlling noise: clip limit and tile grid size. Clip limit sets a threshold for how much the histogram can be stretched in any particular tile, preventing excessive noise enhancement. Tile grid size defines the size of the tiles used for local equalization. Smaller tiles allow for more precise contrast adjustments in areas with significant contrast variations. In our study, we opted for a clip limit of 3 and a tile grid size of $(8 \times 8)$ to strike a balance between effective noise control and detailed contrast enhancement.

After processing the L channel with CLAHE, it is merged back with the original A and B channels. The image is then converted back to the RGB color space, and the CLAHE-enhanced image is displayed. This approach effectively improves the contrast of low-contrast images within our dataset while minimizing the introduction of unwanted noise artifacts.

### 3.2.3. White Balancing (WB)

To tackle potential color irregularities resulting from diverse lighting conditions during image capture, we integrated a precise white balancing [31] process into the ESRGAN image dataset. In UAV-acquired images, inconsistent lighting often introduces color biases that can affect species classification. We specifically employed the Gray-world algorithm, which assumes a well-balanced image should have an average pixel value across all colors (red, green, and blue) close to neutral gray (around RGB value 128).

This process involves converting the image to the LAB color space, where lightness (L) is separated from color information (A and B channels representing red/green and blue/yellow). By calculating the average values of the A and B channels, we identify the color cast. We then adjust these channels to remove the cast and normalize the image. This is achieved by subtracting 128 from their averages and scaling them based on the L channel values. Finally, a multiplier (like 1.2 in this case) is applied to the L channel for fine-tuning brightness. The resulting balanced image is converted back to the standard BGR color space, effectively correcting color casts and enhancing the overall image quality.
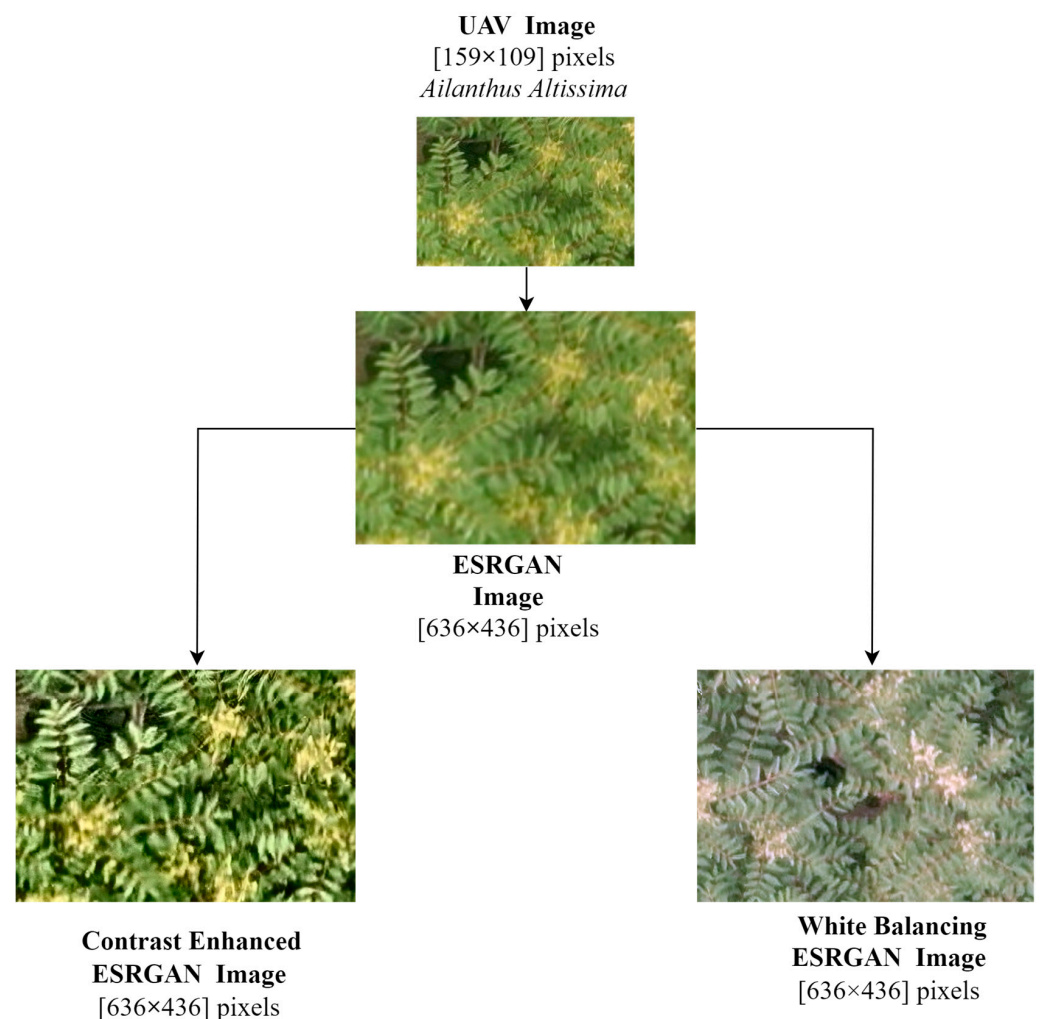
### 3.3. Dataset Creation

To investigate the influence of various image preprocessing techniques on image classification accuracy, we created four distinct datasets, each containing images with different preprocessing applied:

- Base Dataset: This dataset comprises the raw, unprocessed images directly captured by a drone, representing the initial state of the data.
- ESRGAN-Refined Dataset: This dataset incorporates the images enhanced using the ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) algorithm, which improves image resolution. We aimed to evaluate whether increased image resolution through ESRGAN would benefit classification performance.
- Contrast-Enhanced Dataset: Building upon the ESRGAN-refined images, this dataset applies additional contrast enhancement techniques, potentially aiding in the classification process by improving the visibility of subtle details. We investigated whether contrast enhancement, following ESRGAN refinement, could further improve classification accuracy.
- White-Balanced Dataset: This dataset includes the ESRGAN-refined images that have undergone white-balancing procedures, correcting for color casts caused by varying lighting conditions and ensuring a more consistent and natural color representation. We explored whether white balancing, alongside ESRGAN refinement, could enhance the model's ability to accurately classify image features.

This four-dataset approach allows us to systematically evaluate the impact of each image processing technique (ESRGAN refinement, contrast enhancement, and white balancing) on image classification performance. By comparing the classification accuracy achieved on each dataset (potentially visualized in Figure 2, sample image for one class), we can identify the most effective preprocessing techniques for our specific classification task.

**UAV Image**
[159×109] pixels
*Ailanthus Altissima*

**ESRGAN Image**
[636×436] pixels

**Contrast Enhanced ESRGAN Image**
[636×436] pixels

**White Balancing ESRGAN Image**
[636×436] pixels

**Figure 2.** Sample single Ailanthus altissima plant image in the four-image dataset.

*3.4. Feature Extraction*

This section describes the feature extraction process employed in this study, which leverages the VGG-16 architecture, a deep convolutional neural network (CNN) renowned for its efficacy in image classification tasks [32].

### 3.4.1. VGG-16 Architecture

We employ a pre-trained VGG-16 model [32] as the backbone of our feature extraction process. This deep convolutional neural network is known for its robustness and accuracy in extracting relevant features from images. The VGG-16 architecture comprises 16 layers, including 13 convolutional and 3 fully connected layers, structured with ($3 \times 3$) convolutional filters and interspersed with max-pooling layers. This uniform architecture facilitates the progressive extraction of intricate image features through hierarchical representation, as illustrated in Figure 3. Batch normalization applied after each convolutional layer enhances training stability and accelerates convergence rates.
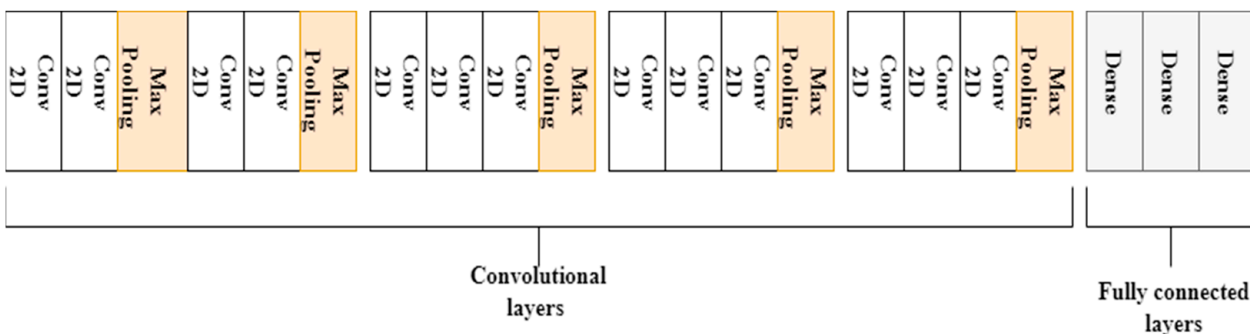


**Figure 3.** VGG-16 architecture map.

### 3.4.2. Feature Extraction Process

It is crucial to note that preprocessing steps such as resizing and normalization, which are essential to prepare images for input into the network, are not part of the VGG-16 architecture itself. These steps are performed prior to feeding the images into the model to ensure consistency in input dimensions and data distribution.

The VGG-16 model then processes the images through convolutional layers to identify patterns, activation functions to introduce non-linearity, batch normalization for training stability, and finally, max pooling to reduce dimensions.

Our methodology extracts the feature vector from the last fully connected layer before the softmax function is applied. This ensures that the feature extraction process remains unaffected by the classification output. The softmax layer, typically used during the training phase for multi-class classification, is not involved in the feature extraction process as described in our study. We maintain the integrity of the feature vectors used for the subsequent classification by omitting the softmax layer during feature extraction.

Ultimately, the output of the last convolutional layer is extracted, yielding a vector representation of the image features. These feature representations, originating from the convolutional layers of the VGG-16 architecture, serve as input data for the classification process.

*3.5. Classification Models*

### 3.5.1. Random Forest

Random forests [33] are ensemble learning models that combine multiple decision trees to improve system performance. Each decision tree in a random forest is trained on a random subset of the training data and makes independent predictions. The final prediction of the random forest is obtained by aggregating the predictions of all the trees, often using the majority rule criteria.

For this study, we configured the random forest model with the following parameters:

- Unpruned Trees: Trees were allowed to grow without constraints on the maximum number of levels, promoting flexibility in capturing complex decision boundaries.
- Minimum Split and Leaf Nodes: The minimum number of data points required in a node before it can be split was set to 2, and the minimum allowed data points in a terminal leaf node was set to 1. These parameters prevent overfitting by ensuring a minimal level of information in each split and leaf.
- Gini Impurity: The Gini index was employed as the criterion for selecting the best splitting feature at each node. The Gini index measures the level of impurity within a node perfectly homogeneous node (all the data points belong to one class) has a Gini index of 0. The model greedily selects the feature that best separates the data into distinct classes, minimizing the overall Gini impurity.
- Number of Trees: We opted for 50 trees in the random forest ensemble. While increasing the number of trees can further enhance performance, we found that 50 trees provided a good balance between accuracy and computational efficiency for our dataset size.

3.5.2. Support Vector Machine (SVM)

SVM [34] is a supervised learning algorithm commonly employed when data are not linearly separable, meaning that classes cannot be distinguished by a straight line or hyperplane in the feature space.

For this SVM implementation, we utilized the radial basis function (RBF) kernel, enabling the algorithm to capture non-linear relationships in the data. The regularization parameter (C) determines the trade-off between a smooth decision boundary and the accurate classification of training points. A higher C value prioritizes correct classification.

For this study, we configured the SVM model with the following key parameters:

- Radial Basis Function (RBF) Kernel: We employed the RBF kernel, which is a popular choice for non-linear SVM applications. The RBF kernel allows the SVM to capture complex, non-linear relationships between the features in the data, making it suitable for effectively separating the image classes.
- Regularization Parameter (C): This parameter controls the trade-off between achieving a smooth decision boundary and accurately classifying all training data points. A higher C value prioritizes the strict classification of training points, which can lead to a more complex decision boundary and potentially increased risk of overfitting. In our case, we carefully tuned the C parameter to a value of 14.5, striking a balance between these competing factors.

3.5.3. Extreme Gradient Boosting (XGBoost)

XGBoost [35] (Extreme Gradient Boosting) is a supervised learning algorithm for tree boosting, renowned for its exceptional prediction performance and efficient computation, making it well suited for classification tasks. XGBoost aims to accurately predict a target variable by combining the predictions of a set of simpler and weaker models, known as decision trees. It minimizes a regularized objective function comprising a convex loss function based on the difference between the predicted and target outputs, and a penalty term for model complexity (i.e., the classification tree functions).

For this study, we configured XGBoost with the following parameters:

- Learning Rate (0.3): This parameter controls the step size taken in each boosting iteration. A smaller learning rate like 0.3 helps prevent overfitting by making smaller adjustments to the model with each step.
- Number of Booster Rounds (100): This parameter determines the number of trees included in the final ensemble. We opted for 100 trees, striking a balance between achieving high accuracy and maintaining computational efficiency.
- Max Tree Depth (6): This parameter limits the maximum depth of each individual tree within the ensemble. Limiting the depth helps control model complexity and reduces the risk of overfitting.

- Subsample (1): This parameter specifies the proportion of training data used to fit each tree. In our case, we used the entire dataset (subsample = 1) to maximize the information available to each tree during training.

### 3.5.4. VGG-16 Neural Network Classifier

Our VGG-16 classifier leverages the pre-trained VGG16 model, known for its performance on the ImageNet dataset. We strategically loaded VGG16 with the "include_top" parameter set to False, excluding the model's final classification layers. This preserves only the powerful convolutional base for efficient feature extraction, as described in Section 3.4.2. To further optimize training, we froze the pre-trained weights in the base model by setting "layer.trainable = False" for each layer. This ensures the model leverages the learned features from VGG16 without modification, significantly reducing training time and computational resources.

Following feature extraction, a custom classifier is built on top of the VGG16 base. A Flatten layer transforms the multi-dimensional feature output into a single vector. This vector is then fed through a dense layer with 1024 neurons and a ReLU activation for introducing non-linearity and enabling complex pattern learning. To prevent overfitting, a Dropout layer randomly sets 20% of the neurons to zero during training. Finally, the classifier concludes with a dense layer containing the number of output classes (7 in our case) and a softmax activation function, which outputs class probabilities for prediction.

### 3.6. Performance Evaluation

The performance of each classification model was assessed by splitting the datasets with expanded features into a training set (comprising 80% of the initial data) and a test set (comprising 20% of the data). For each dataset and each classification model, the following indexes were computed:

$$Acuraccy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

$$Precision = \frac{T_p}{T_p + F_p}$$

$$Recall = \frac{T_p}{T_p + F_n}$$

$$F1\ score = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

where $T_p$ represents True Positive, $T_n$ represents True Negative, $F_p$ represents False Positive, and $F_n$ represents False Negative.

## 4. Result

This study investigates the effectiveness of four machine learning models for classifying images of plant species, focusing on the impact of image preprocessing techniques. We evaluated the performance of three commonly used classifiers Support Vector Machine (SVM), random forest (RF), and XGBoost along with a custom VGG-16 classifier built on top of the pre-trained VGG-16 convolutional base. To assess the impact of data preprocessing, we evaluated the models across four different dataset types, as detailed in Table 1. These datasets included the base dataset, as well as versions enhanced with ESRGAN, CLAHE, and white balancing techniques.

**Table 1.** Accuracy, F1 score, recall, and precision values for image dataset without preprocessing, dataset preprocessed by ESRGAN, dataset preprocessed by ESRGAN and CLAHE, and dataset preprocessed by ESRGAN and white balancing.

| Input Image Dataset Type | Hybrid Classification Model | Accuracy | F1 Score | Recall | Precision |
|---|---|---|---|---|---|
| Image dataset(base) | VGG 16 + RF | 87.59 | 0.874 | 0.875 | 0.877 |
| | VGG 16 + SVM | 95.02 | 0.961 | 0.960 | 0.962 |
| | VGG 16 + XGBoost | 86.20 | 0.862 | 0.860 | 0.870 |
| | Full VGG 16 model | 85.03 | 0.853 | 0.851 | 0.845 |
| Base preprocessed + ESRGAN | VGG 16 + RF | 89.72 | 0.895 | 0.896 | 0.891 |
| | VGG 16 + SVM | 96.71 | 0.960 | 0.965 | 0.961 |
| | VGG 16 + XGBoost | 91.60 | 0.912 | 0.911 | 0.922 |
| | Full VGG 16 model | 89.78 | 0.897 | 0.896 | 0.899 |
| Base preprocessed + ESRGAN and CLAHE | VGG 16 + RF | 93.79 | 0.939 | 0.938 | 0.941 |
| | VGG 16 + SVM | 97.44 | 0.974 | 0.975 | 0.975 |
| | VGG 16 + XGBoost | 92.91 | 0.921 | 0.924 | 0.920 |
| | Full VGG 16 model | 93.79 | 0.939 | 0.938 | 0.939 |
| Base preprocessed + ESRGAN and WB | VGG 16 + RF | 95.25 | 0.953 | 0.953 | 0.954 |
| | VGG 16 + SVM | 97.88 | 0.978 | 0.978 | 0.979 |
| | VGG 16 + XGBoost | 94.52 | 0.946 | 0.946 | 0.948 |
| | Full VGG 16 model | 94.16 | 0.943 | 0.947 | 0.943 |

*4.1. Performance on Base Image Dataset*

On the base image dataset, SVM emerged as the top performer with an accuracy of 95.02%. This performance was accompanied by robust F1 score, recall, and precision metrics, demonstrating SVM's suitability for handling diverse image datasets. Random forest followed closely behind with 87% accuracy, showcasing competitive performance but with slightly lower metrics compared to SVM. XGBoost achieved an accuracy of 86.2%, highlighting its effectiveness but performing lower than SVM and RF in this particular dataset configuration.

The custom VGG-16 classifier achieved an accuracy of 89.71% on the base image dataset. This performance is encouraging, demonstrating the potential of a fully trained VGG-16 model for this classification task.

A comparison of the full VGG-16 model to other hybrid classification models (VGG-16 + SVM, VGG-16 + RF, and VGG-16 + XGBoost) revealed a consistent pattern of underperformance by the full VGG-16 model. For the base image dataset, it achieved an accuracy of 85.03%, while the hybrid models achieved significantly higher accuracies: VGG-16 + SVM (95.02%), VGG-16 + RF (87.59%), and VGG-16 + XGBoost (86.20%).

*4.2. Performance on Preprocessed Datasets*

4.2.1. Base Preprocessed ESRGAN Dataset

Moving to the base preprocessed ESRGAN dataset, SVM continued to lead with an impressive accuracy of 96.7%. This dataset preprocessing approach enhanced SVM's ability to discern patterns in images, resulting in improved classification accuracy and balanced performance across other metrics. Random forest also showed significant improvement in accuracy to 89.7%, underscoring the benefits of preprocessing in boosting classifier performance. XGBoost maintained its competitive stance with an accuracy of 91.6%, demonstrating consistent performance gains with preprocessing but slightly below SVM in accuracy metrics.

The custom VGG-16 classifier achieved an accuracy of 92.34% on the ESRGAN preprocessed dataset. This improvement over the base dataset suggests that the custom VGG-16 model benefits from the enhanced image quality provided by ESRGAN.

With the ESRGAN preprocessed dataset, the full VGG-16 model's accuracy improved to 89.78%. However, it still lagged behind the hybrid models, with the VGG-16 + SVM achieving 96.71%, the VGG-16 + RF achieving 89.72%, and the VGG-16 + XGBoost reaching 91.60%.

4.2.2. Base Preprocessed ESRGAN and CLAHE Dataset

In the base preprocessed ESRGAN and CLAHE dataset, SVM again excelled with an accuracy of 97.4%, marking a notable increase compared to previous configurations. This dataset combination further enhanced SVM's ability to handle nuanced image features, reflected in its high precision, recall, and F1 score. Random forest and XGBoost also showed improvements with accuracies of 93.7% and 92.9%, respectively, indicating their adaptability to enhanced image preprocessing techniques like CLAHE.

The custom VGG-16 classifier achieved an accuracy of 94.87% on the ESRGAN and CLAHE preprocessed dataset. This further improvement suggests that the custom VGG-16 model benefits from the combined enhancements of ESRGAN and CLAHE preprocessing.

The full VGG-16 model recorded an accuracy of 93.79%, which, while much improved, still fell short of the VGG-16 + SVM model's 97.44% and was equal to the VGG-16 + RF model's accuracy of 93.79%. The VGG-16 + XGBoost model also performed better with an accuracy of 92.91%.

4.2.3. Base Preprocessed ESRGAN and White Balancing (WB) Dataset

For the base preprocessed ESRGAN and WB dataset, SVM reached the highest accuracy of 97.88%. This result highlights the significant impact of combining ESRGAN with white balancing to enhance image clarity and standardization. Random forest and XGBoost also performed well, with accuracies of 95.25% and 94.52%, respectively. The full VGG-16 model achieved its highest accuracy of 94.16% in this scenario but remained lower than the VGG-16 + SVM model (97.88%).

The custom VGG-16 classifier achieved an accuracy of 95.72% on the ESRGAN and WB preprocessed dataset. This performance suggests that the custom VGG-16 model is capable of leveraging the benefits of various preprocessing techniques to achieve high accuracy.

*4.3. Comparative Analysis*

Overall, our comparative analysis highlights SVM as consistently outperforming random forest, XGBoost, and VGG-16 classifiers across different dataset types and preprocessing methods in drone-captured imagery classification tasks. Among the image preprocessing techniques evaluated in our study, base preprocessed ESRGAN and white balancing (WB) emerge as particularly effective in enhancing classification accuracy across all the machine learning classifiers.

The effectiveness of the preprocessing techniques, particularly ESRGAN and WB, is evident as they significantly enhance classification accuracy compared to unenhanced datasets. This substantial improvement underscores the importance of ESRGAN in enhancing image clarity and feature extraction, while WB further standardizes color and contrast, making images more suitable for accurate classification by SVM.

The test results demonstrate that the combination of the VGG-16 DL feature extractor and an SVM classifier performed exceptionally well on the dataset enhanced by ESRGAN and white-balancing preprocessing methods. This combination exhibits the highest precision (97.8%) among all the models and dataset combinations, with an F1 score of 0.974, a recall of 0.978, and a precision of 0.979.

As shown in Figure 4, the random forest classifier for the base dataset enhanced by ESRGAN and white balance obtains the second position with accuracy values of 95.25%, a precision of 95.4%, a recall of 95.3%, and an F1 score of 95.3%. The accuracy, precision, recall, and F1 score of the VGG-16 DL feature extractor in conjunction with an XGBoost for the base dataset enhanced by ESRGAN and white balance are 94.5%, 94.8%, 94.6%, and 94.6%, respectively.
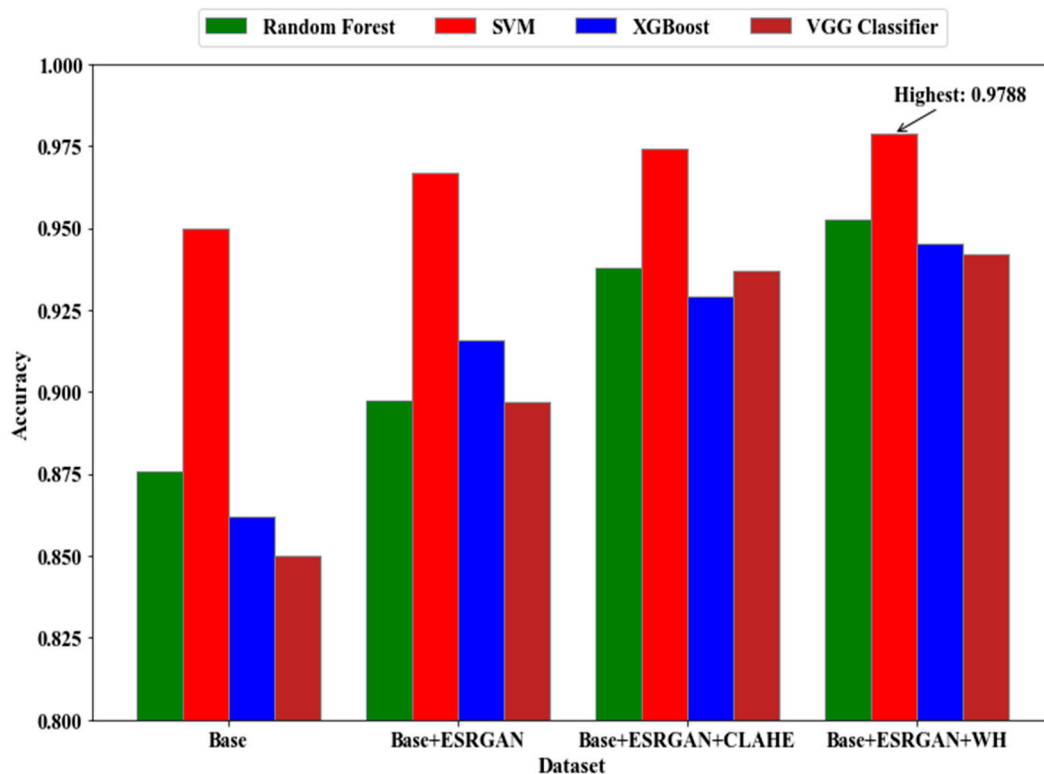
**Figure 4.** Accuracy of the four plant datasets using a combined VGG+16 feature extractor and machine learning classifiers SVM, random forest, XGBoost, and VGG-16 classifier.

*4.4. Confusion Matrices*

Figure 5 presents confusion matrices illustrating the classification results of four models (SVM, random forest, XGBoost, and VGG-16 classifier) across four different dataset enhancements: (1) base dataset, (2) Base + ESRGAN, (3) Base + ESRGAN + CLAHE, and (4) Base + ESRGAN + WB.

Analysis of Misclassifications:

- Base + ESRGAN + WB: The SVM classifier exhibits the lowest misclassification rate with only 6 incorrect classifications, followed by random forest (11) and XGBoost (15). The full VGG-16 model misclassifies 16 images. This dataset enhancement, combining ESRGAN and white balancing, consistently demonstrates the lowest misclassification rates across all models.

- Base + ESRGAN + CLAHE: The SVM classifier shows 7 misclassifications, while random forest and XGBoost misclassify 35 and 22 images, respectively. The VGG-16 classifier misclassifies 15 images. This scenario highlights the potential impact of CLAHE on classifier decision boundaries, leading to a higher number of misclassifications compared to the ESRGAN and white balancing combination.

- Base + ESRGAN: The SVM classifier misclassifies 9 images, while random forest and XGBoost misclassify 28 and 23 images, respectively. The VGG-16 classifier misclassifies 28 images. This suggests that while ESRGAN improves feature extraction, additional preprocessing steps like white balancing contribute to more consistent classification accuracy.

- Base Dataset: The SVM classifier has 17 misclassifications, while random forest and XGBoost misclassify 27 and 22 images, respectively. This highlights the significant improvement achieved through dataset enhancement techniques like ESRGAN and CLAHE in reducing misclassifications and enhancing the overall performance of machine learning models.
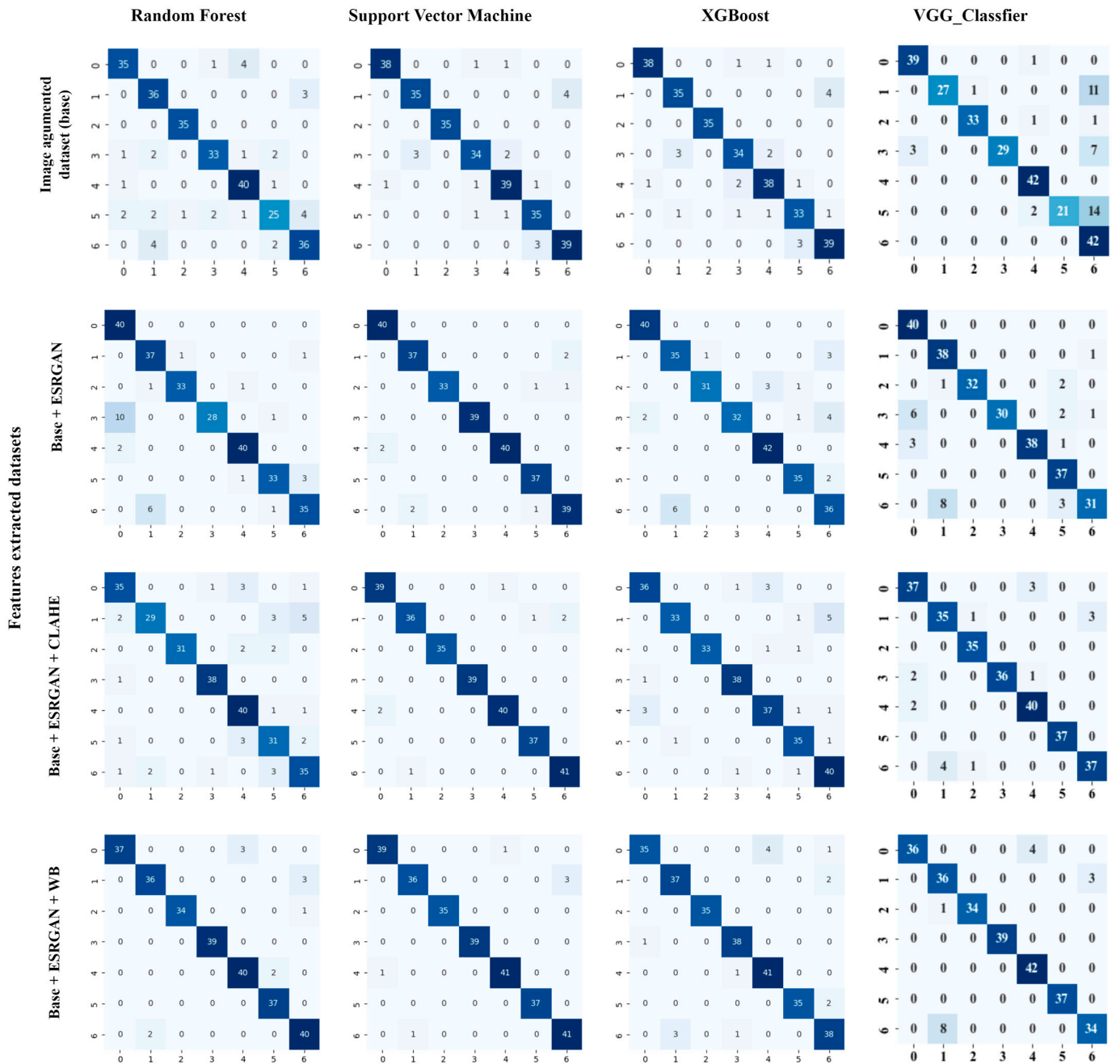
**Figure 5.** Confusion matrices comparing the performance of three machine learning models Support Vector Machine (SVM), random forest (RF), XGBoost, and VGG16 classifier across four datasets: (1) base dataset, (2) Base + ESRGAN, (3) Base + ESRGAN + CLAHE, and (4) Base + ESRGAN + WB. Each row corresponds to a different dataset, and each column represents the classification results from one of the models. The matrices illustrate the number of correct and incorrect classifications for various classes, with darker blue shades indicating higher values, highlighting the models' performance across different data preprocessing techniques.

To validate our image classification method's generalizability, we applied it to the UC Merced Land Use Dataset [36], a publicly available dataset commonly used for land use classification. This dataset comprises 21 classes, each represented by 100 256 × 256-pixel images manually extracted from higher-resolution USGS National Map Urban Area Imagery (1 foot/pixel). We selected three classes for this evaluation. The results demonstrate the effectiveness of our preprocessing steps and hybrid classification approach on diverse image datasets.

Table 2 shows that the base VGG16 + SVM model achieved 95.02% accuracy and an F1 score of 0.947 when applied directly to the dataset. However, incorporating advanced preprocessing techniques (ESRGAN and CLAHE) increased the accuracy to 98.33% and the F1 score to 0.983, highlighting the significant impact of image enhancement. A separate configuration using ESRGAN with white balance (WB) achieved 97.67% accuracy. These findings underscore the importance of preprocessing for optimizing classification performance across different classes within the UC Merced dataset.

**Table 2.** Accuracy, F1 score, recall, and precision values for image dataset without preprocessing, dataset preprocessed by ESRGAN, dataset preprocessed by ESRGAN and CLAHE, and dataset preprocessed by ESRGAN and white balancing.

| Input Image Dataset Type | Hybrid Classification Model | Accuracy | F1 Score | Recall | Precision |
|---|---|---|---|---|---|
| Image dataset(base) | VGG 16 + SVM | 95.02 | 0.947 | 0.956 | 0.950 |
| Base preprocessed + ESRGAN | VGG 16 + SVM | 96.66 | 0.966 | 0.968 | 0.966 |
| Base preprocessed + ESRGAN and CLAHE | VGG 16 + SVM | 98.33 | 0.983 | 0.984 | 0.983 |
| Base preprocessed + ESRGAN and WB | VGG 16 + SVM | 97.67 | 0.975 | 0.979 | 0.976 |

## 5. Discussion

This study investigated the impact of various image preprocessing techniques on the accuracy of plant species classification using UAV-captured imagery. While previous research has demonstrated the effectiveness of individual preprocessing techniques for improving UAV image quality [17,19,20], our research specifically focuses on leveraging these techniques in a cohesive pipeline to enhance the specific task of plant species classification. This is particularly relevant given the challenges posed by UAV image acquisition, such as reduced spatial resolution and image clarity issues [12,14,16].

Our findings highlight the critical role of preprocessing in enhancing image quality, ultimately improving the performance of the chosen classifiers. We observed that techniques like ESRGAN for resolution enhancement [17], CLAHE for contrast improvement [19], and white balancing for color correction [20] significantly improved classification outcomes. Among the tested classifiers, SVM consistently outperformed the others, achieving the highest accuracy, particularly with datasets that underwent preprocessing. Notably, the combination of ESRGAN and white balancing resulted in SVM achieving an impressive 97.88% accuracy, demonstrating its effectiveness with high-quality, preprocessed images. This aligns with previous research demonstrating the effectiveness of super-resolution techniques like ESRGAN in improving image quality for various applications, including crop/weed detection [20].

Furthermore, our comparison of the full VGG16 model with hybrid models consistently revealed superior performance by the hybrid models, particularly VGG16 combined with SVM. This hybrid approach consistently outperformed the full VGG16 model across all the datasets in terms of accuracy, F1 score, recall, and precision, highlighting the benefits of combining VGG16 with other classifiers for enhanced performance. This finding is consistent with previous studies that have explored the benefits of combining deep learning models with traditional machine learning classifiers for improved performance [27].

Our hybrid model, employing VGG-16 for feature extraction and SVM, RF, or XGBoost for classification, utilizes a dataset enhanced by ESRGAN and CLAHE preprocessing. To evaluate its performance, we also compared the hybrid models to well-known deep learning models such as ResNet50 [37], DenseNet121 [38], InceptionV3 [39], EfficientNet-B0 [40], and MobileNetV2 [41]. This comparison was conducted using the base dataset without preprocessing, and all the deep learning models were trained for 15 epochs. The comparison focused on training time, computational resources, and accuracy, as shown in Table 3.

**Table 3.** Performance comparison of different models.

| Model | Training Time (min) | Inference Time (s) | Testing Time (s) | Memory Usage (GB) | Accuracy (%) | F1 Score |
|---|---|---|---|---|---|---|
| VGG-16 (Full) | 15 | 27.01 | 28,61 | 4 | 89.71 | 0.85 |
| ResNet50 | 25 | 16.34 | 18.04 | 6 | 60.2 | 0.59 |
| InceptionV3 | 18 | 11.97 | 14.58 | 5 | 91.5 | 0.91 |
| DenseNet121 | 10 | 10.20 | 11.83 | 3 | 95.3 | 0.95 |
| EfficientNet-B0 | 12 | 4.40 | 5.98 | 2.5 | 94.1 | 0.94 |
| MobileNetV2 | 8 | 6.42 | 7.99 | 2 | 92.8 | 0.93 |
| VGG-16 Feature Extraction + SVM | <2 | 1.74 | 2.88 | 1 | 97.4 | 0.97 |
| VGG-16 Feature Extraction + RF | <2 | 0.60 | 1.74 | 1 | 93.7 | 0.93 |
| VGG-16 Feature Extraction + XGBoost | <2 | 0.30 | 1.47 | 1 | 92.7 | 0.92 |

The hybrid model (VGG-16 feature extraction with SVM) demonstrates a compelling advantage for deployment. It achieved 97.4% accuracy with an inference time of 1.74 s and a testing time of 2.88 s using only 1 GB of memory (Table 3). This significant improvement in memory efficiency and processing speed is crucial for real-time applications and deployment on resource-constrained embedded systems, often essential for in situ agricultural monitoring. The reduced memory footprint and rapid inference and testing times allow for the deployment of smaller, more energy-efficient devices, decreasing costs and increasing the practicality of continuous monitoring. The difference between inference and testing time is noteworthy. Inference time, as reported here, reflects the optimized performance in a production setting, focusing solely on the model's prediction speed once deployed. On the other hand, testing time includes additional overhead such as data loading and metric calculations performed during the evaluation phase. This distinction highlights the practical efficiency gains achieved by our hybrid model in a real-world deployment scenario. This efficiency is particularly notable when compared to models like DenseNet121, which, while achieving a slightly higher accuracy (95.3%), required significantly more training time (10 min), a longer inference time (10.20 s), and a longer testing time (11.83 s), and more memory (3 GB). ResNet50, InceptionV3, EfficientNet-B0, and MobileNetV2 further highlight this trade-off between accuracy, training, inference, testing, and memory usage. The hybrid model's superior efficiency in both memory and speed makes it particularly well suited for real-time applications and resource-constrained environments where rapid processing and minimal memory usage are paramount.

Among the preprocessing techniques evaluated, ESRGAN and white balancing emerged as particularly effective. Their combination significantly enhanced image clarity and detail, facilitating better feature extraction by the VGG-16 model and the subsequent classification by SVM. This combination resulted in the highest overall classification accuracy and the lowest misclassification rates, highlighting its potential for optimizing UAV-based plant species identification tasks. This finding builds upon previous research that demonstrated the effectiveness of white balance adjustments in improving the accuracy of Vegetation Index (VI) calculations under varying light conditions [20].

Our research contributes to the field by providing a comprehensive evaluation of various preprocessing techniques and their integration into different pipelines for plant species classification from RGB UAV images. This study also investigates the performance of hybrid models combining a pre-trained VGG-16 for feature extraction with different classifiers (SVM, RF, XGBoost, and VGG-16 neural network layers), exploring their effectiveness in plant species classification.

While our study presents significant findings, it is important to acknowledge its limitations. The performance of preprocessing techniques such as ESRGAN and CLAHE was evaluated on a specific dataset, which may limit the generalizability of the results. Future work should consider applying these techniques to diverse datasets to further validate their effectiveness. Additionally, exploring other advanced image enhancement methods and integrating them into the classification pipeline could provide deeper insights into optimizing plant species identification from UAV imagery.

## 6. Conclusions

This study addresses the challenges associated with classifying plant species from UAV imagery, particularly those arising from reduced spatial resolution, compromised image clarity, and information loss due to shadows. By implementing a comprehensive multi-step preprocessing pipeline, including techniques such as Enhanced Super-resolution Generative Adversarial Networks (ESRGAN), Contrast-Limited Adaptive Histogram Equalization (CLAHE), and white balance adjustments, we significantly improved image quality, enabling more accurate feature extraction and classification. These advanced preprocessing steps have been shown to substantially enhance the model's performance, contributing to better generalizability across different datasets.

Our methodology leveraged a pre-trained VGG-16 deep convolutional neural network for feature extraction, followed by classification using machine learning models such as Support Vector Machine (SVM), random forest (RF), Extreme Gradient Boosting (XGBoost), and the VGG-16 neural network layer classifier. Among these, the combination of VGG-16 with SVM, using preprocessed images enhanced by ESRGAN and white balance, yielded the highest classification accuracy of 98.33%. This highlights the significant impact of our hybrid approach combining deep learning for feature extraction with machine learning for classification on improving classification performance. Furthermore, our work contributes to the growing body of literature on plant image classification, showcasing the effective collaboration of image enhancement techniques and machine learning models.

Reflecting on these findings, we encourage future researchers to explore various low-resolution datasets and further investigate the effects of image enhancement on classification performance across different plant species and environmental conditions. Future work should also focus on expanding the scope of the model to handle more complex scenarios and diverse plant species, ensuring scalability for broader applications in precision agriculture. In summary, our work provides a practical and effective solution to the challenges of plant identification using UAVs in precision agriculture, demonstrating the value of integrating preprocessing and hybrid classification methodologies.

**Author Contributions:** Conceptualization, G.T., G.G. and A.S.; methodology, G.T.; software, G.T. and F.G.; validation, I.G., G.G., I.S., A.S. and S.A.; formal analysis, G.T., A.S., G.G. and I.S.; investigation, G.T.; resources, G.T. and F.G.; data curation, G.T. and F.G.; writing—original draft preparation, G.T.; writing—review and editing, G.T., A.S., G.G. and I.S.; visualization, A.S. and G.G.; supervision, G.G.; project administration, G.G. All authors have read and agreed to the published version of the manuscript.

**Data Availability Statement:** The dataset presented in this study is available at the link https://doi.org/10.5281/zenodo.8297802. (accessed on 5 September 2023).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Qin, T.; Wang, L.; Zhou, Y.; Guo, L.; Jiang, G.; Zhang, L. Digital technology-and-services-driven sustainable transformation of agriculture: Cases of China and the EU. *Agriculture* **2022**, *12*, 297. [CrossRef]
2. Shahi, T.B.; Dahal, S.; Sitaula, C.; Neupane, A.; Guo, W. Deep Learning-Based Weed Detection Using UAV Images: A Comparative Study. *Drones* **2023**, *7*, 624. [CrossRef]
3. Lee, S.H.; Chan, C.S.; Mayo, S.J.; Remagnino, P. How deep learning extracts and learns leaf features for plant classification. *Pattern Recognit.* **2017**, *71*, 1–13. Available online: https://www.sciencedirect.com/science/article/pii/S003132031730198X (accessed on 7 October 2024). [CrossRef]

4. Geetharamani, G.; Pandian, A. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Comput. Electr. Eng.* **2019**, *76*, 323–338. [CrossRef]

5. Saleem, M.H.; Potgieter, J.; Arif, K.M. Plant Disease Detection and Classification by Deep Learning. *Plants* **2019**, *8*, 468. [CrossRef]

6. Zhang, Z.; Zhu, L. A Review on Unmanned Aerial Vehicle Remote Sensing: Platforms, Sensors, Data Processing Methods, and Applications. *Drones* **2023**, *7*, 398. [CrossRef]

7. Nguyen, V.S.; Jung, J.; Jung, S.; Joe, S.; Kim, B. Deployable hook retrieval system for UAV rescue and delivery. *IEEE Access* **2021**, *9*, 74632–74645. [CrossRef]

8. Li, X.; Tupayachi, J.; Sharmin, A.; Martinez Ferguson, M. Drone-Aided Delivery Methods, Challenge, and the Future: A Methodological Review. *Drones* **2023**, *7*, 191. [CrossRef]

9. Loianno, G.; Kumar, V. Cooperative Transportation Using Small Quadrotors Using Monocular Vision and Inertial Sensing. *IEEE Robot. Autom. Lett.* **2017**, *3*, 680–687. [CrossRef]

10. Mohsan, S.A.H.; Othman, N.Q.H.; Li, Y.; Alsharif, M.H.; Khan, M.A. Unmanned Aerial Vehicles (UAVs): Practical Aspects, Applications, Open Challenges, Security Issues, and Future Trends. *Intell. Serv. Robot.* **2023**, *16*, 109–137. [CrossRef]

11. Jiang, Y.; Wei, Z.; Hu, G. Detection of Tea Leaf Blight in UAV Remote Sensing Images by Integrating Super-Resolution and Detection Networks. *Environ. Monit. Assess.* **2024**, *196*, 1–27. [CrossRef] [PubMed]

12. Seifert, E.; Seifert, S.; Vogt, H.; Drew, D.; van Aardt, J.; Kunneke, A.; Seifert, T. Influence of Drone Altitude, Image Overlap, and Optical Sensor Resolution on Multi-View Reconstruction of Forest Images. *Remote Sens.* **2019**, *11*, 1252. [CrossRef]

13. Bongomin, O.; Lamo, J.; Guina, J.M.; Okello, C.; Ocen, G.G.; Obura, M.; Alibu, S.; Owino, C.A.; Akwero, A.; Ojok, S. UAV Image Acquisition and Processing for High-Throughput Phenotyping in Agricultural Research and Breeding Programs. *Plant Phenome J.* **2024**, *7*, e20096. [CrossRef]

14. Chen, J.; Chen, Z.; Huang, R.; You, H.; Han, X.; Yue, T.; Zhou, G. The Effects of Spatial Resolution and Resampling on the Classification Accuracy of Wetland Vegetation Species and Ground Objects: A Study Based on High Spatial Resolution UAV Images. *Drones* **2023**, *7*, 61. [CrossRef]

15. Šulc, M.; Matas, J. Fine-Grained Recognition of Plants from Images. *Plant Methods* **2017**, *13*, 1–14. [CrossRef] [PubMed]

16. Zali, S.-A.; Mat-Desa, S.; Che-Embi, Z.; Mohd-Isa, W.-N. Post-Processing for Shadow Detection in Drone-Acquired Images Using U-Net. *Future Internet* **2022**, *14*, 231. [CrossRef]

17. Jonak, M.; Mucha, J.; Jezek, S.; Kovac, D.; Cziria, K. SPAGRI-AI: Smart Precision Agriculture Dataset of Aerial Images at Different Heights for Crop and Weed Detection Using Super-Resolution. *Agric. Syst.* **2024**, *216*, 103876. [CrossRef]

18. Ye, Z.; Wei, J.; Lin, Y.; Guo, Q.; Zhang, J.; Zhang, H.; Deng, H.; Yang, K. Extraction of olive crown based on UAV visible images and the U2-Net deep learning model. *Remote Sens.* **2022**, *14*, 1523. [CrossRef]

19. Modak, S.; Heil, J.; Stein, A. Pan sharpening low-altitude multispectral images of potato plants using a generative adversarial network. *Remote Sens.* **2024**, *16*, 874. [CrossRef]

20. Kusnandar, T.; Surendra, K. Camera-Based Vegetation Index from Unmanned Aerial Vehicles. In Proceedings of the 6th International Conference on Sustainable Information Engineering and Technology, Malang, Indonesia, 13–14 September 2021; pp. 173–178.

21. Pandey, A.; Jain, K. An intelligent system for crop identification and classification from UAV images using conjugated dense convolutional neural network. *Comput. Electron. Agric.* **2022**, *192*, 106543. [CrossRef]

22. Reedha, R.; Dericquebourg, E.; Canals, R.; Hafiane, A. Transformer Neural Network for Weed and Crop Classification of High-Resolution UAV Images. *Remote Sens.* **2022**, *14*, 592. [CrossRef]

23. Bouguettaya, A.; Zarzour, H.; Kechida, A.; Taberkit, A.M. Deep learning techniques to classify agricultural crops through UAV imagery: A review. *Neural Comput. Appl.* **2022**, *34*, 9511–9536. [CrossRef]

24. Weiss, K.; Khoshgoftaar, T.M.; Wang, D.D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 1345–1459. [CrossRef]

25. Al Sahili, Z.; Awad, M. The power of transfer learning in agricultural applications: Agrinet. *Front. Plant Sci.* **2022**, *13*, 992700. [CrossRef] [PubMed]

26. Siddharth, T.; Kirar, B.S.; Agrawal, D.K. Plant species classification using transfer learning by pre-trained classifier VGG-19. *arXiv* **2022**, arXiv:2209.03076.

27. Sagi, O.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249. [CrossRef]

28. Tariku, G.; Ghiglieno, I.; Gilioli, G.; Gentilin, F.; Armiraglio, S.; Serina, I. Automated identification and classification of plant species in heterogeneous plant areas using unmanned aerial vehicle-collected RGB images and transfer learning. *Drones* **2023**, *7*, 599. [CrossRef]

29. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.

30. Mishra, A. Contrast limited adaptive histogram equalization (CLAHE) approach for enhancement of the microstructures of friction stir welded joints. *arXiv* **2021**, arXiv:2109.00886.

31. Ancuti, C.O.; Ancuti, C.; De Vleeschouwer, C.; Bekaert, P. Color balance and fusion for underwater image enhancement. *IEEE Trans. Image Process.* **2017**, *27*, 379–393. [CrossRef]

32. Simonyan, Z.-A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

33. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

34. Suthaharan, S. Support Vector Machine. In *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*; Suthaharan, S., Ed.; Springer: Boston, MA, USA, 2016; pp. 207–235. [CrossRef]

35. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

36. Yang, Y.; Newsam, S. Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification. In Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010.

37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June–28 July 2016; pp. 770–778. Available online: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html (accessed on 22 April 2023).

38. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2018**, arXiv:1608.06993. [CrossRef]

39. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June–28 July 2016; pp. 2818–2826. Available online: https://www.cvfoundation.org/openaccess/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html (accessed on 22 April 2023).

40. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.

41. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv* **2019**, arXiv:1801.04381. [CrossRef]