Geliş Tarihi (date of arrival): 2025-05-02
Kabul Tarihi (date of acceptance): 2025-06-30

MEYVECİLİK ARAŞTIRMA ENSTİTÜSÜ MÜDÜRLÜĞÜ
FRUIT RESEARCH INSTITUTE

e-ISSN: 2148-8770
Yıl /Year: 2025
Cilt(Sayı)/Vol.(Issue): 12 (1)
Sayfa/Page: 88-99

Araştırma Makalesi/
Research Article

DOI: 10.51532/meyve.1689356

Meyve Bilimi Fruit Science

Apricot Plant Disease and Pest Detection from Field Images Using Fine-Tuned CNNs and Symptom–Organ Level Labeling

Yahya ALTUNTAŞ[1]* Yusuf KARAKUŞ[2]

[1]Meyvecilik Araştırma Enstitüsü Müdürlüğü, 32500, Eğirdir, Isparta, Türkiye
[2]Kayısı Araştırma Enstitüsü Müdürlüğü, 44090, Yeşilyurt, Malatya, Türkiye

*yahya.altuntas@tarimorman.gov.tr (Corresponding author)

Abstract

Early and accurate detection of plant diseases and pests is critical to preventing yield and quality losses, supporting sustainable agriculture, and ensuring food security. In this study, a novel dataset of 6,081 field images showing disease and pest symptoms on apricot (*Prunus armeniaca*) plants was created. Three pre-trained convolutional neural networks (CNNs), namely AlexNet, GoogLeNet, and ResNet-50, were fine-tuned for the classification task. Instead of a standard labeling strategy, a detailed labeling method was proposed, which considers both symptom type and the affected plant organ. The CNNs were trained on two datasets: a traditional 7-class version and a 13-class version generated using the proposed method. All models were evaluated using 5-fold cross-validation. Among all model and dataset combinations, the highest accuracy of 93.9% was achieved by the ResNet-50 model on the 7-class dataset. Although the proposed labeling method resulted in a slight decrease in classification accuracy, the performance difference remained small even with more classes. These findings indicate that the method is dependable and suitable for practical applications.

Key Words: Apricot (*Prunus armeniaca*), Plant disease and pest detection, Convolutional neural networks (CNNs), Transfer learning, Detailed labeling

İnce Ayarlanmış CNN'ler ve Belirti-Organ Düzeyinde Etiketleme ile Arazi Görüntülerinden Kayısı Bitkisi Hastalık ve Zararlılarının Tespiti

Özet

Bitki hastalık ve zararlılarının erken ve doğru tespiti, verim ve kalite kayıplarının önlenmesi, sürdürülebilir tarımın desteklenmesi ve gıda güvenliğinin sağlanması açısından kritik öneme sahiptir. Bu çalışmada, kayısı (*Prunus armeniaca*) bitkilerinde hastalık ve zararlı semptomlarını gösteren 6.081 adet arazi görüntüsünden oluşan özgün bir veri seti oluşturulmuştur. Ön-eğitimli evrişimsel sinir ağı (CNN) modelleri AlexNet, GoogLeNet ve ResNet-50 sınıflandırma görevi için ince-ayarlanmıştır. Standart bir veri etiketleme stratejisi yerine, hem belirti türünü hem de etkilenen bitki organını dikkate alan ayrıntılı bir etiketleme yöntemi önerilmiştir. CNN modelleri, biri geleneksel 7-sınıflı, diğeri ise önerilen yöntemle oluşturulan 13-sınıflı olmak üzere iki ayrı veri seti üzerinde eğitilmiştir. Tüm modeller beş-katlı çapraz-doğrulama yöntemi ile değerlendirilmiştir. Tüm model ve veri seti kombinasyonları arasında en yüksek doğruluk oranı olan %93,9'a, ResNet-50 modelinin 7-sınıflı veri seti üzerinde elde ettiği sonuçla ulaşılmıştır. Önerilen etiketleme yöntemi sınıflandırma doğruluğunda küçük bir düşüşe neden olsa da, sınıf sayısının artmasına rağmen performans farkı düşük kalmıştır. Bu bulgular, yöntemin güvenilir olduğunu ve pratik uygulamalar için uygunluğunu göstermektedir.

Anahtar kelimeler: Kayısı (*Prunus armeniaca*), Bitki hastalık ve zararlı tespiti, Evrişimsel sinir ağları (CNN), Transfer öğrenme, Ayrıntılı etiketleme.

Introduction

Plant diseases and pests cause significant yield and quality losses in agricultural production. These losses vary depending on the spread and stage of the diseases and the population of the pests, and can reach up to 100%, even resulting in plant deaths. Furthermore, due to the damage they inflict on various plant organs, they can negatively impact the productivity of subsequent years, not just the current year's yield. Preventing these losses and increasing the success of agricultural pest and disease management can be achieved by accurately diagnosing diseases in their early stages or before the pest population exceeds the economic threshold for damage (Mohanty et al., 2016).

Diagnosing plant diseases and pests is a task that requires both time and expertise. The symptoms may vary depending on the plant's phenological stage and environmental conditions. This further complicates the diagnosis of plant diseases and pests. Diagnoses made by plant protection experts are not always accurate. This situation can lead to delayed or incorrect treatments, which may result in more severe crop damage. Due to these challenges, artificial intelligence (AI) solutions have become a necessity to support fast and accurate diagnoses. AI-powered plant disease and pest diagnosis systems support sustainable agriculture and contribute to food security in the long term.

In this study, the aim is to detect diseases and pests on apricot (*Prunus armeniaca*) plants using deep learning methods with images obtained under field conditions. In this regard, a novel image dataset has been created, which includes disease symptoms at different stages, the levels of damage caused by pests, and the effects on various plant organs. As prediction models, commonly used pre-trained CNN architectures such as AlexNet (Krizhevsky et al., 2017), GoogLeNet (Szegedy et al., 2015), and ResNet-50 (He et al., 2016) were fine-tuned and adapted to the problem. Additionally, an alternative labeling strategy based on combinations of diseases/pests and plant organs has been proposed, and the effects of this detailed labeling structure on classification performance have been evaluated.

In recent years, the number of deep learning-based approaches for detecting plant diseases and pests has been increasing (Ahmad et al., 2023; Liu and Wang, 2021; Shoaib et al., 2025; Tugrul et al., 2022). Deep learning-based models, which are widely used in this field, exhibit superior performance in image classification and object detection tasks. In particular, adapting CNN-based architectures to different datasets with transfer learning enables working with limited examples and increases the overall performance of the model (Altuntaş et al., 2019; Turkoglu et al., 2022; Yao et al., 2024).

Ferentinos (2018) classified various plant diseases using different CNN architectures and achieved very successful results on the PlantVillage dataset. Similarly, Mohanty et al. (2016) demonstrated the advantages of transfer learning by retraining pre-trained networks instead of training CNN models directly.

Turkoglu et al. (2020) proposed a CNN model for apricot disease detection. They examined how different convolution filter sizes affected classification performance. The highest accuracy, 98.20%, was achieved using a 9×9 kernel. Falaschetti et al. (2022) developed a lightweight CNN-based plant disease detection system running on a low-cost, low-power embedded device. The system was tested on both binary and multi-class classification tasks. It achieved accuracy rates of 98.10% and 95.24%, respectively. Yao et al. (2024) utilized data augmentation and transfer learning in their work on tea leaf blight detection. They particularly stressed that under conditions of limited samples, relying solely on data augmentation might prove inadequate, and that transfer learning offers notable advantages. Shafik et al. (2024) introduced a pair of new models for detection plant diseases. They put to use fine-tuned CNN models for extracting deep features. Their trials, carried out on PlantVillage dataset, uncovered that both the early fusion and ensemble learning models demonstrated strong predictive performance, achieving 96.74% and 97.79% accuracy rates. In their study, Ashurov et al. (2024) introduce a modified depthwise CNN architecture that incorporates squeeze-and-excitation blocks along with improved residual skip connections for plant disease detection. According to the authors, the model achieves 98% accuracy across diverse plant species and disease types. Perumal et al. (2024) proposed an CNN-based approach for detecting plant diseases. They optimized model parameters and used visual interpretation techniques to better understand the network's decisions. The method achieved strong classification performance and showed promise for real-time use in agriculture.

Within the scope of transfer learning, fine-tuned CNN models have been successfully applied on different datasets. However, in addition to using these models directly for classification, their use as feature extractors has also found a response in the literature. For example, Too et al. (2019) combined deep features with traditional methods by classifying the intermediate layer features extracted from various CNN architectures with Support Vector Machines (SVM). Altuntaş and Kocamaz, (2021) combined the deep features obtained from 3 CNN models and classified them with SVM and managed to detect tomato diseases and pests with high performance.

However, a significant portion of existing studies has been conducted with limited datasets and mostly under controlled laboratory conditions (Moupojou et al., 2023). This study aims to analyze apricot diseases and pests in a more realistic scenario using a unique dataset created from images obtained under field conditions.

In this context, the study makes the following original contributions: (i) A novel image dataset of apricot diseases and pests obtained under field conditions has been created. (ii) The impact of detailed class labels based on disease/pest and plant organ combinations on classification performance has been investigated. (iii) AlexNet, GoogLeNet, and ResNet-50 CNN models have been fine-tuned and adapted to the apricot plant disease and pest detection problem, and their performance has been evaluated.

The remaining sections of this paper are organized as follows. The technical details of the proposed model are presented in the *Materials and Methods* section. The experimental results are reported in the *Results* section. Discussion, general conclusions and recommendations for future work are provided in the *Discussion and Conclusion* section.

Materials and Methods

Dataset collection

The study was conducted during the 2021–2022 period through fieldwork in producer orchards located in the Malatya, Kayseri, and Aksaray provinces of Türkiye, primarily in the trial and

production orchards of the Apricot Research Institute Directorate. The objective was to create a unique image dataset for use in disease and pest detection studies. Healthy and disease or pest infected apricot orchards were surveyed, and images of visible disease symptoms and pest damage on various plant organs were collected.
Images were captured under natural field conditions using digital cameras and smartphones of different brands and models, resulting in variations in lighting, framing, and distance. Care was taken to ensure that disease symptoms and pest damage were clearly visible and, where possible, centered in the frame. All collected images were labeled by a subject expert through visual diagnosis.
The dataset comprises images showing disease symptoms and pest damage on apricot leaves, fruits, shoots, branches, and stones. It includes examples from different developmental stages of three diseases (shot hole, monilia, and sharka) and three pest types (aphids, plum scale, and leaf blister mite). In total, the dataset contains 6,081 images. The detailed distribution of images across diseases, pests, plant organs, and classes is provided in Table 1, offering an overview of the dataset's structure and balance.

Table 1. Detailed information of the dataset, including class and organ-specific image counts
Çizelge 1. Sınıf ve organ bazlı görüntü sayılarını içeren veri setine ait ayrıntılı bilgiler

Class Name	Scientific Name	Leaf	Fruit	Shoot / Branch	Stone	Total Images
Aphid	*Myzus persicae*	420	0	0	0	420
Healthy	-	449	665	86	204	1,404
Leaf Blister Mite	*Acalitus phloeocoptes*	515	0	0	0	515
Monilia	*Monilinia laxa*	0	0	336	0	336
Plum Scale	*Parthenolecanium corni*	0	0	739	0	739
Sharka	*Plum pox virus (PPV)*	398	375	0	143	916
Shot Hole	*Wilsonomyces carpophilus*	974	777	0	0	1,751
Total Images		2,756	1,817	1,161	347	6,081

The leaf images consist of samples of shot hole and sharka diseases, as well as aphid and leaf blister mite pests, along with healthy ones. Shot hole disease presents as round, red lesions surrounded by a light-colored halo on young leaves. These lesions gradually turn brownish-centered, reddish-brown spots, and after 5-10 days, they fall off, creating holes in the leaf. Sharka disease causes scattered line and halo-shaped symptoms around the secondary veins of the leaves. Aphids cause the leaves they feed on to curl and form red spots. Leaf blister mites form round or oval horn-like galls on the undersurface of the leaves. The gall tissue on the underside of the leaf causes irregular, raised, yellowish-green to reddish spots on the upper side (Tarım ve Orman Bakanlığı, 2022). Example images of the relevant classes are presented in Figure 1.
The fruit images consist of samples of shot hole and sharka diseases, along with healthy ones. Shot hole disease creates depressions on the fruit surface, accompanied by lesions with a light-colored halo around them. The symptoms of sharka disease on fruits are bright yellow rings or deep wounds extending to the stone (Tarım ve Orman Bakanlığı, 2022). Example fruit images from the dataset are presented in Figure 2.
Shoot and branch images include samples of monilia disease and plum scale pest, along with healthy individuals. Shoots infected with monilia disease turn brown; thin shoots dry completely, while thicker ones develop sunken, elliptical or elongated cracks. The plum scale pest forms visible colonies on the trunks and thicker branches (Tarım ve Orman Bakanlığı, 2022). Example shoot/branch images from the dataset are presented in Figure 3.
Stone images consist of samples of sharka disease, along with healthy ones. Dark spots surrounded by yellow or cream-colored rings on the stone are typical symptoms of sharka disease (Tarım ve Orman Bakanlığı, 2022). Healthy stones have smooth and uniform surfaces. Example stone images from the dataset are presented in Figure 4.

Data preprocessing
Among the pre-trained CNN models used in this study, AlexNet requires input images of 227×227 pixels, whereas GoogLeNet and ResNet-50 require 224×224 pixels. Therefore, the original images in the dataset, which had varying resolutions, were first cropped to a square format while preserving their aspect ratio. During cropping, equal portions were removed from both ends of the longer side, based on the difference in length between the two dimensions. Subsequently, the cropped images were resized to meet the input requirements of the respective models.

Figure 1. Images of sample leaves from the dataset, visually depicting symptoms and damage caused by aphids (top-left), leaf blister mites (top-right), shot hole (bottom-left), and sharka (bottom-middle). A healthy one is also presented (bottom-right).

Şekil 1. Yaprak biti (sol-üst), yaprak uyuzu (sağ-üst), çil (sol-alt) ve şarka (orta-alt) hastalık ve zararlılarının sebep olduğu simptom ve belirtileri gösteren veri setine ait örnek yaprak görüntüleri. Sağlıklı bir yaprak da sunulmuştur (sağ-alt).

Figure 2. Sample fruit images. These display symptoms of shot hole (left) and sharka (middle) diseases, alongside a healthy one (right).

Şekil 2. Örnek meyve görüntüleri. Bu görsellerde çil (sol) ve şarka (orta) hastalıklarına ait simptomlar ile birlikte sağlıklı bir meyve (sağ) gösterilmektedir.

Figure 3. Shoot and branch images illustrating symptoms of monilia disease (left) and plum scale pest infestation (middle), plus a healthy sample (right).

Şekil 3. Monilya hastalığına ait semptomları (sol) ve erik koşnili zararlısı belirtilerini (orta) gösteren sürgün ve dal görüntülerinin yanı sıra sağlıklı bir örnek (sağ).

Figure 4. Apricot stone images. Symptoms of sharka disease are visible on the left, shown with a healthy stone on the right.

Şekil 4. Kayısı çekirdeği görüntüleri. Şarka hastalığına ait simptomlar sol tarafta görülmekte, sağda ise sağlıklı bir çekirdek gösterilmektedir.

CNN Architectures and fine-tuning approach

Convolutional Neural Network (CNN) architectures, which demonstrate superior performance in solving computer vision problems such as image classification, object recognition, and object tracking, differ from traditional neural networks by incorporating convolutional blocks, where features are discovered, and pooling layers, which reduce the dimensionality of the data (Lecun et al., 2015).

In CNN architectures, feature extraction is performed through convolution, activation functions, and pooling layers. This process enables the extraction of meaningful features from raw data. In the convolutional layer, the image is processed using specific filters, resulting in visual features such as horizontal, vertical, and angular edges, as well as smoothed and sharpened versions of the image. The pooling operation, on the other hand, reduces the image size using functions like average or maximum value pooling. This process increases the computational efficiency of the network, thereby improving overall performance.

The extracted features are then transformed into vector form and classified through the fully connected layer. In the final stage, the error value calculated at the network's output is used in the backpropagation algorithm to update the convolutional filters and the weights of the fully connected layer. Through this process, the model gains the ability to make more accurate predictions during the learning process.

CNN models require large amounts of data as they automatically learn the representation of the data (He et al., 2019). Additionally, training these models requires high computational costs (Kornblith et al., 2019). To overcome these disadvantages, transfer learning techniques have been developed. Transfer learning is a machine learning method in which the knowledge gained by solving one problem is reused as a starting point to solve a different problem. Pre-trained CNN models are used as starting points for solving different problems, allowing high classification performance to be achieved on smaller datasets with lower computational costs (Altuntaş and Kocamaz, 2021; Kornblith et al., 2019).

A brief overview of the CNNs employed in this study is presented below: AlexNet considered one of the foundational architectures in deep learning. It consists of a total of eight layers, including five convolutional layers for feature extraction and three fully connected layers for classification. Its relatively simple structure laid the groundwork for later CNN advancements.

GoogLeNet introduced the innovative Inception module, which enables the network to capture multi-scale features within a single layer by combining multiple convolutional operations. Although it comprises 22 layers, GoogLeNet maintains efficiency through the strategic use of 1×1 convolutions for dimensionality reduction.

ResNet-50 is a deep residual network with 50 layers, specifically designed to overcome the vanishing gradient problem often encountered in very deep architectures. By employing residual connections, it facilitates more effective gradient flow and faster convergence. ResNet-50 is widely adopted in transfer learning due to its balance of depth, performance, and computational efficiency.

Data labeling approach

The image data obtained in this study were diagnosed by the subject matter expert within our research team. During the data labeling phase, each image was labeled according to the disease symptoms or pest indications it showed. Disease and pest names were used as class labels, resulting in a dataset with 7 classes: 3 diseases, 3 pests, and 1 healthy class.

Diseases and pests can cause symptoms and damage on multiple plant organs. These symptoms and damages may visually differ significantly depending on the plant organ. To ensure that the dataset aligns with real-world conditions, images were collected from different plant organs. Therefore, each image was categorized based on which plant organ the primary feature belonged to. The dataset consists of images obtained from 4 main plant organs: leaf, fruit, shoot/branch, and stone.

Since some diseases and pests cause symptoms and damage on multiple plant organs, visual differences appear between these symptoms and damages. This results in high intra-class variance. High intra-class variance can make it challenging for the model to make accurate predictions (Wei et al., 2022). Therefore, to improve the classification performance of the deep learning-based prediction models (fine-tuned CNN), a second data labeling approach was proposed with detailed data labeling. In the first dataset, 7 classes were created using disease and pest names, while in the second dataset, 13 classes were formed by combining disease and pest names with plant organs (disease-pest x plant organ). This data labeling strategy aims to increase inter-class variance while reducing intra-class variance.

Experimental setup

To measure the classification performance of the CNN models used in the study, the k-fold cross-validation procedure was applied. In the hold-out validation procedure, performance evaluation based on a single training-test split can be either fortunate or unfortunate. k-fold cross-validation reduces the variance caused by data splitting by including all the samples in the test set once. In addition, results may vary even on the same data split due to factors such as dropout layers in CNN architectures, random weight initialization, or the order of mini-batches. Evaluating the models with different folds creates random training conditions, allowing the changes resulting from the stochastic nature of deep learning to be reflected in the average. The results are presented as mean and standard deviation values. In this way, a more reliable performance evaluation can be made. In this study, the number of folds was set to 5, considering the number of samples and their class distribution.

An evaluation framework is proposed to compare the predictive performance of CNN models and to fairly evaluate the effects under different data labeling approaches. This evaluation framework ensures that all models are trained with the same training, validation and test samples on different data labeling approaches. In this framework, the 13-class labeled dataset is divided into 5 folds. In order to prevent underrepresented classes from being left out by chance, the partitioning process is carried out by taking into account the class ratios in each fold. Then, each fold is divided into two subsets, 75% of the samples in the fold are for training and 25% of the samples in the fold are for validation. First, the entire first fold was used as the test set, while the training subsets of the remaining four folds were combined to form the training set, and their validation subsets were merged to form the validation set. The training process was then carried out using these sets. Then, the entire second fold was used as the test set, while the training subsets of the remaining four folds were merged to create the training set, and their validation subsets were combined to form the validation set. The training process was then carried out using these sets. This process is repeated five times, each time using a different fold for testing. Thus, it is guaranteed that each model is trained with the same training, validation and test sets.

After the training of all CNN models used in the study was completed on the 13-class labeled dataset, the five folds created for the 13-class dataset were converted into folds for the 7-class version. The detailed class labels within each fold were mapped back to their original categories. For example, SharkaXFruit, SharkaXLeaf, and SharkaXStone classes were converted to Sharka class. The same process was applied to other detailed classes. Model trainings were repeated for each fold as explained above. Thus, not only CNN models but also data labeling approaches were partitioned with the same training, validation and test sets.

Training models for a fixed number of epochs may lead to overfitting or, conversely, underfitting. Therefore, an early stopping procedure was adopted as a regularization technique. Models were evaluated on the validation set in each epoch. Training was stopped if no improvement in validation accuracy was observed during five consecutive epochs.

Evaluation metrics

In this study, accuracy (Acc.), precision (Pre.), recall (also known as sensitivity), and F1 score performance metrics were used to evaluate and compare the classification performances of the fine-tuned CNN models. These metrics are calculated using the values of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) obtained from the confusion matrix. The confusion matrix summarizes the number of correct and incorrect predictions made by the model on the test set.

In multi-class classification problems, performance metrics are calculated separately for each class. For the class under evaluation, correct predictions are considered as TP, while incorrect predictions for that class are counted as FN. All other classes are treated as negative. Among these, instances that do not belong to the positive class and are correctly not predicted as such are classified as TN. Conversely, instances from negative classes that are incorrectly predicted as positive are counted as FP.

The mathematical formulas for the performance metrics used in this study are given below.

$$Acc. = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$Pre. = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ score = \frac{2 * TP}{2 * TP + FP + FN} \quad (4)$$

In addition, the overall performance of the model is assessed using the overall accuracy metric, whose formula is also provided below.

$$Overall\ Acc. = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} (TP_i + FP_i)} \quad (5)$$

Here, N refers to the number of classes, while TP_i and FP_i represent the number of correct and incorrect positive predictions for the i^{th} class, respectively.

Results

All procedures in this study were conducted using MATLAB© R2020b. The computer used for the study was equipped with a 2.6 GHz i5 processor, a 512 GB SSD hard drive, 8 GB DDR4 system memory, and a 4 GB graphics card.

This study aimed to accurately detect apricot diseases and pests from images of different plant organs captured under field conditions using fine-tuned CNN models. To support this goal, two different labeling strategies were employed: a conventional 7-class structure and a more detailed 13-class structure that incorporates both disease-pest types and affected plant organs. The dataset was evaluated using stratified 5-fold cross-validation. Each model was evaluated using the same data partitions across folds, ensuring a fair basis for comparison between models and labeling methods. The results are reported as the mean and standard deviation of performance metrics across all folds.

Stochastic Gradient Descent with Momentum (SGDM) was used as the optimization method for fine-tuning the CNN architectures (Murphy, 2012). The hyperparameters were set as follows: The final fully connected layers of the models were replaced with fully connected layers having output sizes of 7 and 13, respectively, to match the number of classes in each dataset. The maximum number of epochs was set to 100, the mini-batch size to 32, and the learning rate to 10^{-4}.

The validation procedure was performed once at the end of each epoch. If no improvement in validation performance was observed for five consecutive validation runs, training was terminated early. Table 2 summarizes the total training time and total number of epochs spent across the five folds, along with the mean and standard deviation of validation accuracy for each model and dataset.

Table 2. Total training time, total number of epochs, and validation accuracy (mean ± standart deviation) obtained from the models after fine-tuning procedure
Çizelge 2. Modellerin ince-ayarlama işlemi sonrasında elde edilen toplam eğitim süresi, toplam eğitim döngüsü sayısı ve doğrulama başarımı (ortalama ± standart sapma)

Model	Dataset	Total Training Time	Total Epochs	Val. Accuracy (%, Mean ± SD)
AlexNet	7-class	3 hr 28 min 59 sec	91	88.28±0.63
	13-class	3 hr 16 min 23 sec	85	87.48±0.81
GoogLeNet	7-class	11 hr 48 min 24 sec	105	91.87±0.60
	13-class	10 hr 48 min 36 sec	102	91.60±0.83
ResNet-50	7-class	36 hr 37 min 22 sec	133	94.35±0.34
	13-class	32 hr 31 min 28 sec	128	93.51±0.64

Table 3. Classification results of the fine-tuned AlexNet model
Çizelge 3. İnce-ayarlanmış AlexNet modeline ait sınıflandırma sonuçları

DataSet	Class Name	TP	FP	FN	TN	Acc. (%)	Pre. (%)	Recall (%)	F1 scr (%)	Overall Acc. (%)
	Aphid	361	30	59	5,631	98.54 ±0.48	92.48 ±4.67	85.95 ±4.64	89.02 ±3.60	
	Healthy	1,211	179	193	4,498	93.88 ±0.55	87.14 ±1.41	86.25 ±1.94	86.68 ±1.24	
	Leaf Blister Mite	423	81	92	5,485	97.16 ±0.31	83.95 ±2.31	82.13 ±1.48	83.03 ±1.76	
7-class	Monilia	304	38	32	5,707	98.85 ±0.16	89.00 ±2.55	90.48 ±3.88	89.66 ±1.48	88.80±0.66
	Plum Scale	705	60	34	5,282	98.45 ±0.29	92.18 ±1.32	95.40 ±2.11	93.75 ±1.25	
	Sharka	775	137	141	5,028	95.43 ±0.32	85.02 ±2.04	84.61 ±1.92	84.79 ±1.05	
	Shot Hole	1,621	156	130	4,174	95.30 ±0.65	91.22 ±1.06	92.58 ±1.53	91.89 ±1.14	
	Aphid	358	38	62	5,623	98.36 ±0.38	90.49 ±3.63	85.24 ±3.53	87.74 ±2.84	
	HealthyXBranch	59	8	27	5,987	99.42 ±0.15	89.76 ±10.50	68.56 ±10.80	76.93 ±6.68	
	HealthyXFruit	592	90	73	5,326	97.32 ±0.41	86.90 ±3.06	89.02 ±2.16	87.91 ±1.76	
	HealthyXLeaf	334	86	115	5,546	96.70 ±0.46	79.69 ±4.23	74.37 ±5.70	76.81 ±3.59	
	HealthyXStone	196	3	8	5,874	99.82 ±0.07	98.56 ±2.14	96.07 ±2.20	97.27 ±1.03	
	Leaf Blister Mite	439	75	76	5,491	97.52 ±0.31	85.56 ±3.54	85.24 ±2.70	85.34 ±1.72	
13-class	Monilia	303	23	33	5,722	99.08 ±0.27	93.00 ±3.16	90.18 ±2.70	91.55 ±2.45	88.18±0.64
	Plum Scale	713	56	26	5,286	98.65 ±0.35	92.73 ±1.15	96.48 ±2.54	94.55 ±1.47	
	SharkaXFruit	318	58	57	5,648	98.11 ±0.54	85.17 ±7.05	84.80 ±3.48	84.79 ±3.63	
	SharkaXLeaf	317	94	81	5,589	97.12 ±0.46	77.18 ±3.97	79.64 ±2.80	78.38 ±3.30	
	SharkaXStone	138	8	5	5,930	99.78 ±0.05	94.65 ±2.62	96.53 ±3.45	95.51 ±0.91	
	Shot HoleXFruit	695	71	82	5,233	97.48 ±0.24	90.85 ±2.33	89.45 ±3.86	90.06 ±1.17	
	Shot HoleXLeaf	900	109	74	4,998	96.99 ±0.42	89.38 ±3.61	92.40 ±2.49	90.79 ±1.02	

Table 4. Classification results of the fine-tuned GoogLeNet model
Çizelge 4. İnce-ayarlanmış GoogLeNet modeline ait sınıflandırma sonuçları

DataSet	Class Name	TP	FP	FN	TN	Acc. (%)	Pre. (%)	Recall (%)	F1 scr (%)	Overall Acc. (%)
7-class	Aphid	388	33	32	5,628	98.93 ±0.33	92.28 ±3.55	92.38 ±3.33	92.27 ±2.36	91.43±0.74
	Healthy	1,245	127	159	4,550	95.30 ±0.51	90.76 ±0.77	88.67 ±2.79	89.68 ±1.29	
	Leaf Blister Mite	456	69	59	5,497	97.89 ±0.52	86.93 ±3.36	88.54 ±4.63	87.67 ±3.16	
	Monilia	307	24	29	5,721	99.13 ±0.30	92.74 ±2.34	91.37 ±3.58	92.03 ±2.84	
	Plum Scale	714	35	25	5,307	99.02 ±0.28	95.35 ±1.15	96.62 ±2.48	95.96 ±1.21	
	Sharka	788	142	128	5,023	95.56 ±0.73	84.88 ±3.79	86.03 ±2.01	85.40 ±2.09	
	Shot Hole	1,662	91	89	4,239	97.04 ±0.54	94.83 ±1.53	94.92 ±0.74	94.87 ±0.92	
13-class	Aphid	391	27	29	5,634	99.08 ±0.34	93.58 ±2.77	93.10 ±3.08	93.32 ±2.45	91.30±0.67
	HealthyXBranch	71	12	15	5,983	99.56 ±0.11	87.34 ±9.90	82.49 ±10.29	83.94 ±4.00	
	HealthyXFruit	620	78	45	5,338	97.98 ±0.34	88.92 ±2.31	93.23 ±4.01	90.95 ±1.64	
	HealthyXLeaf	387	89	62	5,543	97.52 ±0.50	81.49 ±4.73	86.19 ±1.33	83.73 ±2.90	
	HealthyXStone	199	5	5	5,872	99.84 ±0.10	97.58 ±1.65	97.56 ±2.99	97.54 ±1.54	
	Leaf Blister Mite	444	49	71	5,517	98.03 ±0.44	90.23 ±3.87	86.21 ±3.78	88.10 ±2.54	
	Monilia	307	22	29	5,723	99.16 ±0.07	93.34 ±1.18	91.36 ±1.97	92.32 ±0.71	
	Plum Scale	711	36	28	5,306	98.95 ±0.28	95.27 ±2.42	96.21 ±2.73	95.69 ±1.18	
	SharkaXFruit	318	35	57	5,671	98.49 ±0.11	90.17 ±1.76	84.80 ±3.60	87.34 ±1.21	
	SharkaXLeaf	322	72	76	5,611	97.56 ±0.41	82.07 ±5.53	80.91 ±1.53	81.38 ±2.40	
	SharkaXStone	139	6	4	5,932	99.84 ±0.10	95.99 ±3.51	97.24 ±2.89	96.56 ±2.01	
	Shot HoleXFruit	730	38	47	5,266	98.60 ±0.37	95.10 ±2.22	93.95 ±1.96	94.50 ±1.45	
	Shot HoleXLeaf	913	60	61	5,047	98.01 ±0.11	93.92 ±2.23	93.74 ±2.57	93.78 ±0.40	

Table 3 presents the classification outcomes of the fine-tuned AlexNet model on the 7-class and 13-class datasets. Confusion matrix values are shown as totals across all five folds, while the performance metrics are summarized using mean and standard deviation.
Table 4 summarizes the results obtained with the fine-tuned GoogLeNet model for both labeling approaches. The confusion matrices reflect total counts across the five folds, and the performance values are reported as averages with standard deviations.
Table 5 provides the classification performance of the ResNet-50 model using the two labeling schemes. Confusion matrix totals are based on all five folds, and the associated metrics are expressed as mean and standard deviation.

Discussion and Conclusion

This study aimed to accurately detect apricot diseases and pests from images of different plant organs captured under field conditions using fine-tuned CNN models. For this purpose, our research team constructed an original dataset consisting of 6,081 images including healthy samples and samples affected by three diseases and three pests. In addition to the 7-class dataset, a second 13-class dataset was created by combining disease-pest types with plant organ labels through a detailed labeling approach.

Table 5. Classification results of the fine-tuned ResNet-50 model
Çizelge 5. İnce-ayarlanmış ResNet-50 modeline ait sınıflandırma sonuçları

DataSet	Class Name	TP	FP	FN	TN	Acc. (%)	Pre. (%)	Recall (%)	F1 scr (%)	Overall Acc. (%)
7-class	Aphid	402	19	18	5,642	99.39 ±0.44	95.60 ±4.45	95.72 ±2.74	95.63 ±3.15	93.90±0.39
	Healthy	1,312	125	92	4,552	96.43 ±0.43	91.32 ±1.26	93.45 ±1.31	92.36 ±0.94	
	Leaf Blister Mite	463	28	52	5,538	98.68 ±0.25	94.39 ±2.74	89.90 ±2.79	92.05 ±1.54	
	Monilia	304	21	32	5,724	99.13 ±0.29	93.62 ±3.88	90.46 ±2.30	91.99 ±2.60	
	Plum Scale	717	40	22	5,302	98.98 ±0.23	94.76 ±1.93	97.02 ±1.77	95.86 ±0.93	
	Sharka	831	94	85	5,071	97.06 ±0.45	90.02 ±3.74	90.72 ±2.22	90.30 ±1.26	
	Shot Hole	1,681	44	70	4,286	98.12 ±0.24	97.46 ±1.11	96.00 ±0.71	96.72 ±0.40	
13-class	Aphid	403	21	17	5640	99.37 ±0.37	95.17 ±3.84	95.95 ±3.22	95.51 ±2.65	93.27±1.18
	HealthyXBranch	64	13	22	5982	99.42 ±0.19	83.01 ±2.54	74.38 ±14.83	77.89 ±9.08	
	HealthyXFruit	628	69	37	5347	98.25 ±0.60	90.33 ±4.82	94.43 ±2.17	92.26 ±2.49	
	HealthyXLeaf	401	61	48	5571	98.21 ±0.52	86.84 ±4.06	89.30 ±3.47	88.03 ±3.50	
	HealthyXStone	198	6	6	5871	99.80 ±0.13	97.10 ±2.53	97.05 ±2.07	97.06 ±1.84	
	Leaf Blister Mite	465	27	50	5539	98.73 ±0.22	94.62 ±2.57	90.29 ±3.29	92.34 ±1.34	
	Monilia	305	24	31	5721	99.10 ±0.17	92.79 ±2.81	90.77 ±1.97	91.74 ±1.52	
	Plum Scale	711	39	28	5303	98.90 ±0.12	94.82 ±1.11	96.21 ±1.77	95.50 ±0.55	
	SharkaXFruit	338	25	37	5681	98.98 ±0.24	93.13 ±2.08	90.13 ±2.60	91.59 ±1.99	
	SharkaXLeaf	358	42	40	5641	98.65 ±0.27	89.54 ±2.61	89.94 ±2.83	89.71 ±2.10	
	SharkaXStone	137	6	6	5932	99.80 ±0.15	95.79 ±2.94	95.79 ±4.53	95.75 ±3.25	
	Shot HoleXFruit	737	30	40	5274	98.85 ±0.36	96.10 ±1.62	94.85 ±1.98	95.46 ±1.44	
	Shot HoleXLeaf	927	46	47	5061	98.47 ±0.24	95.31 ±1.77	95.17 ±1.18	95.23 ±0.71	

The motivation behind this detailed labeling was to reduce intra-class variance and increase inter-class variance, thereby improving classification performance. Accordingly, the AlexNet, GoogLeNet, and ResNet-50 CNN architectures were fine-tuned and adapted to the problem, and their performances were compared. The results showed that these three models achieved overall accuracy rates of 88.80%, 91.43%, and 93.90%, respectively, on the 7-class dataset. On the 13-class dataset, the models achieved 88.18%, 91.30%, and 93.27% accuracy, respectively. All three models produced highly successful results. A trend was observed where deeper architectures outperformed shallower ones in terms of classification accuracy, suggesting a potential advantage of increased representational capacity. However, when tested on the 13-class labeled dataset created through detailed labeling, each model exhibited a slight, statistically insignificant drop in classification performance. One potential reason for this decline is the increased visibility of class imbalance as the number of classes increased. Furthermore, although the images were labeled based on plant organs, they were captured under real field conditions, often containing multiple plant parts in a single image. The background complexity and the presence of different plant organs together may be other contributing factors to the performance drop.

In future studies, it is planned to create a more balanced dataset by incorporating additional images or applying data augmentation techniques. To improve the classification performance of the detailed 13-class dataset, the use of attention mechanisms or hierarchical classification approaches will be considered. Additionally, hybrid or ensemble learning methods that combine the feature extraction capabilities of different CNN architectures will be explored to further enhance classification success.

Acknowledgment
This study was carried out as part of the completed project titled "*Detection of Diseases and Pests on Apricot Trees Based on Field Images Using Deep Learning Techniques*" (Project No: TAGEM/TSKAD/B/21/A9/P7/5030), funded and supported by the General Directorate of Agricultural Research and Policies (TAGEM) of the Ministry of Agriculture and Forestry of the Republic of Türkiye. We would like to thank TAGEM for their financial support.
We also extend our gratitude to Erciyes University Faculty of Agriculture and the Ortaköy District Directorate of Agriculture and Forestry for their collaboration and valuable contributions during the fieldwork.

Data Availability
The dataset used in this study was created during the completed TAGEM project and is currently being reused and further expanded within the scope of a broader, ongoing research project that includes additional fruit species and vineyards.
At this stage, due to class imbalance and the dataset's active use in ongoing analyses and model development, the data is not publicly available. However, we support open science and plan to make the dataset publicly available once it becomes sufficiently complete and balanced.
Until then, the dataset may be shared with researchers upon reasonable request, provided that the request meets relevant ethical standards and institutional requirements.

References

Ahmad A, Saraswat D, El Gamal A, 2023. A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. Smart Agricultural Technology 3(June 2022), 100083. https://doi.org/10.1016/j.atech.2022.100083

Altuntaş Y, Cömert Z, Kocamaz AF, 2019. Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach. Computers and Electronics in Agriculture 163: 104874. https://doi.org/10.1016/j.compag.2019.104874

Altuntaş Y, Kocamaz AF, 2021. Deep Feature Extraction for Detection of Tomato Plant Diseases and Pests based on Leaf Images. Celal Bayar University Journal of Science 17(2): 145–152. https://doi.org/10.18466/cbayarfbe.812375

Ashurov AY, Al-Gaashani MSAM, Samee NA, Alkanhel R, Atteia G, Abdallah HA, Saleh Ali Muthanna M, 2024. Enhancing plant disease detection through deep learning: a Depthwise CNN with squeeze and excitation integration and residual skip connections. Frontiers in Plant Science 15: 1–16. https://doi.org/10.3389/fpls.2024.1505857

Falaschetti L, Manoni L, Di Leo D, Pau D, Tomaselli, V, Turchetti C, 2022. A CNN-based image detector for plant leaf diseases classification. HardwareX 12: e00363. https://doi.org/10.1016/j.ohx.2022.e00363.

Ferentinos KP, 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

He K, Girshick R, Dollar P, 2019. Rethinking imageNet pre-training. Proceedings of the IEEE International Conference on Computer Vision 2019-Octob(ii): 4917–4926. https://doi.org/10.1109/ICCV.2019.00502

He K, Zhang X, Ren S, Sun J, 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 770–778.

Kornblith S, Shlens J, Le QV, 2019. Do Better ImageNet Models Transfer Better? Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2661–2671.

Krizhevsky A, Sutskever I, Hinton GE, 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM 60(6): 84–90. https://doi.org/10.1145/3065386

Lecun Y, Bengio Y, Hinton G, 2015. Deep learning. Nature 521(7553): 436–444. https://doi.org/10.1038/nature14539

Liu J, Wang X, 2021. Plant diseases and pests detection based on deep learning: a review. Plant Methods 17(1): 1–18. https://doi.org/10.1186/s13007-021-00722-9

Mohanty SP, Hughes DP, Salathé M, 2016. Using deep learning for image-based plant disease detection. Frontiers in Plant Science 7: 1–10. https://doi.org/10.3389/fpls.2016.01419

Moupojou E, Tagne A, Retraint F, Tadonkemwa A, Wilfried D, Tapamo H, Nkenlifack M, 2023. FieldPlant: A Dataset of Field Plant Images for Plant Disease Detection and Classification With Deep Learning. IEEE Access 11: 35398–35410. https://doi.org/10.1109/ACCESS.2023.3263042

Murphy KP, 2012. Machine learning: a probabilistic perspective. The MIT Press.

Perumal VK, Supriyaa T, Santhosh PR, Dhanasekaran S, 2024. CNN based plant disease identification using PYNQ FPGA. Systems and Soft Computing 6: 200088. https://doi.org/10.1016/j.sasc.2024.200088

Shafik W, Tufail A, De Silva Liyanage C, Apong RAAHM, 2024. Using transfer learning-based plant disease classification and detection for sustainable agriculture. BMC Plant Biology 24(1): 1–19. https://doi.org/10.1186/s12870-024-04825-y

Shoaib M, Sadeghi-Niaraki A, Ali F, Hussain I, Khalid S, 2025. Leveraging deep learning for plant disease and pest detection: a comprehensive review and future directions. Frontiers in Plant Science 16: 1–19. https://doi.org/10.3389/fpls.2025.1538163

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A, 2015. Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 1–9.

Tarım ve Orman Bakanlığı, 2022. Kayısı entegre mücadele teknik talimatı [Technical guidelines for the integrated control of apricot pests and diseases] (in Turkish).

Too EC, Yujian L, Njuki S, Yingchun L, 2019. A comparative study of fine-tuning deep learning models for plant disease identification. Computers and Electronics in Agriculture 161: 272–279. https://doi.org/10.1016/j.compag.2018.03.032

Tugrul B, Elfatimi E, Eryigit R, 2022. Convolutional Neural Networks in Detection of Plant Leaf Diseases: A Review. Agriculture (Switzerland) 12(8). https://doi.org/10.3390/agriculture12081192

Turkoglu M, Hanbay K, Sarac Sivrikaya I, Hanbay D, 2020. Derin Evrişimsel Sinir Ağı Kullanılarak Kayısı Hastalıklarının Sınıflandırılması. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 9(1), 334–345. https://doi.org/10.17798/bitlisfen.562101

Turkoglu, M., Yanikoğlu, B., Hanbay, D. (2022). PlantDiseaseNet: convolutional neural network ensemble for plant disease and pest detection. Signal, Image and Video Processing 16(2): 301–309. https://doi.org/10.1007/s11760-021-01909-2

Wei XS, Song YZ, Aodha O Mac, Wu J, Peng Y, Tang J, Yang J, Belongie S, 2022. Fine-Grained Image Analysis With Deep Learning: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(12): 8927–8948. https://doi.org/10.1109/TPAMI.2021.3126648

Yao X, Lin H, Bai D, Zhou H, 2024. A Small Target Tea Leaf Disease Detection Model Combined with Transfer Learning. Forests 15(4). https://doi.org/10.3390/f15040591