Sinopfbd, 2025; 10(2): 497-512 https://doi.org/10.33484/sinopfbd.1749697

Sinop Üniversitesi Fen Bilimleri Dergisi
*Sinop Uni J Nat Sci*

E-ISSN: 2564-7873

https://dergipark.org.tr/tr/pub/sinopfbd

# Disease Detection from Grape Plant Leaves Using Transfer Learning Methods

**Fatih YÜCALAR[1] and Ramazan YILDIRIM[1]**

How to cite: Yücalar, F., & Yıldırım, R. (2025). Disease detection from grape plant leaves using transfer learning methods. *Sinop Üniversitesi Fen Bilimleri Dergisi, 10*(2), 497-512. https://doi.org/10.33484/sinopfbd.1749697

**Research Article**

**Corresponding Author**
Fatih YÜCALAR
fatih.yucalar@cbu.edu.tr

**ORCID of the Authors**
F.Y: 0000-0002-1006-2227
R.Y: 0009-0004-2047-4323

**Received:** 24.07.2025
**Accepted:** 06.10.2025

**Abstract**

Today, the agricultural sector faces significant challenges due to population growth and limited resources. Enhancing productivity and minimizing losses is of great importance for the sustainability of agriculture. Therefore, leveraging technological advancements plays a critical role, particularly in the development of sustainable farming practices. Among these advancements, artificial intelligence (AI) stands out with its potential to contribute significantly to agricultural production. The primary objective of this study is to provide farmers with fast and accurate information regarding plant health, thereby preventing the spread of diseases and optimizing agricultural output. In line with this goal, AI-based image processing techniques were employed. Specifically, this study focuses on detecting grapevine leaf diseases namely powdery mildew (*Erysiphe necator*), downy mildew (*Plasmopara viticola*), and grapevine rust mite (*Eriophyes vitis*) using AI. Disease detection was carried out using leaf images, which were then used for classification. A hybrid dataset was constructed using a combination of publicly available images and manually collected samples captured via smartphone cameras in vineyards, fields, and gardens. This diverse and balanced dataset was used to train several CNN-based transfer learning models, including AlexNet, DarkNet53, Inception-ResNet-V2, Inception-V3, MobileNet-V3, ResNet50, ResNet101, VGG16, and VGG19 architectures. Among these, Inception-ResNet-V2 achieved the best performance with an accuracy of 97.45%, a training loss of 8.19%, a test accuracy of 93.00%, and a test loss of 20.60%. These results demonstrate that the model performs well in detecting diseases from grapevine leaves during both training and testing phases.

**Keywords:** Artificial intelligence, transfer learning, plant disease detection, grape leaf diseases, convolutional neural networks

## Aktarımlı Öğrenme Yöntemleri ile Üzüm Bitkisi Yaprağından Hastalık Tespiti

[1]Manisa Celal Bayar University, Hasan Ferdi Turgutlu Technology Faculty, Department of Software Engineering

**Öz**

Günümüzde tarım sektörü, nüfus artışı ve kaynakların sınırlı olması gibi zorluklardan etkilenmektedir. Tarım sektörü için verimliliği artırmak ve kayıpları en aza indirmek büyük önem taşımaktadır. Bu nedenle, teknolojinin getirdiği yeniliklerden yararlanmak, özellikle sürdürülebilir tarım uygulamalarının geliştirilmesinde kritik bir rol oynamaktadır. Yapay zekâ, bu yeniliklerin başında gelmekte olup, tarımsal üretime katkı sağlama potansiyeline sahiptir. Bu çalışmanın temel amacı, bitki sağlığı konusunda çiftçilere hızlı ve doğru bilgiler sağlayarak, hastalıkların yayılmasını önlemek ve tarımsal üretimi optimize etmektir. Bu hedef doğrultusunda, yapay zekâ tabanlı görüntü işleme tekniklerinden yararlanılmıştır. Bu kapsamda, üzüm bitkisi yaprağı üzerinden bağ küllemesi (*Erysiphe necator*), mildiyö (*Plasmopara viticola*) ve bağ uyuzu (*Eriophyes vitis*) hastalıklarının yapay zekâ ile tespiti sağlanmıştır. Hastalık tespiti için yaprak görüntüleri kullanılmış ve bu görüntüler üzerinden sınıflandırma gerçekleştirilmiştir. Çalışma kapsamında, bir kısmı hazır olarak temin edilen, bir kısmı ise bağ, tarla, bahçe gibi ortamlardan cep telefonu kamerası ile manuel olarak elde edilen çeşitli ve dengeli örneklerden oluşan bir karma veri seti oluşturulmuştur. Oluşturulan bu karma veri seti, CNN tabanlı aktarımlı öğrenme yöntemlerinden AlexNet, DarkNet53, Inception-ResNet-V2, Inception-V3, MobileNet-V3, ResNet50, ResNet101, VGG16 ve VGG19 mimarileri üzerinde eğitilmiştir. Eğitim ve test işlemleri sonucunda; %97.45 doğruluk, %8.19 eğitim kaybı, %93.00 test doğruluğu ve %20.60 test kaybı değerleri ile en başarılı model olarak Inception-ResNet-V2 belirlenmiştir. Bu sonuç, modelin hem eğitim hem de test verilerinde üzüm bitkisi yaprağı üzerinden hastalık tespiti için yüksek performans gösterdiğini ortaya koymaktadır.



**Anahtar Kelimeler:** Yapay zekâ, transfer öğrenmesi, bitki hastalığı tespiti, üzüm yaprağı hastalıkları, evrişimli sinir ağları

## Introduction

Agriculture remains one of the fundamental pillars of the global economy. Millions of people worldwide are directly employed in the agricultural sector, and economies of many countries rely heavily on it. In developing countries, agriculture contributes significantly to GDP and provides a large share of employment. However, the sector faces several challenges, including climate change, diminishing water resources, declining soil fertility, and plant diseases. These issues threaten not only agricultural output but also food security and farmers' income. Among these, plant diseases pose a major risk to productivity [1]. They can hinder growth, reduce quality and yield, and sometimes cause total crop loss. Therefore, early detection and effective treatment are critical to minimize losses and improve productivity. Integrating modern technologies, particularly AI-based analytical methods, into agriculture offers new solutions for managing these diseases [2]. Studying grapevines is highly relevant in this context. Grapes are widely cultivated and have substantial economic value. They are consumed fresh and used in processed products such as wine, vinegar, and dried fruit. Under certain environmental conditions, grapevine leaves are susceptible to various diseases. Common diseases include powdery mildew (*Erysiphe necator*), downy mildew (*Plasmopara viticola*), and grapevine rust mite (*Eriophyes vitis*) [3]. These diseases, caused by fungi or mites, can spread rapidly, especially under humid conditions, reducing both quality and yield. Timely and accurate diagnosis is therefore essential for effective intervention. Leaves are among the most important indicators for early disease detection. Visual symptoms such as discoloration, spots, drying, and vein disruptions provide diagnostic clues. However, these symptoms are not always easily detected by the human eye. In large vineyards, inspecting all plants regularly is time-consuming and labor-intensive, increasing the likelihood of missed symptoms. Here, AI-assisted image processing can significantly aid detection. AI methods analyze leaf features color, texture, shape at the pixel level and convert them into digital signals. These representations are processed using Convolutional Neural Networks (CNNs) trained on labeled image datasets [4]. CNNs can detect even microscopic changes that humans might miss. They automatically learn features such as color variations, texture patterns, and morphology with high accuracy, identifying both visible and subtle early-stage anomalies. CNNs have been successful due to their ability to learn hierarchical relationships among image features [5]. They can differentiate disease types and recognize disease progression stages. Classification results can be delivered to farmers in real-time via mobile apps or cloud-based decision systems. This enables early intervention and reduces potential yield loss. Compared to traditional inspection, AI-based methods are faster and more accurate. They form a vital component of decision support systems, making early diagnosis more accessible and preserve plant health. Early and accurate disease detection on grape leaves is crucial for timely intervention and reduced crop loss. However, deep learning models require large, balanced datasets to perform well. Creating such datasets in agriculture is time-consuming and costly. Environmental conditions, lighting variations, and camera quality add further complexity, affecting model generalization. Transfer learning addresses these issues [6]. It allows models pretrained on large datasets (e.g., ImageNet) to be retrained for specific tasks [7]. Fundamental visual features learned by the model edges, textures, gradients can then be reused, reducing training time and enabling high performance even with limited data. The remainder of this paper is organized as follows. The next section reviews related work on plant disease detection using deep learning and transfer learning. Materials and methods are then presented, covering dataset preparation, augmentation, and model development. Experimental results from training and evaluating various CNN architectures follow. The discussion highlights the approach's strengths and limitations, and the paper concludes with a summary and suggestions for future research.

## Related Work

Automated detection of grapevine leaf diseases is crucial for maintaining agricultural productivity and crop quality. However, research in this area remains limited in number and scope compared to other plant species. The existing, albeit limited, studies show that deep learning and transfer learning techniques yield promising results in identifying grape leaf diseases. Despite these positive outcomes, there's a clear need to develop more comprehensive and effective methods, including larger datasets and extensive model comparisons. In this context, methods developed by various researchers for different plant species and leaf diseases can provide valuable insights for studies specifically focusing on grape leaves. This section aims to summarize the current literature in this field. Fang et al. [8] utilized a Convolutional Neural Network (CNN), a deep learning algorithm, for the detection of apple leaf diseases. Specifically, they opted for VGG16, a CNN-based architecture commonly used in deep learning and known for its successful performance in complex visual recognition tasks, to enhance classification accuracy. The researchers combined the center loss function with the softmax loss function during model training. This approach is frequently employed to improve classification performance by making the distinctions between classes more pronounced. The dataset used in their study comprised 5373 images of diseased apple leaves and 1683 images of healthy leaves. The diseased leaf images included common ailments such as anthracnose blight, cedar rust, leaf rust, gray spot, black spot, and black rot. This diversity allowed for the evaluation of the model's ability to recognize various disease types. The results demonstrated that the VGG16 architecture offered high classification success, with the model's accuracy rates ranging from 95% to 99.70% [8]. In their research, Yaman and Tuncer [9] combined deep learning and machine learning methods to identify diseases in tree and plant leaves. They collected a dataset of 726 images of walnut leaves, categorized into two classes: healthy and diseased. Deep learning models were then used to extract meaningful features from these images. The study tested 17 different deep learning models, ultimately selecting DarkNet53 and ResNet101 as the two highest-performing models. The features extracted by these two models were then combined to create a hybrid feature set, offering a more robust representation. These hybrid features were subsequently evaluated using machine learning classifiers, achieving a high success rate of 99.58% [9]. Sladojević et al. [10] developed a CNN-based model capable of accurately identifying both healthy plants and 13 different diseased plant species from leaf images. To enhance classification accuracy, their study included a class for background images in addition to plant leaves. This approach allowed the model to produce more robust results when faced with visual complexity. A total of 30880 images were used for training, and 2589 images for testing. The Caffe deep learning framework was used to develop the CNN, and the model architecture was designed with 8 learning layers: 5 convolutional layers and 3 fully connected layers. The results demonstrated that the model achieved precision values ranging from 91% to 98% and an overall accuracy rate of 96.3% [10]. Wagle et al. [11] developed a deep learning model for detecting diseases in tomato plants, leveraging data augmentation techniques. For this purpose, they used healthy and diseased tomato leaf images from the PlantVillage dataset, with disease categories selected based on their agricultural prevalence and impact in India. During the model development process, transfer learning architectures such as ResNet50, ResNet18, and ResNet101 were integrated with a softmax classification layer. The dataset was enhanced by applying various data augmentation techniques, including noise, blur, positional shifts, and color variations. This significantly improved the model's generalization ability and reduced the risk of overfitting. It was observed that the applied data augmentation methods considerably increased classification accuracy. The ResNet101 model, trained with the augmented dataset, demonstrated the best performance, achieving 99.99% training accuracy and 95.83% test accuracy [11]. Rao et al. [12] investigated the detrimental effects of agriculturally significant microbial diseases on food security in the Indian agricultural sector. Within this scope, they focused on grape and mango plants, performing automatic plant disease detection from raw images using deep learning and transfer learning techniques. Their study successfully detected diseases in grape and mango leaves using a balanced dataset comprising 8438 images of both diseased and healthy leaves. For automatic feature extraction and classification, they applied the AlexNet transfer learning model, which achieved an accuracy of 99% for grape leaves and 89% for mango leaves. They subsequently developed an Android application named "JIT CROPFIX" to make this research accessible to farmers. This application aims to enhance agricultural productivity by providing farmers with a quick and practical means of identifying plant diseases [12]. In the study conducted by Ajra et al. [13], a method involving the AlexNet and ResNet-50 models was proposed for disease detection using image data of tomato and potato leaves. The leaf images obtained from the Kaggle dataset were subjected to preprocessing, data augmentation, and feature extraction before classification. Experimental results demonstrated that the ResNet-50 model achieved an accuracy of 97.3%, while the AlexNet model reached 95.9%. This study aims to contribute to the early protection of plant health and the improvement of agricultural production [13]. Nader et al. [14] developed a deep learning-based ensemble approach for the classification of grapevine leaf diseases. Their study focused on three significant and commonly observed diseases in grapevines: Black Measles, Black Rot, and Leaf Blight. To achieve this, pre-trained VGG16, VGG19, and Xception models from ImageNet were employed using transfer learning and retrained on grape leaf images from the PlantVillage dataset. The developed ensemble model combined the strengths of these CNN architectures, resulting in a more robust classification performance with higher accuracy. Furthermore, data preprocessing and data augmentation techniques were applied to improve the overall performance of the model. The experimental results demonstrated that the proposed ensemble structure outperformed both the individual models and other methods in the literature, achieving an impressive accuracy rate of 99.82% [14].

## Materials and Methods

This section outlines the materials and methodological steps employed in the study, including the preparation of the dataset, the generation of an augmented hybrid dataset using data augmentation techniques, and the development of the deep learning model.

## Preparation of the Dataset

The preparation of the dataset used in this study involved a two-stage process. In the first stage, previous research on grapevine leaf diseases was reviewed. From these studies, existing datasets containing images of both diseased and healthy grapevine leaves were gathered [15–18]. These pre-existing datasets served as the initial data source for model training. In the second stage, in addition to the existing datasets, various natural cultivation areas such as vineyards, gardens, and orchards were visited. High-resolution images were manually captured using a mobile phone. Specifically, the image acquisition process was carried out using an iPhone 11 smartphone camera. Photographs were taken in daylight under sunny and well-lit conditions, ensuring natural illumination of the leaves. Images were captured at close range to clearly highlight leaf texture, color, and vein details, with a resolution consistent with the default camera settings of the device. These conditions were chosen to minimize shadows and improve the visibility of disease-related symptoms. The collected images were then classified according to the disease types examined in this study, ensuring the inclusion of samples from both diseased and healthy grapevine leaves. All of the collected data were combined to create a more diverse and balanced hybrid dataset, encompassing both real-world field data and pre-existing datasets.

## Replication of a Hybrid Dataset Through Data Augmentation

A data augmentation method was employed to enhance the generalization capability of the training model to enrich the hybrid dataset generated from the data collection process, and thereby overfitting was prevented. In this context, various transformation operations such as rotation, horizontal and vertical shifting, zooming, and brightness adjustment were applied using the "ImageDataGenerator" class provided by the Keras library. These operations aimed to improve the model's robustness when confronted with different variations of the images. As illustrated in Figure 1, these augmentation techniques generated diverse variations of the same leaf sample, thereby further enhancing the robustness of the dataset.

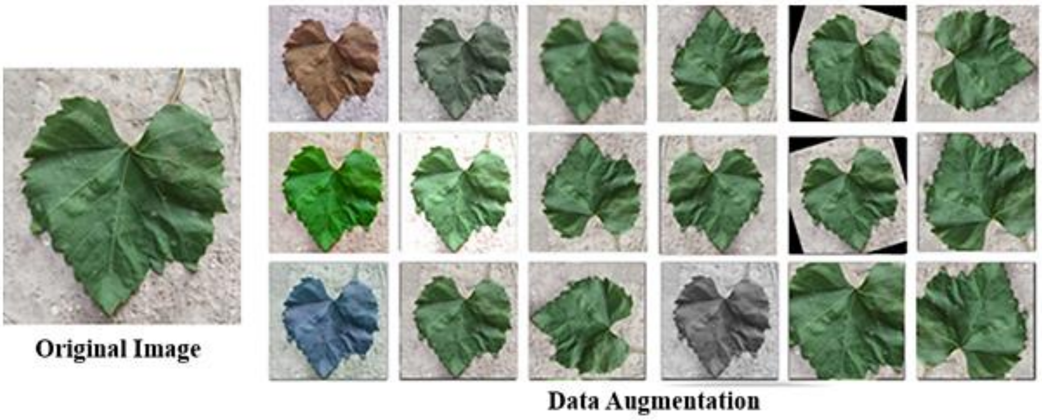

***Figure 1.*** *Data augmentation example*

In this study, data augmentation techniques were applied to all four classes Powdery Mildew, Downy Mildew, Grapevine Rust Mite, and Healthy Leaves. Since the number of images per class was nearly balanced after data collection, augmentation was performed uniformly across classes to maintain this balance. Specifically, each original image was augmented by generating two additional transformed samples using random combinations of rotation (±25°), horizontal/vertical shifting (up to 10%), zooming (up to 15%), and brightness variation (±20%). As a result, the dataset size for each class approximately tripled, ensuring both diversity and balance across the four classes. Following the application of data augmentation methods, the resulting hybrid dataset was divided into four distinct classes. The distribution of training and test data for each class was determined as follows:

- *Powdery Mildew (Erysiphe Necator) class:* A total of 903 images, with 801 for training and 102 for testing.

- *Downy Mildew (Plasmopara Viticola) class:* A total of 898 images, with 797 for training and 101 for testing.
- *Grapevine Rust Mite (Eriophyes Vitis) class:* A total of 897 images, with 797 for training and 100 for testing.
- *Healthy Leaf class:* A total of 907 images, with 806 for training and 101 for testing.

This balanced distribution ensured equal representation for each class during the model's learning process, which was a crucial step towards improving classification performance.

## Model Development

The model development process began with the selection of suitable transfer learning architectures and continued with the tuning of appropriate hyperparameters for each model. Subsequently, training was conducted on the selected deep learning architectures, and the classification performance of the models was evaluated. It is important to note that techniques such as early stopping, dropout, or additional regularization were not employed in this study. The primary objective was to perform a comparative evaluation of different transfer learning architectures under identical training conditions. The use of early stopping could have led to inconsistent results by allowing some models to terminate training prematurely while others continued for a longer number of epochs. Similarly, applying dropout or other regularization strategies would have required additional hyperparameter tuning, which might have favored certain architectures and undermined the fairness of the comparison. Instead, overfitting was mitigated through the construction of a balanced dataset and the application of extensive data augmentation, ensuring that performance differences among models were primarily due to their intrinsic architectural characteristics.

## Selection of Transfer Learning Methods

Based on the information obtained through literature research, the transfer learning methods to be used in training the project's artificial intelligence model were identified. The review focused particularly on CNN-based transfer learning models, which have demonstrated the most successful results in the classification of plant diseases. Accordingly, the architectures identified in the literature as exhibiting high performance AlexNet [13], DarkNet53 [19], Inception-ResNet-V2 [20], Inception-V3 [21], MobileNet-V3 [21], ResNet50 [13], ResNet101 [22], VGG16 [19], and VGG19 [20] were deemed suitable for use in model training. These selected models possess the potential to meet the project's requirements in terms of both classification accuracy and computational efficiency. Multiple criteria influenced the selection of these models. First and foremost, their prior success in achieving high accuracy in earlier studies was taken into account. Additionally, the architectural design, layer depth, number of parameters, and computational costs vary among the models. This diversity allows for a comprehensive comparison of model performances in terms of both accuracy and processing time. For instance, AlexNet offers a relatively simple architecture with advantages such as shorter training time and lower computational cost. On the other hand, ResNet50 and ResNet101 architectures, with their residual connections, effectively overcome learning difficulties even in deep network structures. Inception-V3 and Inception-ResNet-V2 stand out with their modular designs and parameter efficiency, enabling the model to learn more complex patterns with fewer parameters. MobileNet-V3, which is highly suitable for mobile deployment, provides advantages in scenarios requiring low resource consumption. Lastly, VGG16 and VGG19, with their fixed filter sizes and regular layer structures, offer architectures that are easy to understand and implement, although they have higher computational costs compared to other models. For all these reasons, using models with varying architectural characteristics during the model training phase is expected to enhance the project's success and help determine which architecture is best suited for the target problem.

## Hyperparameter Optimization and Model Training

At this stage, the hyperparameters for all transfer learning models to be used were defined, and the model training was conducted in a two-phase process. In the first phase, the models were trained for 10 epochs to evaluate their general performance. Those that demonstrated satisfactory results were retrained for 20 epochs in the second phase for a more comprehensive analysis. Throughout both training phases, the hyperparameters were kept constant: the batch size was set to 32, and the learning rate to 0.0001. The input image size was configured as 299×299 pixels for the Inception-V3 and Inception-ResNet-V2 architectures, and 224×224 pixels for all other architectures. Since the dataset had already been split into training and test subsets, validation was performed directly on the test data. The results of this training process will be presented in detail in the following section, in terms of accuracy, loss, training time, and classification performance. The training parameters used in this study are summarized in Table 1 to ensure clarity and reproducibility of the experimental setup. These parameters were kept consistent across all experiments unless otherwise stated.

***Table 1.*** *Training parameters used in the study*

| Parameter | Value | Description |
|---|---|---|
| Optimizer | Adam | Selected for its adaptive learning rate optimization and suitability for CNN training |
| Learning Rate | 0.0001 | Fixed throughout the training process |
| Batch Size | 32 | Number of images processed per training iteration |
| Number of Epochs | Phase 1: 10 epochs<br>Phase 2: 20 epochs | Initial evaluation in the first phase, extended training in the second phase |
| Input Image Size | 299 × 299 (Inception-V3, Inception-ResNet-V2) 224 × 224 (other CNN models) | Adjusted according to the specific requirements of each model |

## Experimental Results

In the initial phase of model training, the augmented hybrid dataset was used to train all the selected transfer learning models. Each model was trained under identical conditions to ensure that the resulting performance outcomes were comparable. Model performance was evaluated based on key metrics such as training accuracy, training loss, validation accuracy, and validation loss [23]. These metrics were used to assess both the model's effectiveness during training and its ability to generalize to unseen test data. Training accuracy indicates the proportion of correct classifications on the training set, while training loss reflects the magnitude of prediction errors. Likewise, validation accuracy represents the model's accuracy on the test set, and validation loss measures the prediction error on these unseen samples [24]. Evaluating these metrics collectively provides comprehensive insight into the model's learning capacity, generalization performance, and potential risk of overfitting. Therefore, these core metrics were used together to compare model performance. Table 2 summarizes the comparative results obtained after 10 epochs of training.

***Table 2.*** *Model performance results after 10 epochs of training*

| Models | Training Accuracy | Training Loss | Validation Accuracy | Validation Loss |
|---|---|---|---|---|
| AlexNet | 0.9317 | 0.1944 | 0.5875 | 2.3570 |
| DarkNet53 | 0.9501 | 0.1415 | 0.2400 | 4.6175 |
| Inception-ResnetV2 | 0.9470 | 0.1704 | 0.9125 | 0.2817 |
| InceptionV3 | 0.9275 | 0.2368 | 0.9075 | 0.3236 |
| MobileNetV3 | 0.9166 | 0.3012 | 0.9025 | 0.2641 |
| ResNet50 | 0.2851 | 1.4075 | 0.3250 | 1.3754 |
| ResNet101 | 0.2866 | 1.4080 | 0.4300 | 1.3809 |
| VGG16 | 0.9917 | 0.0293 | 0.9325 | 0.2905 |
| VGG19 | 0.9774 | 0.0824 | 0.9200 | 0.2785 |

As seen in the Table 2, signs of overfitting were observed in the AlexNet, DarkNet53, ResNet50, and ResNet101 models. This was clearly demonstrated by their relatively high training accuracy but poor validation performance. Due to this negative impact on generalization ability, these models were excluded from further evaluation in subsequent stages. Following the elimination of the overfitting models, the remaining five models Inception-ResNet-V2, Inception-V3, MobileNet-V3, VGG16, and VGG19 underwent a second training phase. In this stage, each model was retrained for 20 epochs, and their performances on both the training and validation datasets were analyzed. Table 3 presents the performance metrics obtained from these trainings, while Figure 2 illustrates the epoch-wise changes in accuracy and loss for each model.

***Table 3.*** *Model performance results after 20 epochs of training*

| Models | Training Accuracy | Training Loss | Validation Accuracy | Validation Loss |
|---|---|---|---|---|
| Inception-ResnetV2 | 0.9745 | 0.0819 | 0.9300 | 0.2060 |
| InceptionV3 | 0.9527 | 0.1673 | 0.9100 | 0.2853 |
| MobileNetV3 | 0.9429 | 0.1799 | 0.9225 | 0.2124 |
| VGG16 | 0.9717 | 0.0698 | 0.9475 | 0.2863 |
| VGG19 | 0.9932 | 0.0258 | 0.9275 | 0.2938 |

After examining Table 3 and Figure 2 following the 20-epoch training process, it was determined that this training yielded more successful results compared to the 10-epoch training. In particular, a noticeable improvement in overall performance was observed in some models alongside increases in both training and validation accuracies. However, upon closer analysis of the VGG16 and VGG19 models, although both exhibited high accuracy and low loss values, the graphical evaluation revealed a growing gap between training and validation metrics, indicating signs of overfitting. Consequently, the VGG16 and VGG19 models were excluded from further evaluation, and the final assessment proceeded only with models that did not exhibit overfitting. In this study, the dataset was carefully balanced, with each class containing approximately equal numbers of samples. Therefore, the primary evaluation metrics accuracy and loss were considered sufficient for assessing model performance. Accuracy provides a direct measure of the proportion of correctly classified samples, while loss reflects the magnitude of prediction errors during training and testing. It should be noted that if the dataset had been imbalanced, accuracy alone might not have adequately represented the model’s true performance. In such cases, additional metrics such as Precision, Recall, and F1-Score would be critical to evaluate the model’s ability to correctly identify samples from minority classes. Since the current dataset is balanced, these additional metrics were deemed supplementary rather than necessary for performance evaluation. As depicted in Figure 2, VGG16 and VGG19 exhibited a widening gap between training and validation curves, which is a strong indication of overfitting, whereas Inception-ResNet-V2 maintained stable convergence.

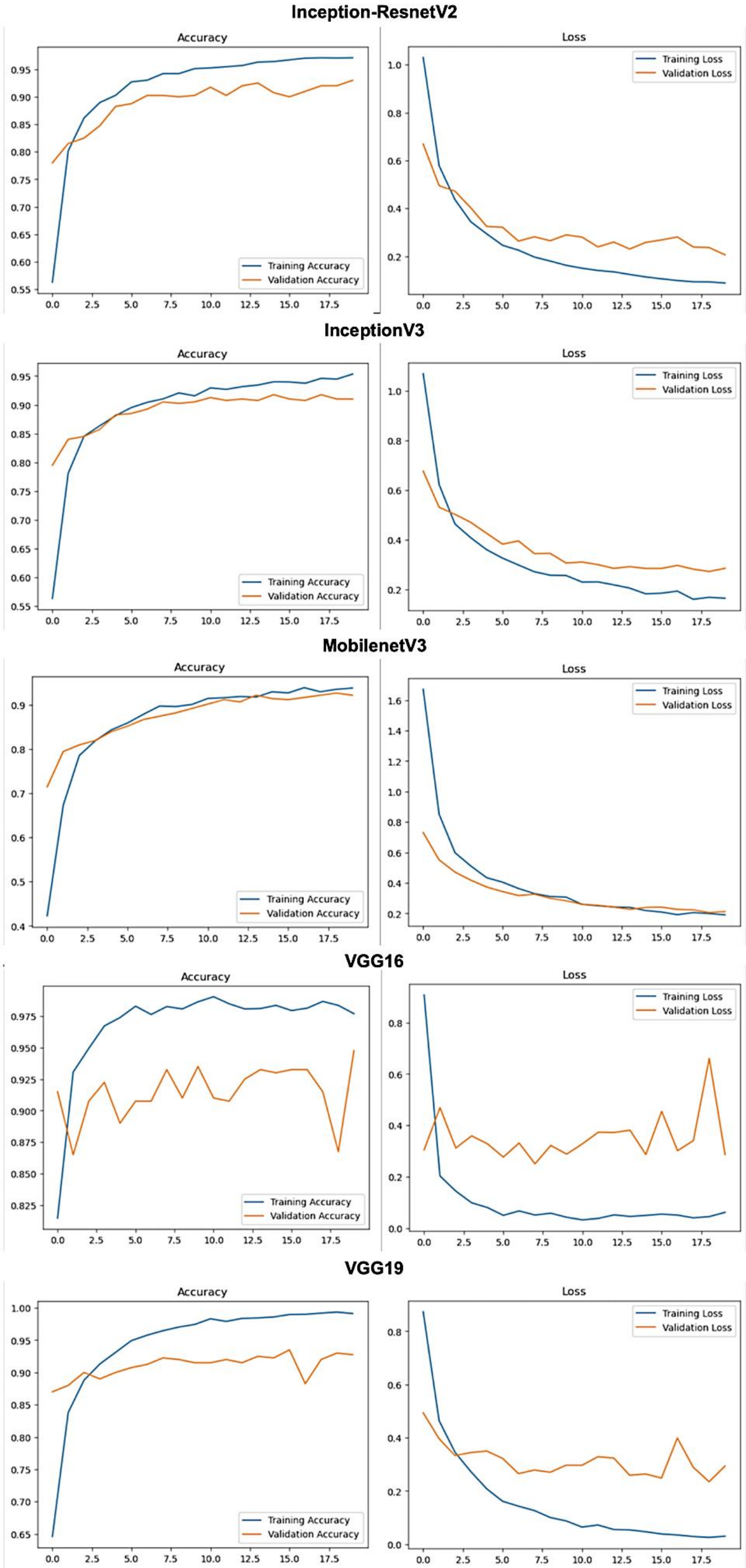


***Figure 2.*** *Model graphs after 20 epochs of training*

After the 20-epoch training process, the remaining models Inception-ResNet-V2, Inception-V3, and MobileNet-V3 were re-evaluated based on the data presented in the tables and graphs. The analyses revealed that the Inception-ResNet-V2 model demonstrated the best performance. Evaluation metrics such as the Confusion Matrix [25] and the Receiver Operating Characteristic (ROC) curve [26] were utilized to examine the classification success of this model in greater detail and to assess its accuracy across different classes. These analyses are presented in Figure 3.

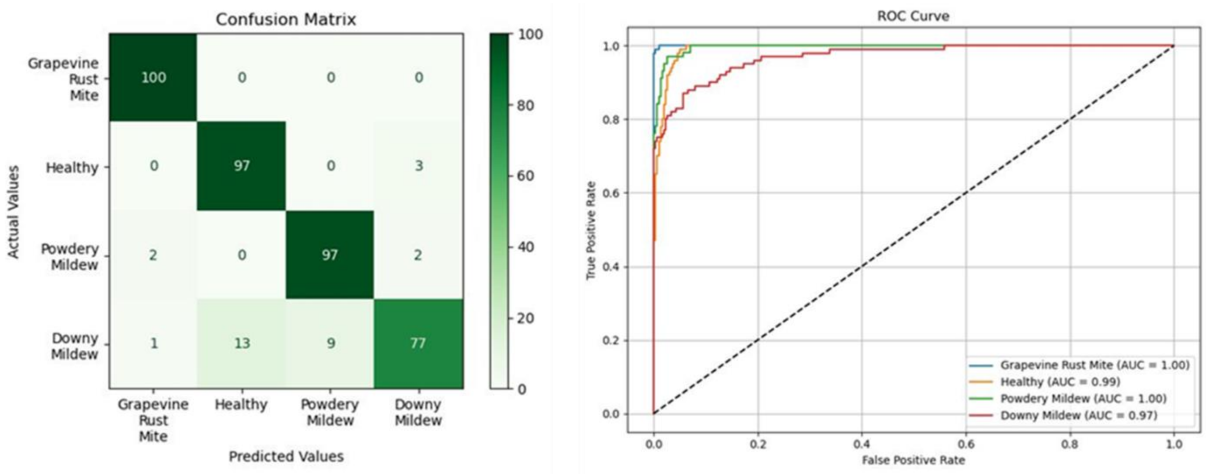


***Figure 3.*** *Confusion matrix and ROC curve of the Inception-ResNet-V2 Model*

The confusion matrix in Figure 3 clearly shows that the Inception-ResNet-V2 model classified healthy leaves and powdery mildew samples with high accuracy, while minor misclassifications occurred in downy mildew and rust mite categories. Additionally, the ROC curve demonstrates strong discriminative power, with AUC values approaching 1.0 across all classes.

## Discussion

In this study, the classification of grape leaf diseases was carried out using various transfer learning-based CNN architectures. Comparative analyses showed that the Inception-ResNet-V2 model was the most successful, due to its high accuracy and stability. The model achieved a training accuracy of 97.45% and a test accuracy of 93.00%, demonstrating balanced and reliable performance in both learning and generalization. These results show that, despite the relatively small dataset, transfer learning methods can effectively address agricultural imaging problems where data availability is limited. Performance differences among models were influenced by architectural depth, parameter count, and hardware requirements. For example, deeper architectures such as VGG16, VGG19, and ResNet101 achieved high training accuracy; however, their test accuracy declined, indicating overfitting. This excessive adaptation to training data reduced their ability to generalize to new samples. In contrast, Inception-ResNet-V2 combined high accuracy with strong generalization, benefiting from both architectural depth and balanced parameter optimization. Although part of the dataset was manually collected to increase diversity, variations in image quality created learning difficulties in certain classes.

Differences in lighting, background, damaged leaves, or varying angles made accurate learning more challenging. Additionally, class imbalance led to underrepresentation of some diseases during training, as shown in Figure 3, resulting in decreased classification performance for those classes. Misclassification rates were higher for classes with fewer samples. While data augmentation partially mitigated these issues, class imbalance and image variability remain significant factors affecting model success. Hardware limitations and long training times were another challenge. Despite using Manisa Celal Bayar University laboratory facilities, each model required about five hours to train. This underscores the importance of time management, particularly when evaluating multiple models. High hardware demands also pose constraints for academic research with limited resources. In summary, the success of deep learning models depends not only on architecture but also on data quality, class balance, and computational resources.

## Conclusion and Future Work

In this study, various CNN architectures based on transfer learning were compared for the detection of grape leaf diseases. Taking into account both high accuracy and model stability, the Inception-ResNet-V2 architecture was identified as the most suitable model. Its balanced and consistent performance on both the training and test datasets was a decisive factor in its selection. The high accuracy results obtained demonstrate that transfer learning methods can be effectively applied in domains such as agriculture, where labeled data is often limited. The use of data augmentation techniques and the manual collection of field data enhanced the model's generalization capability, enabling it to make accurate predictions on previously unseen samples. The Inception-ResNet-V2 model shows strong potential for integration into future mobile applications. Through such an application, farmers could detect grape leaf diseases quickly and accurately using their mobile devices, enabling early intervention and minimizing crop losses. This approach would not only help reduce economic losses but also support the sustainability of agricultural production. The study highlights the tangible benefits of artificial intelligence-based solutions in the agricultural sector and provides a solid foundation for practical implementation. In particular, the widespread adoption of such technologies in developing countries could significantly enhance the effectiveness of digital agriculture practices. Future research could focus on improving model accuracy by employing larger and more balanced datasets. Furthermore, ensemble methods that combine multiple models rather than relying on a single architecture may yield more robust and generalizable results. In previous studies, hybrid deep learning models that combine complementary architectures have been reported to achieve higher classification accuracy. This indicates that such an approach could also enhance the robustness of grapevine leaf disease detection. These methods leverage the strengths of individual models while minimizing their limitations. Additionally, moving beyond purely image-based CNN architectures to hybrid systems that incorporate time-series or sensor-based data could further improve accuracy and broaden the system's applications. Similarly, the integration of multimodal data, such as image and sensor-based information, has been shown to improve recognition performance in various domains. Extending the current dataset with sensor inputs may therefore provide more reliable outcomes in future applications. Such systems could evolve into comprehensive decision-support tools by considering environmental factors such as temperature, humidity, and soil properties. Ultimately, the proposed model has significant potential not only for classifying leaf diseases but also for integration into broader agricultural health monitoring and disease management systems.

***Acknowledgements –***

***Funding/Financial Disclosure*** The authors have not received any financial support for the research. authorship or publication of this study.

***Ethics Committee Approval and Permissions*** This study does not require ethics committee permission or any special permission.

***Conflict of Interests*** No conflict of interest or common interest has been declared by the authors.

***Authors Contribution*** The authors contributed equally to the study.

## References

[1] Gai, Y., & Wang, H. (2024). Plant disease: A growing threat to global food security. *Agronomy*, *14*(8), 1615. https://doi.org/10.3390/agronomy14081615

[2] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147,* 70-90. https://doi.org/10.1016/j.compag.2018.02.016

[3] Pertot, I., Caffi, T., Rossi, V., Mugnai, L., Hoffmann, C., Grando, M. S., Gary, C., Lafond, D., Duso, C., Thiery, D., Mazzoni, V., & Anfora, G. (2017). A critical review of plant protection tools for reducing pesticide use on grapevine and new perspectives for the implementation of IPM in viticulture. *Crop Protection, 97,* 70-84. https://doi.org/10.1016/j.cropro.2016.11.025

[4] Ferentinos, K. P. (2018) Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture, 145,* 311-318. https://doi.org/10.1016/j.compag.2018.01.009

[5] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for ımage-based plant disease detection. *Frontiers in Plant Science, 7*, 1419. https://doi.org/10.3389/fpls.2016.01419

[6] Prashanthi, B., Praveen Krishna, A. V., & Mallikarjuna Rao, C. (2025). A comparative study of fine-tuning deep learning models for leaf disease ıdentification and classification. *Engineering, Technology & Applied Science Research, 15*(1), 19661-19669.

[7] Mishra, A., Mishra, A., Tewari, A. K., & Gangrade, J. (2023, December). *Deep Transfer Learning for Tomato Leaf Diseases Detection and Classification using Pre-trained Models*. 2023 9$^{th}$ International Conference on Signal Processing and Communication (ICSC), Noida, India. https://ieeexplore.ieee.org/document/10441215

[8] Fang, T., Chen, P., Zhang, J., & Wang, B. (2019, July). Identification of apple leaf diseases based on convolutional neural network. In *International Conference on Intelligent Computing* (pp. 553-564). Cham: Springer International Publishing.

[9] Yaman, O., & Tuncer, T. (2022). Bitkilerdeki yaprak hastalığı tespiti için derin özellik çıkarma ve makine ögrenmesi yöntemi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 34*(1), 123-132. https://doi.org/10.35234/fumbd.982348

[10] Sladojevic S, Arsenovic M, Anderla A, Culibrk D, & Stefanovic D. (2016). Deep neural networks-based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience, 2016,* 3289801. https://doi.org/10.1155/2016/3289801

[11] Wagle, S. A., Sampe, J., Mohammad, F., & Ali, S. H. M. (2021). Effect of data augmentation in the classification and validation of tomato plant disease with deep learning methods. *Traitement du Signal, 38*(6), 1657-1670. https://doi.org/10.18280/ts.380609

[12] Rao, U. S., Swathi, R., Sanjana, V., Arpitha, L., Chandrasekhar, K., & Naik, P. K. (2021). Deep learning precision farming: grapes and mango leaf disease detection by transfer learning. *Global Transitions Proceedings, 2*(2), 535-544. https://doi.org/10.1016/j.gltp.2021.08.002

[13] Ajra, H., Majid, M. A., Shohidul Islam, Md., & Abdullah, D. (2025). Leaf disease detection in plant care using CNN architecture: AlexNet and resnet-50 models. *International Journal on Advanced Science, Engineering and Information Technology (IJASEIT), 15*(1), 283–292. https://doi.org/10.18517/ijaseit.15.1.19944

[14] Nader, A., Khafagy, M. H., & Hussien, S. A. (2022). Grape leaves diseases classification using ensemble learning and transfer learning. *International Journal of Advanced Computer Science and Applications*, *3*(7), 563-571. https://doi.org/10.14569/IJACSA.2022.0130767

[15] Ananya Giliyal, A. (2022). Grape Disease Identification -Resnet50. [Dataset]. Kaggle. Grape_disease. Access Date: May,12, 2025. https://www.kaggle.com/datasets/pushpalama/grape-disease

[16] Roboflow. (2023). Grape Leaf Disease Computer Vision Project. [Dataset]. https://universe.roboflow.com/tru-projects-cqcql/grape-leaf-disease-dataset

[17] Hidiat. (2022). Grape Leaf Disease Classification with InceptionV3. [Dataset]. Kaggle. https://www.kaggle.com/code/yemi99/grape-leaf-disease-classification-with-inceptionv3/input

[18] Uygun, T. (2024). Vineyard Leaf Scab (*Eriophyes vitis* Pagst.) Damage. [Dataset]. Mendeley Data. V1, doi: 10.17632/79f2wmdpmr.1

[19] Akinci, T. C., Ekici, S., Kabir, M., & Martinez-Morales, A. A. (2024). HybridPlantNet23: A Scientific Insight into the Power of Ensemble Modelling using VGG16 and Darknet53 for Plant Disease Classification. *Preprints.* https://doi.org/10.20944/preprints202407.0556.v1

[20] Suksukont, A. (2025). Integration of InceptionResNetV2 with VGG19 for sugarcane leaf disease recognition. *RMUTSB Academic Journal, 13*(1), 99–109. https://doi.org/10.64989/rmutsbj.2025.266380

[21] Ataman, F., & Eroğlu, H. (2024). Comparative ınvestigation of deep convolutional networks in detection of plant diseases. *Turkish Journal of Nature and Science, 13*(3), 37-49. https://doi.org/10.46810/tdfd.1477476

[22] Chai, M. X., Fam, Y. D., Octaviano, Q. N., Pee, C.-Y., Wong, L. K., Mohd Hilmi Tan, M. I. S., & See, J. (2024, December). *Improved Cassava Plant Disease Classification with Leaf Detection*, *2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC),* Macau, Macao. https://ieeexplore.ieee.org/document/10849031

[23] Terven, J., Cordova-Esparza, D. M., Ramirez-Pedraza, A. Chavez-Urbiola, E. A., & Romero-González J. A. (2025). A comprehensive survey of loss functions and metrics in deep learning. *Artificial Intelligence Review, 58*(195). https://doi.org/10.1007/s10462-025-11198-7.

[24] Ghojogh, B., & Crowley, M. (2019). The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. *arXiv*. https://doi.org/10.48550/arXiv.1905.12787

[25] Bozuyla, M., & Özçift, A. (2022). Developing a fake news identification model with advanced deep language transformers for Turkish COVID-19 misinformation data. *Turkish Journal of Electrical Engineering and Computer Sciences, 30*(3), 908-926. https://doi.org/10.3906/elk-2106-55

[26] Karasulu, B., Yücalar, F., & Borandağ, E. (2022). A hybrid approach based on deep learning for gender recognition using human ear images. *Journal of the Faculty of Engineering and Architecture of Gazi University, 37*(3), 1579-1594. https://doi.org/10.17341/gazimmfd.945188