Computer and Decision Making

An International Journal

www.comdem.org
eISSN: 3008-1416

CNN Models Approaches for Robust Classification of Apple Diseases

Ismail Kunduracioglu [1,*]

[1] Computer Engineering, Faculty of Engineering, Igdir University, 76000, Igdir, Turkey

ARTICLE INFO

Article history:
Received 28 September 2024
Received in revised form 15 October 2024
Accepted 19 October 2024
Available online 19 October 2024

Keywords:

Artificial intelligence; Deep-Learning; CNN; Plant diseases; Apple diseases

ABSTRACT

In this research, the performance of several CNN models for diagnosing diseases in apple leaves is critically assessed. Due to this, early identification of diseases in agriculture is essential to boost productivity and avoid losses. In this context, disease identification was carried out with CNN models including ResNet50 and InceptionV4, Xception, DenseNet121, EfficientNetV2_m, and VGG13. While diagnosing the performance of each model the key parameters like Accuracy, Precision, Recall, and F1 score were computed. The accuracy values obtained show that all the discussed models are very accurate. However, it should be pointed out that the EfficientNetV2_m model is the best one in terms of accuracy and F1 score of 100%. This has shown that the EfficientNetV2_m gives a better detection of diseases in apple leaves compared to other models. The outcomes of the research reveal that in comparison to the conventional machine learning methodologies, the outcomes of the deep learning methods can prove to be much more beneficial for the agricultural sector by presenting a vision of how they can be incorporated and employed for improving the different sectors of management in the agricultural field. Summing up, the present research has shown the efficiency of the deep learning models in detecting diseases on apple leaves and has made a substantial contribution to the development of this line of study. In regards to the recommendations, the authors propose considering increasing the size of the sets and switching to the utilization of more advanced algorithms to enhance the present model. Thus, diseases found in agriculture will be identified easily and may also be more efficient in creating efficiency in the agriculture business.

1. Introduction

Agriculture still gets a lot of things done for the global economy. It makes it easier for various sectors to replenish sufficient food supply to people as well as earn their living [1]. Although the apple plantation is economically important and highly distributed across the globe, it might be even more standing because of the profit earned from the cultivation of other crop species. Nonetheless, new apple production, mainly caused by diseases, threatens the global supply chain of the product. The disease's effect on the fruit production quantity and quality might be very serious [2]. Immediate and precise identification of infection is a top priority for successful treatment. This is also a critical factor in keeping the apple farming sector economically sustainable [3].

E-mail address: ismail.kunduracioglu@igdir.edu.tr
https://www.doi.org/10.59543/comdem.v1i.10957

The Convolutional Neural Networks (CNN) application in agriculture has shown great success in different ways. For example, Zhang et al. created a deep CNN architecture that was state-of-the-art in recognizing diseases in maize leaves [4]. Also, CNNs have been used to detect diseases in mango and wheat leaves, thus corroborating their efficacy in disease identification [5], [6], [7]. These achievements demonstrate the capacity of CNNs to bring a major change in disease diagnosis in agriculture, which would be faster, more accurate, and less dependent on human specialists [8]. Historically apple's disease diagnosis has been quite dependent on manual inspections by trained doctors who do all the work themselves, which in turn makes the process both labor-intensive and error-prone [9]. The visual symptoms of the different diseases can often look the same, thus it is difficult to tell the diagnosis correctly, especially in field conditions. Moreover, the medical field has managed to push the use of automated methods to the forefront because of their better performances and reliability. Particularly, deep learning and image recognition technologies have experienced some considerable advancements. CNN is a type of deep learning model that has been effectively used in plant disease classification showing that they can identify diseases equally well as humans from image data [4], [5].

This research specifically aims to reveal the classification of apple leaf diseases by using CNN architectures. Studies conducted in the past times have the same magnitude of importance. Hence, the correct disease diagnosis is the real solution to minimizing agricultural losses and promoting economic growth [10]. Web-based learning methods help not only the automation of the procedure of diagnosis but also provide the effectiveness of the system by revealing some difficult patterns that conventional techniques may overlook [11]. Machine learning methods have been applied to the classification of many agricultural items, including apples, by using the image data that has been analyzed [12]. The smarter tech growing in agriculture, e.g. IoT and 5G networks, is endowing real-time data processing and decision-making [13], [14] . These technologies make it thereby the continuous monitoring of crops possible enabling timely interventions and efficient disease management [10], [15]. The integration of these CNNs with these technologies could highly improve accuracy in apple disease diagnosis, thus, diminishing the use of pesticides and downtime to the environment [10]. This research work tries to bring new results into this branch of science by creating a solid CNN model for identifying apple leaf diseases. Using a rich dataset of apples, the ultimate goal of the model is to properly classify the disease even if the photos are affected by conditions such as uneven illumination, occlusions, and music video backgrounds [16]. The results of this project can be a fortune for the apple sector by laying down a first-rate device for disease detection that is early enough and contributes to the sustainability of apple production.

In conclusion, implementing deep learning technologies in agriculture worldwide, especially for disease detection, reveals the enormous potential for the sector to advance. This analysis is not only concerned with apple disease diagnostics which are more accurate and effective but also the implications of this technology for other crops and agricultural practices as a whole. The outcomes can result in the development of agricultural disease management strategies that will be the reason behind crop yield increase, economic losses decrease, and the sustainability of farming practices establishment [17].

1.1 Literature review

The latest improvements in deep learning technology are responsible for the greater accuracy and tighter apple disease classification. A study from 2019 called LeNet-5-based CNN input the data containing specific types of apple tree diseases and convincing or healthy apple leaves and got a

98.54% accuracy rate on the Black Rot, Rust, Apple Scab, and healthy leaf sets [18]. In the same year, another study proposed the VGG-INCEP model which utilized the structures of VGG and Inception in a random way to detect for example ATLDs like Mosaic, Rust, Grey Spot, Brown Spot, and Alternaria Leaf Spot. The model was 97.14% accurate while it also had a real-time detection model with an average mean accuracy of 78.80% [19]. Image classification has been moving from advanced deep learning techniques. In the early stages, image classification was performed by humans through the process of manual feature extraction in addition to classifiers such as Naive Bayes, Support Vector Machines (SVMs), and K Nearest Neighbours (KNN) [8]. These methods established the groundwork, but their dependence on manual feature extraction limited their accuracy and efficiency. The emergence of deep learning represented a significant revolution, bringing in the automation of both feature extraction and classification. Nowadays, CNNs have been accepted as the leading architecture because of their competence in learning to derive hierarchical feature representations directly from data [20]. Earlier CNN models like LeNet-5 revealed the strength of deep learning in the field of image classification with its excellent performance in handwritten digit recognition [21]. The introduction of AlexNet in 2012 represented a breakthrough, showcasing the effectiveness of deep CNNs in large-scale image classification competitions like ILSVRC 2012.

Recent investments in deep learning have practically bolstered the apple disease classification system in both aspects of precision and speed. In 2019, a model named LeNet-5-based CNN made by Baranwal et al. was able to recognize three types of apple tree leaf diseases (ATLDs) and healthy apple leaves with 98.54% accuracy on a dataset with Black Rot, Rust, Apple Scab, and healthy leaves [18]. The same year, Jiang et al. presented the VGG-INCEP model which employed a VGG and Inception structures pipeline to randomly find ATLDs like Mosaic, Rust, Grey Spot, Brown Spot, and Alternaria Leaf Spot. The performance of the model was 97.14% and it was the real-time detection model which had a mean average accuracy of 78.80% [19]. Image classification has been moving from the advanced deep learning techniques [22]. Later works have taken the foundations of these studies even further. Sladojevic et al. used CaffeNet to classify 13 types of leaf diseases and got an accuracy of 96.3% [7]. Jiang et al. introduced the INAR-SSD method for real-time diagnosis and detection of five common apple leaf diseases, achieving 78.80% mean average precision and 23.13 frames per second [19]. Zhang et al. constructed a global pooling dilated CNN to diagnose six types of cucumber leaf diseases, with an accuracy of 94.65%, thus, outperforming deep CNNs based on traditional methods [23]. Innovative data augmentation techniques have solved the problem regarding the small size and diversity of datasets [24]. Conventional approaches, including rotation, mirroring, translation, and scale transformation techniques, have been illustrated to generate datasets. However, these methods often fail to alter the superpixel information and retain similarities in color, brightness, and texture [25].

In summary, the evolution from traditional image classification methods to advanced deep learning techniques, particularly CNNs, has greatly enhanced the accuracy and efficiency of apple disease classification. Recent research has demonstrated notable improvements. Kim et al. introduced a CNN-based superpixel classification method for apple diseases, achieving 92.43% accuracy and an F1 score of 0.93 [26]. Verma et al. applied deep learning techniques to various crops, including tomatoes and potatoes, achieving an accuracy of 96.24% [27]. Khan et al. used transfer learning with CNNs, reaching a high accuracy of 97.18% for apple disease classification [3]. In 2020, Zhong and Zhao proposed three different loss functions for DenseNet-121, which demonstrated accuracy rates of 93.51%, 93.31%, and 93.71% on a dataset containing various apple diseases and healthy leaves [28]. These results surpassed those of the cross-entropy loss function. Additionally, Yu and Son. introduced a region of interest-based DCNN, which achieved an accuracy of 84.3% on 404 images of Brown Spot, Alternaria Leaf Spot, and healthy leaves [29]. Agarwal et al. employed a VGG16-based CNN to achieve a validation accuracy of 93.3% [30]. Savla et al. enhanced ResNet18 models to achieve accuracies of 95.2% and 97.2% [31]. Additionally, hybrid models and novel architectures have been explored. For instance, the VGG-INCEP model achieved a mean average accuracy of 78.80% for real-time disease detection [19]. DenseNet-121-based approaches with various loss functions achieved accuracy rates between 93.51% and 93.71% [28]. Other innovations include region of interest-based DCNNs and feature extraction combined with optimization algorithms [29], [32]. These advancements provide valuable tools for modern agriculture.

2. Methodology

2.1 Deep Learning

Deep learning, which resembles a highly developed version of traditional neural networks such as the human brain, particularly the neocortex that processes brain signals through a hierarchical arrangement over time [33], has ancient roots in traditional neural networks. Such inspiration saw the birth of models implementing controls based on the information hierarchy that can allow such deep networks to hand off data at high-level abstractions, automatically, and with the help of complex neural network architectures [34]. Moreover, on the other hand, deep learning models are not like the traditional neural networks which have only one hidden layer. In support of this is the fact that the detachability of the data representations and the solving of complex problems are enhanced by having multiple hidden layers [35].

CNNs are the most popular and appropriate architectures for deep learning, which are tailored for processing only 2D data like images. CNNs employ techniques of convolutional layers, pooling layers, and activating functions such as ReLU. Obtaining data of a lower dimension and extracting the features necessary for efficient performance are also included [36]. The networks that depend on the various layers combination have been critical in the tasks of handwritten digit recognition, image classification, and detection of diseases. This is because they are able to achieve high accuracy in classification problems [35]. CNN architecture combines models like AlexNet and SqueezeNet, thus, data representation is optimized by the spatial relationship in the input data consequently reducing the size of the parameter to be learned and enhancing the training [37], [38]. This ability to automatically extract relevant features without manual intervention underscores the transformative impact of deep learning in fields like computer vision and medical diagnostics [34], [39].

2.2 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) are really powerful deep learning algorithms that are superb for the fields of image and video processing. CNNs differ from a conventional artificial neural network in the sense that they learn the data spatial hierarchies, which in turn, gives them strength for working on large datasets [1]. CNNs have three main types of layers: convolutional, pooling, and fully connected. The convolutional layers introduce the input data to the filters so that they can extract the prominent features. These filters usually target low-level features such as edges, corners, and textures. The pooling layers lower the dimensionality of the data while still keeping the key features, thus, decreasing the computational budget. Techniques like max-pooling take this dimensionality reduction technique by choosing the maximum value in each filter's output. Slowly but surely, the fully connected laid out the classification of the data through the use of features that the network has learned. These layers correspond to the full connections that are made between all the neurons and are used in the classical neural networks. They are the ones that determine the category to which the input data belongs.

One of the major strengths of CNN is that feature engineering is no longer necessary. These, therefore, gained the best features by means of which they learned the desired outcome with very high accuracy, particularly in tasks like image classification, object detection, and facial recognition. Some deep learning models, such as AlexNet and VGGNet, have greatly aided in the adoption of deep learning technology. Ever since doubling down on image recognition competitions, CNNs have become the go-to technology in the fields of computer vision and medical imaging.

2.2.1 ResNet

The Residual Network (ResNet) is a qualitative leap in one of the most fundamental problems of deep learning, teaching to train very deep neural networks. The original presentation of ResNet by [40] provided a solution to the issue of vanishing gradients which is one of the common problems faced when training deep networks. ResNet's main creativity is found in dimensioned learning via residual learning through “skip connections” or “shortcut connections.” The output of a layer is added to the output of a deeper layer, effectively bypassing one or more layers.

In this approach, the network can be taught functions of a residual nature concerned with the layer inputs, instead of just the unreferenced functions learning. The training of deeper networks is not only facilitated by stabilizing but it is also possible to create models consisting of hundreds or thousands of layers that have a relatively simple structure. ResNet-50 is a deeper version of ResNet, which contains 50 layers in total. Besides that, the architecture is based on complicated residual blocks, which have an extra convolutional layer in each block. The extra depth given by this depth provides ResNet-50 with the capability to obtain more detailed representations, hence, the performance of ResNet-50 in the high accuracy tasks is better. Yet, its greater depth with the use of residual connections ensures that ResNet-50 can be trained easily without the issues usually associated with very deep networks like vanishing gradients. ResNet-50 has become a natural model for many computer vision activities from object detection to segmentation and many more.

The introduction of ResNet and its variants like ResNet-34 and ResNet-50 has had a profound impact on the development of deep learning models, setting new standards for both accuracy and efficiency in neural network design.

2.2.2 Inception

Inception networks are one of the major innovations brought by CNN architectures, introduced by Szegedy et al. which are well-known for their new method of feature extraction that is based on the use of multiple convolutional filter sizes and pooling operations combined in a single layer [41]. The idea of such a design is that the network can gather various kinds of features at multiple scales of the image at once, which would allow it to recognize tough image patterns accordingly.

In the case of the Inception model, the insufficiency of single-sized filter convolution is overcome through the use of “Inception modules.” Sequences of convolutions and pooling operations executed in parallel (1x1, 3x3, and 5x5) are examples of the Inception modules. From this plan, the total scope of the data input is covered and then the appropriate modules are summed up. This puts the network in a position to respond accordingly to input image features, thus it reduces the image processing requirements.

Inception v4 is a cutting-edge convolutional neural network architecture developed by Google, taking inspiration from its predecessors Inception v1, v2, and v3 while also incorporating the efficiency found in ResNet. This innovative model is defined by its intricate inception blocks, consisting of multiple modules that utilize parallel convolutional layers with different kernel sizes. This design helps the model to have the capacity to capture data at diffident resolution levels which greatly improves feature extraction from images. What sets Inception v4 apart is its depth; with over 55 layers, it can learn complex and abstract features, leading to impressive performance on challenging tasks. The network begins with a stem module which precisely performs down sampling of the input image as well as extraction of the important features. After this, the reduction modules effectively down-sample the spatial dimensions of the feature maps used to lessen the computational intensity. Inception v4 employs three distinct types of inception blocks A, B, and C each designed to process feature maps in unique ways. In an effort to solve the vanishing gradient problem, an auxiliary classifier is implemented, creating another way for gradients to propagate during training. Moreover, the architecture includes connections inspired by ResNet, which facilitate smoother gradient flow and alleviate training challenges often encountered in very deep networks. Overall, Inception v4 showcases remarkable advancements over earlier models, achieving high accuracy across various image classification benchmarks like ImageNet. It has all the aspects of being complex enough to solve difficult image recognition problems but is not computationally too heavy.

2.2.3 Xception

Xception (Extreme Inception) is a CNN architecture that is more advanced than the Inception model. But it includes some more important innovations. It was addressed by Chollet [42]. The key updates to Xception include the use of depthwise separable convolutions which greatly enhances both model performance and efficiency. Depthwise separable convolutions are implemented in two steps: separating the convolution operation into two procedures, namely Depthwise convolution and Pointwise convolution. This is the method that gives a better result than traditional convolutions where the fewer number of parameters and computations is called for. The meaning of depthwise convolution is that a single filter is applied to each input channel. Hence, pointwise convolution is a 1x1 convolution that combines the outputs of depth-weighted convolution. This approach allows Xception to have a more complicated feature set by using fewer parameters, thus it is more efficient and effective in detecting fine patterns from input data.

Compared to its predecessors, such as Inception-v3, Xception provides a more streamlined and efficient approach to feature extraction. Although Inception-v3 utilizes several different-sized convolutional filters in its Inception modules, Xception simplifies the structure by using only depthwise separable convolutions. Therefore, the model generated by this method had less number of parameters and a better computational efficiency which is the reason why the model learned how to be as accurate as other well-known models while being less resource-consuming. The modular design and the efficiency enhancements of the Inception models have already become the new normal for CNNs, and thus their contribution to image recognition and other computer vision applications has been significantly boosted.

2.2.4 DenseNet

In the deep learning field, DenseNet can be referred as an effective method to build the convolutional neural network (CNN) and it is put forward by Huang et al. [43]. This architecture envisages that as depth increases the flow of information and feature sharing should also be improved. Unlike other network architectures, DenseNet is constructed under the idea that each layer shares inputs with all previous layers. This structure also reduces the information loss as the depth of the networks is improved while allowing an improved learning process with less parameters.

This is due to the method called DenseNet which has one unique feature in comparison with others "dense connectivity". Each layer is connected with every preceding layer which improves data transferring and acquisition of multiple-layer patterns by the network. Nevertheless, this architecture often yields the best results when implemented with the least number of parameters since there is less computational strength needed. For these reasons, DenseNet achieves high outcomes in image classification tasks and other deep learning processes. Of all the components of DenseNet, it is made up of several groups of layers known as "Dense Blocks". There are several convolutional layers in each Dense Block and the connection between the layers provide each layer smooth access to features learned by the subsequent layers. This also makes learning processes efficient and quicker and at the same time avoids over-fitting and general improvement of models. Specifically for different application fields like image classification and object detection, DenseNet reveals interesting performance. Third, another indigenous aspect of DenseNet is that each layer directly interacts with not only the subsequent but also all preceding layers. Such structure enables each layer to take the data being received from successive layers as additional input which makes feature recycling possible. But this may also lead to the reinjection of some relatively unimportant characteristics slightly modified in each reiteration. In order to solve this problem DenseNet121 used the Squeeze-and-Excitation (SE) blocks thus the model pays more attention to details that are stored in the channel features. While Deep Learning Networks seek to become more effective while increasing their depth, DenseNet-121 will make use of limited connections between layers. In each layer, only inputs from other preceding layers are gained and then the layer passes its feature maps to the other layer. Therefore, the 'i' layer gets an input 'I' and holds all the feature maps of the first convolutional blocks. DenseNet has more connections than other traditional deep learning models but it needs fewer parameters than other models and hence it is efficient. Therefore, DenseNet has earned its rightful place among methods discussed in the field of deep learning because of the performances it offers. As it has been used in several applications including image classification and object recognition the structure has a huge implication for deep learning models and studies. The unique feature of DenseNet makes deeper networks improve their learning capacity and hence can solve complex problems.

2.2.5 EfficientNet-v2

EfficientNet-v2 is specifically derived from the EfficientNet series provided by Google to solve more complex problems such as image classification [44]. The new 'model scaling method', the multidimensional compound model scaling was launched when presenting the EfficientNet series in 2019 and has been warmly embraced in academia. This approach aims to achieve a balanced trade-off between speed and accuracy by simultaneously scaling three key dimensions: In this case, there is network depth, network width, and also image resolution. These principles are extended in EfficientNet-v2 with the intent to enhance the model's performance. Models in EfficientNet-v2 series have been described as EfficientNet-v2s, EfficientNet-v2m, EfficientNet-v2l, and EfficientNet-v2xl with every model having different width and depth options based on the specific task at hand. Out of these, we have seen that the EfficientNet-v2s model is the most significant model, which is of small size and high accuracy. With the help of unique architecture and advanced training techniques involved in this model, the identification performance and training modes improve.

Another major change carried out in the present work as compared to other EfficientNet models is presented by the improvements introduced to the network and its training procedure. While EfficientNet-v1 models used a uniform scale, EfficientNet-v2 use a minimal scale with additional scalable parameters. This has brought about overall training time reduction to one fourth of the previous training time and a reduction of parameters to 1/6.8. The MBConv blocks at the later stages of EfficientNet-v2 are different from the Fused-MBConv blocks that are formed for enhancing feature transfer inside the network. Such architectural improvements make EfficientNet-v2 capable of giving better performance on distinct tasks. For instance, in cases with image classification, the training speed has been enhanced by 11 times, while the training parameters are drastically smaller. Further, the training strategy is improved; The EfficientNet-v2 has an improved progressive learning approach in the training procedure, specifically the learning rate dependent on the size of the training image that makes it train faster and offers improvement in accuracy. Therefore, EfficientNet-v2 is generally a more significant improvement over the original EfficientNet series in the perspective of the training efficiency and model parameters. All these improvements put EfficientNet-v2 in a suitable position for research and industrial purposes.

2.2.6 VGG

VGG networks (VGGNet), the mode constructed by the team at Oxford University referred to as the Vision Geometry Group are among the most CNNs [45]. These are designed to eliminate design and optimization complexities and to increase model depth with the help of a very simple structure. The main purpose of these networks is to systematically introduce the depth for improving the performance and to apply identical 3x3 convolutional filters over all layers of convolution. In fact, these small kernels are capable of capturing spatial relationships between pixels while letting the network construct deeper structures. The other major property of VGGNet is that the network is of a regular structure at the layer level. After each two of the convolutional layers, the max–pooling layers are used this helps to down sample the feature map, this will help to reduce computational cost as well as rate of over training. It’s because the model structure is built from containing layers that are repeated, so it becomes possible for the network to learn more intricate features.

VGG-13 is one of the models in the family of VGGNet; they are modelled with thirteen convolutional layers and three fully connected ones. Hence, kernels of size 3×3 with a consistent stacking in each of the convolutional layers help the model to grasp intricate details and spatial correspondences into images and, therefore, to build deeper feature learning for more intricate categorizations of images. The architecture is often formed in succession of a convolution layer, and two layers of a pooling layer whereby only one is a max pooling layer. For example, the first two groups included 64 channels, the next two included 128 channels, and the last two included 256 and 512 channels, correspondingly. In fact, the max-pooling layers decrease the size of the feature maps by 2 times, thus reducing computational requirements during training. This structure is used five times, which greatly extends the depth and learning of the network. Following the convolutional layers, there are three fully connective layers with 4096 neurons in the first two and 250-1024 neurons of a softmax format for the final output. The final layer computes the probability distribution among the predicted classes, and the softmax activation function is used for classification output.

In conclusion, VGG-13 demonstrates strong performance in tasks such as image classification and object recognition. Although deeper variants like VGG-16 and VGG-19 may be more effective at learning complex features, VGG-13 is often preferred due to its lower computational cost and simpler structure.

2.3. Dataset

The PlantVillage dataset, a highly detailed compilation of images intended for the purpose of plant disease classification, has played a key role in our research. Specifically, we utilized the dataset to detect and classify various apple diseases. The apple section of the dataset comprises multiple classes including healthy apples as well as apples with diseases such as apple scab, black rot, and cedar apple rust (Fig.1). The dataset provides each of these classes with a large enough number of images, enabling the training of deep learning models to reach the disease classification accuracy level as shown in (Table 1).

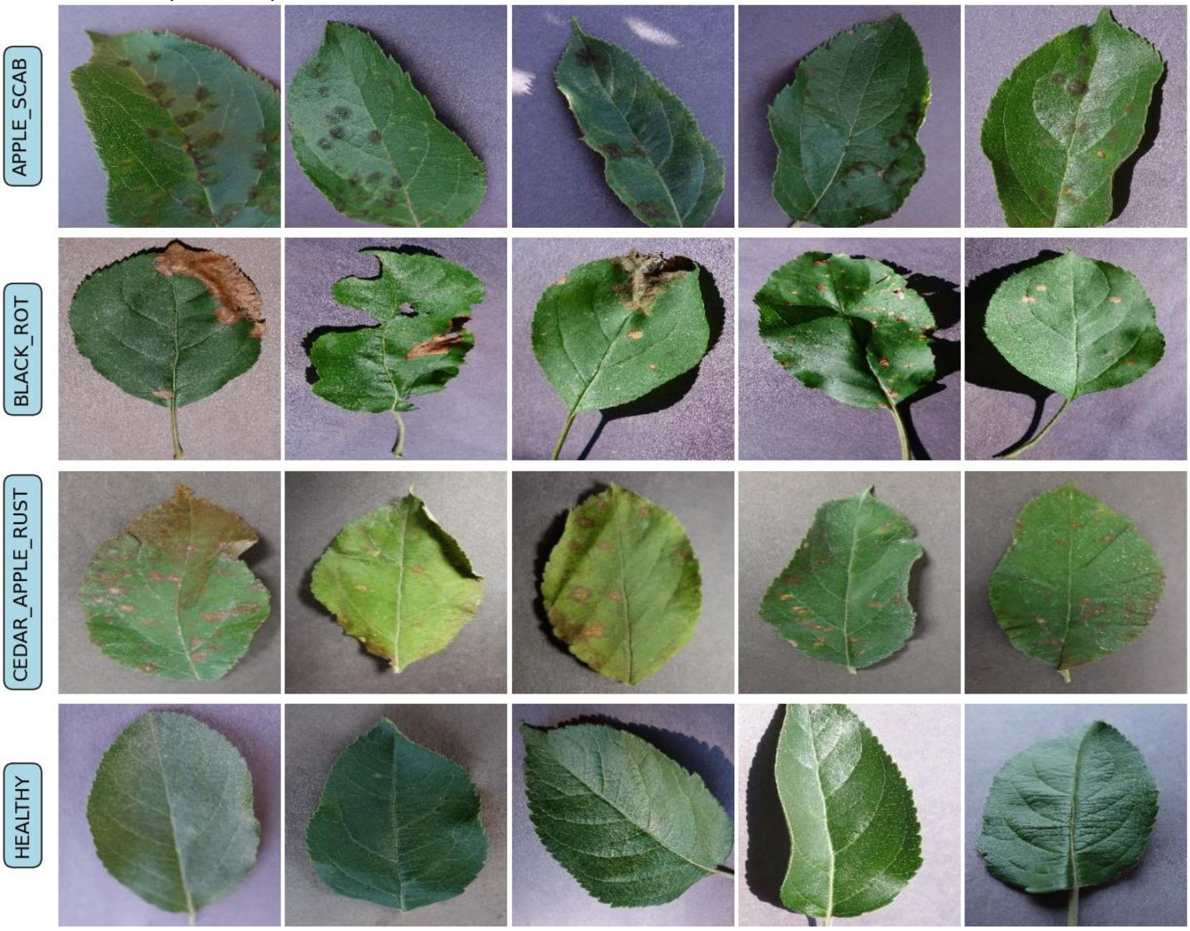

Fig.1 Randomly selected samples of apple leaf images from PlantVillage dataset

Table.1
Apple diseases and their counts

Class Names	Total (100%)	Train (70%)	Validation (15%)	Test (15%)
Apple scab	630	441	94	95
Black rot	621	435	93	93
Cedar apple rust	275	193	41	41
Healthy	1645	1152	247	246

Table 1 shows how the images of diseased apples and healthy apples are distributed in different datasets for training, validation, and testing. The Training (70%), Validation (15%), and Test sets (15%) are added for each category to obtain a balanced evaluation of the model's performance. The large training set helps the model develop a deep understanding, while the balanced validation and test sets allow the evaluation of the model's generalization performance. The database with 3,171 apple images is quite small in size. Since the size is very small, the use of data augmentation techniques becomes of paramount importance in improving the accuracy of the model and making sure that the results are closer to real-world conditions.

2.4 Data preprocessing

Transfer learning is a critical machine-learning technique that helps to use knowledge from one area to improve the performance of a similar but often different domain. This method is clearly favored, especially, in the data scarce cases, e.g. when the data labeling is expensive. In agriculture applications, it is impossible to predict the diseases of tomatoes if the data collection and labeling process will be so difficult that defective images of even a few effective ones will result in overfitting and uncertainty. The overfitting due to transfer learning is one of the issues, so, this is where pre-trained models like ResNet34 come into play which have been trained on huge datasets such as ImageNet. Fine-tuned learned properties are utilized for the problem at hand, and hence the pre-trained model can retain its parameters and only need to be retrained on the final layers which are task-specific. In fact, the method speeds up the training process convergence as well as alleviates overfitting through reusing the knowledge acquired.

In practice, this technique significantly reduces the dependency on large amounts of labeled data, enhances training efficiency, and improves the overall performance of the model on the new task. Transfer learning, thus, represents a critical advancement in machine learning, offering substantial benefits in data-scarce environments and complex problem domains.

Data augmentation is an essential technique in machine learning and deep learning which is used to improve model performance and curtail overfitting. This method, in particular, aids in the problem of insufficient image data where the real dataset is artificially expanded and made diverse by data simulation of different conditions in the real world, for instance, light conditions, angles, and deformation of leaves. Several augmentation techniques such as adjusting image brightness, flipping images horizontally, and moving images at particular angles are utilized. Taking, for example, techniques like rotating images to the left if 45 and 90 degrees and stretching along are the ones that come to mind to aid in the dataset expansion Sun. Being that these augmentation techniques are used, therefore, the dataset is augmented in every category which in turn allows the model to generalize more effectively in scenarios and thus it improves its robustness in the real world.

3. Results and Discussion

In this section, we present the results obtained from various deep learning models, specifically CNN architectures, for detecting apple leaf diseases. The dataset utilized in this study comprises a diverse collection of apple leaf images, representing several disease categories, including Apple Scab, Powdery Mildew, and Healthy leaves. This comprehensive dataset allows for an in-depth evaluation of the models' performance in accurately identifying and classifying these diseases. Traditionally, data processing has been done either through training-test or training-validation split. However, these approaches may fail to effectively demonstrate the models' ability to extrapolate unseen data. Differently from previous works, the approach used a 70%/15%/15% division into training, validation, and test sets to guarantee that the datasets were split and not overlapping. For this reason, this strategy meant that each model was assessed solely by its results on the unseen test set to give a clear approximation of generalization ability. In each of the models, only a single training cycle was performed, and the model that achieved the highest validation was chosen. This model was then evaluated on the test data in order to evaluate the performance of this model on other data samples.

There approaches of CNN models were trained and optimized based on the apple leaf disease dataset. The experiments performed on these models derived how efficient it is to classify diverse diseases of apple leaves with high accuracy, precision, recall, and F1-score. The results confirm the usefulness of CNN models for diagnosing apple leaf diseases and outlining feasible strategies for controlling them in apple plantations. These findings suggest that with a high degree of accuracy, precision, recall and F-score, these models may be extremely useful for the identification and differentiation of apple leaf diseases (Table 2).

Table 2.
Model performance metrics for apple leaf diseases

Model	Accuracy%	Precision%	Recall%	F1-Score%
ResNet50	98.95	98.39	98.34	98.60
InceptionV4	99.79	99.74	99.90	99.82
Xception	99.79	99.42	99.74	99.57
DenseNet121	99.79	99.74	99.90	99.82
EfficientNetV2_m	1.00	1.00	1.00	1.00
VGG13	99.79	99.90	99.73	99.81

Accuracy is therefore calculated from the model by the true positive and the true negative values against the total models in use. As is visibly observed, all the models possess high accuracy values. The Maximum accuracy is recorded at an accuracy of 100 % from the EfficientNetV2_m. Other models demonstrate high accuracy values and they vary from 98.95 % and up to 99.79 %. This suggests that the models are for the most part running effectively.

Precision expresses the ratio of true positive predictions to the total positive predictions. In Table 2, the EfficientNetV2_m model has the highest precision value of 100%. The precision values of the other models range from 98.39% to 99.90%, indicating that classification errors are quite low.

Recall indicates the ratio of true positive predictions to the actual positives. The EfficientNetV2_m model also demonstrates the highest recall value of 100%, while the other models show variation between 98.34% and 99.90%. High recall values are critical for disease detection, as they emphasize the importance of not missing any diseases.

The F1-Score is the harmonic mean of precision and recall values. This helps maintain a balance. The EfficientNetV2_m model stands out with the highest F1 score of 100%, while the other models range between 98.57% and 99.82%. A high F1 score indicates a reduction in both false positives and false negatives.

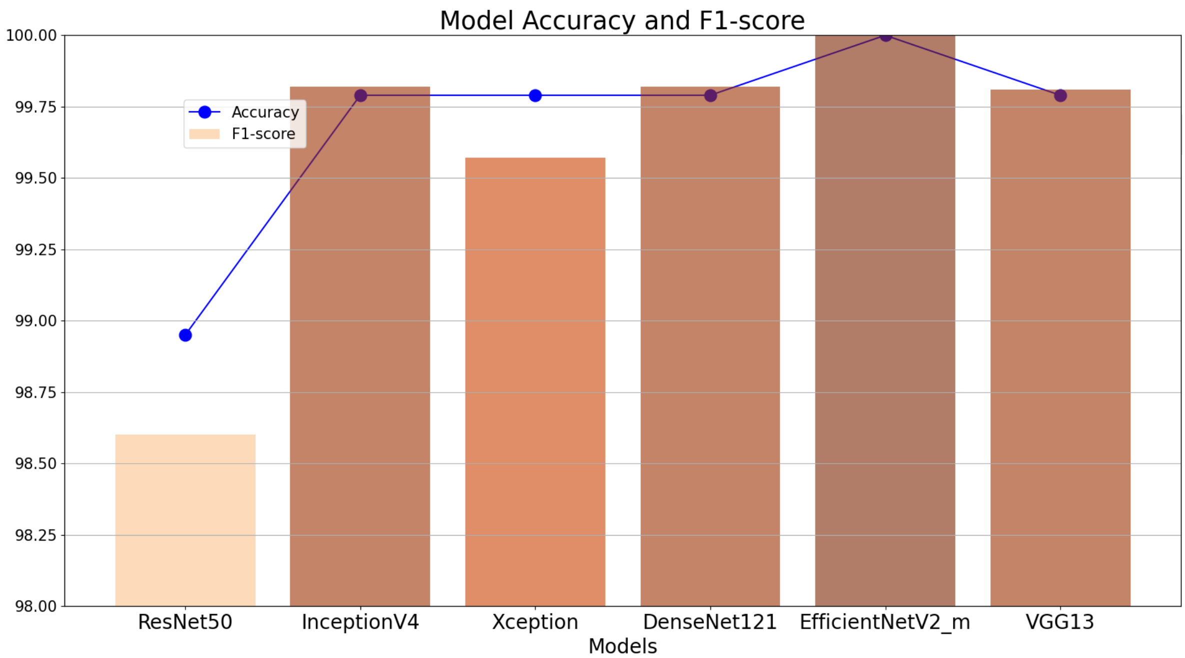

Fig. 2 Accuracy and F1-score values for CNN models

When looking at the models, it can be said that they generally exhibit high accuracy and F1-score (Fig. 2). However, the EfficientNetV2_m model shows exceptional results with an accuracy and F1-score of 100%.

The confusion matrices for several models, including those that achieved the highest accuracy and one that performed relatively lower, are presented in Fig. 3. The models with high accuracy are displayed in dark blue, while the one with lower accuracy is indicated in light blue.

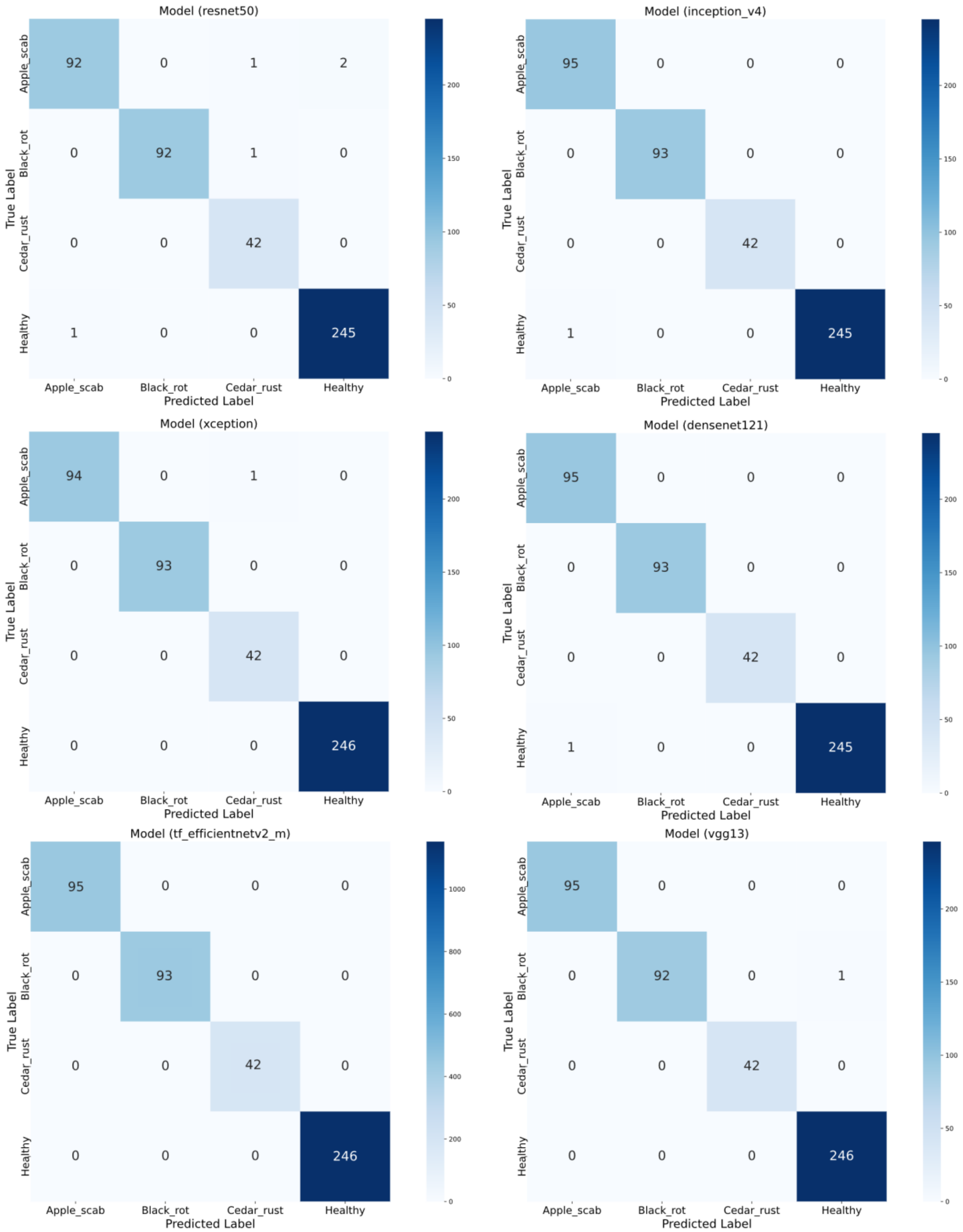

Fig. 3 Confusion matrices for CNN models

In Fig.3, focusing on the ResNet50 model, we observe that it correctly classified 92 instances in the apple_scab class, 92 in the black_rot class, 42 in the cedar_rust class, and 245 in the healthy class. These results highlight the ResNet50 model's effectiveness in identifying plant diseases, particularly in accurately recognizing healthy leaves. The Inception_V4 model also demonstrates strong performance, with 95 correct classifications for apple_scab, 93 for black_rot, 42 for cedar_rust, and 245 for healthy. This further validates the model's effectiveness in accurately distinguishing between the various categories of plant diseases, which will now and again deliver a similar performance. Similarly, the Xception model achieved 94 correct classifications in the apple_scab class, 93 in the black_rot class, 42 in the cedar_rust class, and 246 in the healthy class. The results from the Xception model indicate a high level of accuracy in detecting and classifying plant diseases, mirroring the performance of previous models. During the analysis, DenseNet121 provided equally good performance; it was able to classify 95 apple_scab, 93 black_rot, 42 cedar_rust, and 245 healthy images. The high level of consistencies observed between the confusion matrices tends to reveal that DenseNet121 is very reliable in disease identification. Another model called EfficientNetV2_m also provided exceptional performance with 95 instances correctly classified as apple_scab, 93 for black_rot, 42 for cedar_rust, and 246 for healthy. Its performance expounds the model's capability of identifying and categorizing plant diseases. Lastly in the VGG13 model, we obtained 95 for apple_scab, 92 for black_rot, 42 for cedar_rust, and 246 for healthy. Still, it revealed the detection rate of black_rot was slightly lower, but still, the presented algorithm's efficiency indicates its ability to detect various types of plant diseases. Conclusively, all the analyzed models reveal high accuracy in plant diseases classification, and while having their distinctive Class-specific performance. However, it is worth mentioning that the model, named EfficientNetV2_m, has obtained a perfect result.

4. Conclusion

In this study, methods of the deep learning approach and their suitability for disease detection in apple leaves as well as their relevant scores accuracy, precision, recall, and F1-score were examined. For all the models that have been calibrated, it has emerged that all possess high accuracy levels with the EfficientNetV2_m model receiving 100% on both accuracy and F1-score. This shows not only is the model good at pinpointing diseases' existence but also helps in reducing both false positives and false negatives. The high precision values of all models demonstrate the models' ability to accurately classify the positive cases while the high recall values underscore the need to detect all the diseases, to avoid missing any infected leaves. In general, the study emphasizes the utility of deep learning in agricultural fields especially in disease detection as valuable for the improvement of the disease status of crops. There are various possible directions for improvement for future work: incorporating more features using different sources of information and using ensemble methods. More importantly, the application of these models within the field environments with actual agricultural operations will reveal their feasibility and performance in the different environment scenarios. Extension to this line of research involves extensive validation of these models on different, and big, datasets to establish their versatility, and reliability under different environmental situations and diseases. Nevertheless, extending these models to incorporate options for real-time monitoring systems and applications for mobile devices may expand the field even further, and the creation of an overall platform for disease control and precision agriculture may be developed in the not-too-distant future. Besides making a noteworthy contribution to the domain of plant pathology this study arms farmers and agricultural practitioners with important resources to undertake optimal disease control measures.

References

[1] I. Kunduracioglu and I. Pacal, “Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases,” *Journal of Plant Diseases and Protection*, Jun. 2024, doi: 10.1007/s41348-024-00896-z.

[2] I. Pacal, “Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model,” *Expert Syst Appl*, vol. 238, p. 122099, Mar. 2024, doi: 10.1016/j.eswa.2023.122099.

[3] A. I. Khan, S. M. K. Quadri, and S. Banday, “Deep learning for apple diseases: classification and identification,” *International Journal of Computational Intelligence Studies*, vol. 10, no. 1, p. 1, 2021, doi: 10.1504/IJCISTUDIES.2021.113831.

[4] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, “Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks,” *IEEE Access*, vol. 6, pp. 30370–30377, 2018, doi: 10.1109/ACCESS.2018.2844405.

[5] U. P. Singh, S. S. Chouhan, S. Jain, and S. Jain, “Multilayer Convolution Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease,” *IEEE Access*, vol. 7, pp. 43721–43729, 2019, doi: 10.1109/ACCESS.2019.2907383.

[6] Z. Lin *et al.*, “A Unified Matrix-Based Convolutional Neural Network for Fine-Grained Image Classification of Wheat Leaf Diseases,” *IEEE Access*, vol. 7, pp. 11570–11590, 2019, doi: 10.1109/ACCESS.2019.2891739.

[7] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification,” *Comput Intell Neurosci*, vol. 2016, pp. 1–11, 2016, doi: 10.1155/2016/3289801.

[8] A. E. Rad and A. Ehsanirad, “Plant Classification Based on Leaf Recognition,” 2010. [Online]. Available: http://sites.google.com/site/ijcsis/

[9] M. K. Tripathi and D. D. Maktedar, “A role of computer vision in fruits and vegetables among various horticulture products of agriculture fields: A survey,” Jun. 01, 2020, *China Agricultural University*. doi: 10.1016/j.inpa.2019.07.003.

[10] Y. Tian, E. Li, Z. Liang, M. Tan, and X. He, “Diagnosis of Typical Apple Diseases: A Deep Learning Method Based on Multi-Scale Dense Classification Network,” *Front Plant Sci*, vol. 12, Oct. 2021, doi: 10.3389/fpls.2021.698474.

[11] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” Jul. 2012.

[12] M. K. Tripathi and D. D. Maktedar, “Recent machine learning based approaches for disease detection and classification of agricultural products,” in *2016 International Conference on Computing Communication Control and automation (ICCUBEA)*, IEEE, Aug. 2016, pp. 1–6. doi: 10.1109/ICCUBEA.2016.7860043.

[13] S. Liu *et al.*, “Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer’s Disease,” *IEEE Trans Biomed Eng*, vol. 62, no. 4, pp. 1132–1140, Apr. 2015, doi: 10.1109/TBME.2014.2372011.

[14] İ. Kunduracioğlu, “Examining the Interface of Lego Mindstorms Ev3 Robot Programming,” *Journal of Educational Technology & Online Learning*, vol. 1, no. 1, pp. 28–46, 2018, doi: 10.31681/jetol.372826.

[15] X. Liu, M. Jia, Z. Na, W. Lu, and F. Li, “Multi-Modal Cooperative Spectrum Sensing Based on Dempster-Shafer Fusion in 5G-Based Cognitive Radio,” *IEEE Access*, vol. 6, pp. 199–208, 2018, doi: 10.1109/ACCESS.2017.2761910.

[16] X. Liu, F. Li, and Z. Na, “Optimal Resource Allocation in Simultaneous Cooperative Spectrum Sensing and Energy Harvesting for Multichannel Cognitive Radio,” *IEEE Access*, vol. 5, pp. 3801–3812, 2017, doi: 10.1109/ACCESS.2017.2677976.

[17] İ. Paçal and İ. Kunduracıoğlu, “Data-Efficient Vision Transformer Models for Robust Classification of Sugarcane,” *Journal of Soft Computing and Decision Analytics*, vol. 2, no. 1, pp. 258–271, Jun. 2024, doi: 10.31181/jscda21202446.

[18] S. Baranwal, S. Khandelwal, and A. Arora, “Deep Learning Convolutional Neural Network for Apple Leaves Disease Detection,” *SSRN Electronic Journal*, 2019, doi: 10.2139/ssrn.3351641.

[19] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, “Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks,” *IEEE Access*, vol. 7, pp. 59069–59080, 2019, doi: 10.1109/ACCESS.2019.2914929.

[20] K. Sabanci, A. Kayabasi, and A. Toktas, “Computer vision-based method for classification of wheat grains using artificial neural network,” *J Sci Food Agric*, vol. 97, no. 8, pp. 2588–2593, Jun. 2017, doi: 10.1002/jsfa.8080.

[21] K. Sabanci, A. Toktas, and A. Kayabasi, “Grain classifier with computer vision using adaptive neuro-fuzzy inference system,” *J Sci Food Agric*, vol. 97, no. 12, pp. 3994–4000, Sep. 2017, doi: 10.1002/jsfa.8264.

[22] I. Pacal and I. Kunduracioglu, “Advanced deep learning approach for early detection of potato leaf diseases using Efficient Channel Attention mechanism,” in *International Symposium On Architecture, Engineering And Design (ISAED)*, Diyarbakır/Türkiye, Jul. 2024, pp. 76–86.

[23] S. Zhang, S. Zhang, C. Zhang, X. Wang, and Y. Shi, “Cucumber leaf disease identification with global pooling dilated convolutional neural network,” *Comput Electron Agric*, vol. 162, pp. 422–430, Jul. 2019, doi: 10.1016/j.compag.2019.03.012.

[24] İ. Kunduracıoğlu and İ. Paçal, “Deep Learning-Based Disease Detection in Sugarcane Leaves: Evaluating EfficientNet Models,” *Journal of Operations Intelligence*, vol. 2, no. 1, pp. 321–235, Aug. 2024, doi: 10.31181/jopi21202423.

[25] E. VEZIROGLU, I. PACAL, and A. COŞKUNÇAY, “Derin Evrişimli Sinir Ağları Kullanılarak Pirinç Hastalıklarının Sınıflandırılması,” *Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 13, no. 2, pp. 792–814, Jun. 2023, doi: 10.21597/jist.1265769.

[26] M. Kim, “Apple Leaf Disease Classification Using Superpixel and CNN,” 2021, pp. 99–106. doi: 10.1007/978-3-030-71051-4_8.

[27] R. Verma and V. Singh, “Leaf Disease Identification Using DenseNet,” 2022, pp. 500–511. doi: 10.1007/978-3-030-95711-7_42.

[28] Y. Zhong and M. Zhao, “Research on deep learning in apple leaf disease recognition,” *Comput Electron Agric*, vol. 168, p. 105146, Jan. 2020, doi: 10.1016/j.compag.2019.105146.

[29] H.-J. Yu and C.-H. Son, “Apple Leaf Disease Identification through Region-of-Interest-Aware Deep Convolutional Neural Network,” Mar. 2019.

[30] M. Agarwal, R. K. Kaliyar, G. Singal, and S. Kr. Gupta, “FCNN-LDA: A Faster Convolution Neural Network model for Leaf Disease identification on Apple’s leaf dataset,” in *2019 12th International*

Conference on Information & Communication Technology and System (ICTS), IEEE, Jul. 2019, pp. 246–251. doi: 10.1109/ICTS.2019.8850964.

[31] D. Savla, V. S. Dhaka, G. Rani, and M. Oza, "Apple Leaf Disease Detection and Classification Using CNN Models," 2022, pp. 277–290. doi: 10.1007/978-981-19-2719-5_26.

[32] J. S. H. Al-bayati and B. B. Üstündağ, "Evolutionary Feature Optimization for Plant Leaf Disease Detection by Deep Neural Networks," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, p. 12, 2020, doi: 10.2991/ijcis.d.200108.001.

[33] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier]," *IEEE Comput Intell Mag*, vol. 5, no. 4, pp. 13–18, Nov. 2010, doi: 10.1109/MCI.2010.938364.

[34] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, New York, NY, USA: ACM, Nov. 2014, pp. 675–678. doi: 10.1145/2647868.2654889.

[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[36] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to Forget: Continual Prediction with LSTM," *Neural Comput*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000, doi: 10.1162/089976600300015015.

[37] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[38] H. Durmus, E. O. Gunes, and M. Kirci, "Disease detection on the leaves of the tomato plants by using deep learning," in *2017 6th International Conference on Agro-Geoinformatics*, IEEE, Aug. 2017, pp. 1–5. doi: 10.1109/Agro-Geoinformatics.2017.8047016.

[39] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput Electron Agric*, vol. 147, pp. 70–90, Apr. 2018, doi: 10.1016/j.compag.2018.02.016.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[41] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, "Rethinking the Inception Architecture for Computer Vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[42] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," Oct. 2016.

[43] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708. [Online]. Available: https://github.com/liuzhuang13/DenseNet.

[44] M. Tan and Q. V Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *International conference on machine learning*, 2019, pp. 6105–6114.

[45] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014.