Turkish Journal of Agricultural and Natural Sciences 12 (3): 679–686, 2025
https://doi.org/10.30910/turkjans.1695283

TÜRK TARIM ve DOĞA BİLİMLERİ DERGİSİ

TURKISH JOURNAL of AGRICULTURAL and NATURAL SCIENCES

www.dergipark.gov.tr/turkjans

Original Article

Comparative Performance Analysis of Deep Learning Based CNN Models for Diagnosis of Corn Leaf Diseases

Adnan Gökten [1], Erkut Tekeli [2]

[1]Department of Computer Technologies, Kozan Vocational School, Çukurova University ,
[2]Faculty of Computer and Information Sciences, Department of Software Engineering, Adana Alparslan Türkeş Science and Technology University

[1]https://orcid.org/ 0000-0001-8988-9720, [2]https://orcid.org/ 0000-0001-9468-5378

✉: agokten@cu.edu.tr

ABSTRACT

This study comparatively analyses the effectiveness of deep learning based on the convolutional neural network (CNN) models in the diagnosis of corn leaf diseases. Accurate and rapid diagnosis of leaf diseases, which cause serious economic losses in agricultural production, is critical for production efficiency. In this context, different CNN architectures (AlexNet, GoogLeNet, ResNet, DenseNet, EfficientNet, MobileNet and ConvNeXt) were used to perform classification on Corn leaf images. Model performances were evaluated with metrics such as accuracy, F1 score, precision and ROC-AUC. The results showed that modern architectures (especially ConvNeXtNet) provide higher performance. These findings support the practical applicability of artificial intelligence-supported automatic diagnosis systems in agriculture.

Key words: Deep learning, Corn leaf diseases, CNN, Image classification, ROC-AUC, Precision

Mısır Yaprak Hastalıklarının Teşhisinde Derin Öğrenme Tabanlı CNN Modellerinin Karşılaştırılmalı Performans Analizi

ÖZ

Bu çalışma, mısır yaprak hastalıklarının teşhisinde derin öğrenme temelli evrişimsel sinir ağı (CNN) modellerinin etkinliğini karşılaştırmalı olarak analiz etmektedir. Tarımsal üretimde ciddi ekonomik kayıplara neden olan yaprak hastalıklarının doğru ve hızlı teşhisi, üretim verimliliği açısından kritik öneme sahiptir. Bu bağlamda, farklı CNN mimarileri (AlexNet, GoogLeNet, ResNet, DenseNet, EfficientNet, MobileNet ve ConvNeXt) kullanılarak mısır yaprağı görüntüleri üzerinde sınıflandırma işlemi gerçekleştirilmiştir. Model performansları doğruluk, F1 skoru, keskinlik ve ROC-AUC gibi metriklerle değerlendirilmiştir. Sonuçlar, modern mimarilerin (özellikle ConvNeXtNet) daha yüksek başarı sağladığını göstermiştir. Bu bulgular, tarımda yapay zekâ destekli otomatik tanı sistemlerinin pratikte uygulanabilirliğini desteklemektedir.

Anahtar kelimeler: Derin öğrenme, Mısır yaprak hastalıkları, CNN, Görüntü sınıflandırma, ROC-AUC, Keskinlik

INTRODUCTION

The agricultural sector is of critical importance in terms of ensuring food security on a global scale, as well as contributing to economic development. The efficiency of agricultural production is of great importance in terms of meeting global food needs and ensuring sustainable production. As posited by Lenk et al. (2007), agriculture fulfils a dual function in human society. In addition to providing sustenance, it also serves to supply raw materials that are indispensable to economic growth, employment, and industry. However, the agricultural sector is confronted with numerous challenges, including climate change, declining soil fertility, and the emergence of pests and diseases (Saleem et al., 2024).

Corn is an agricultural crop that is widely cultivated after wheat and rice in terms of cultivation area worldwide, and it occupies a significant position in the food industry (Zhang et al., 2018). In addition to its application in food production, the crop has a wide range of uses, including in the manufacture of animal feed, biofuel and industrial raw materials. It is evident that, owing to its elevated nutritional value and extensive range of applications, the crop occupies a pivotal position within the domain of global agriculture (Erenstein et al., 2022). However, the yield of Corn is subject to the influence of a variety of environmental and biological factors. In particular, foliar diseases, which are prevalent in Corn plants, have been shown to result in significant losses in production and to have a detrimental effect on agricultural productivity (Bruns, 2017; Chen et al., 2021).

Early diagnosis of plant diseases is imperative for minimising losses in agricultural production and ensuring sustainable yield. Conventional agricultural practices typically rely on observational methods for disease detection, but these approaches are susceptible to human error and often require a considerable time investment (Mehta et al., 2023). Incorrect or delayed diagnoses of diseases have the potential to result in producers utilising chemical pesticides in excess or in insufficient amounts. This situation gives rise to both economic losses and environmental problems. Consequently, the utilisation of more sophisticated and automated systems has become imperative in the domain of agricultural disease diagnosis (Megersa et al., 2023).

In recent years, information technologies such as artificial intelligence and machine learning have emerged as significant areas of research and development.

The objective of this study is to comparatively evaluate the performance of different convolutional neural network (CNN) architectures in the image-based classification of corn leaf diseases. Within the scope of the study, widely used CNN models—namely AlexNet, GoogLeNet, ResNet, DenseNet, EfficientNet, MobileNet, and ConvNeXt—were employed for the classification task. The performance of each model was analyzed using evaluation metrics such as accuracy, F1-score, precision, and ROC-AUC. Accordingly, the study aims to identify the most suitable deep learning architecture that can contribute to the development of artificial intelligence-based diagnostic systems in the field of agriculture.

MATERIALS AND METHODS

The dataset utilised in the present study was obtained from the Kaggle platform and comprises 4188 labelled images, encompassing the following classes: Common Rust (1 306 images, Figure 1), Gray Leaf Spot (574 images, Figure 2), Blight (1146 images, Figure 3) and Healthy (1162 images, Figure 4). The data is segmented into 80% for training, 10% for validation, and 10% for testing. The training was executed within the Google Colab environment, utilising the Python programming language and TensorFlow/Keras libraries. All CNN models (AlexNet, GoogLeNet, ResNet, DenseNet, EfficientNet, MobileNet and ConvNeXt) were trained by means of the transfer learning method. Each model was run for 10 epochs.

Figure 1. Examples of the common rust disease images contained within the dataset.

Figure 2. Examples of grey leaf spot disease images contained within the dataset.

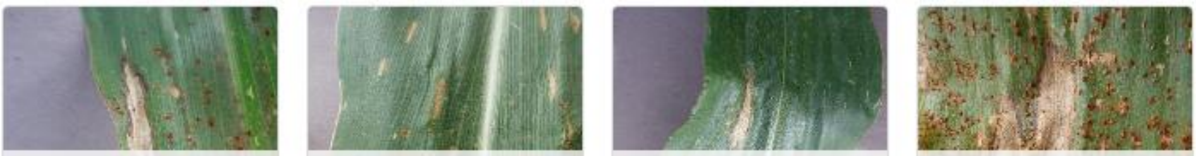

Figure 3. Examples of leaf blight disease images contained within the dataset.

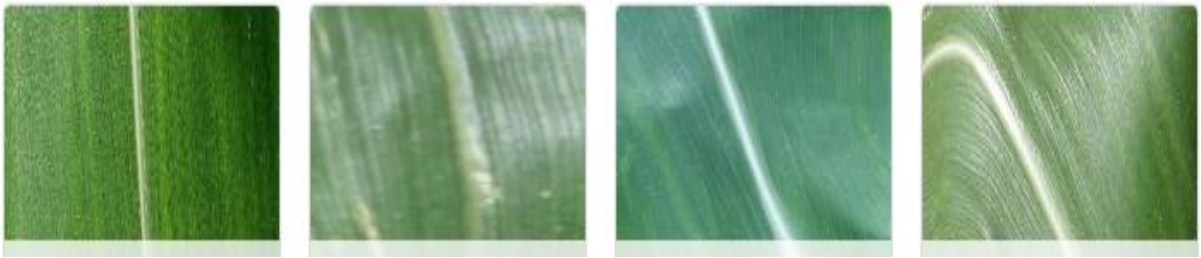

Figure 4. Examples of healthy Corn leaf images found in the dataset

The performance of the model was evaluated based on several metrics, including accuracy, precision, the F1 score, and the ROC curve (AUC).

Accuracy

The term 'accuracy' is a widely used metric for evaluating the performance of a classification model. The term is most commonly defined as the ratio of correct predictions to the total number of predictions. In summary, the precision of a model is calculated as the percentage of instances that it correctly classifies.

Accuracy=(True Positive + True Nagative)/(Total Number of Predictions) (Eq 1.).

Precision

It is a concept that evaluates the overall performance of a model in the context of machine learning and deep learning. It is usually calculated using accuracy and loss values. Precision indicates how well the model generalises, i.e. how effectively it can apply the knowledge learnt from the training dataset to new, unseen data.

Precision= True Positive /(True Positive + False Positive) (Eq.2).

F1-score

This metric is of significant importance in the evaluation of the performance of a model in machine learning, particularly in the context of classification problems. The F1 score is a metric that calculates the equilibrium between precision and recall. Precision and recall seek to achieve equilibrium between false positive and false negative predictions of a model, a strategy that is especially beneficial in the context of imbalanced datasets.

Calculation:

Precision and recall are frequently considered to be opposing objectives; enhancing one may concomitantly diminish the other. The F1 score provides an overall measure of performance by balancing these two metrics. The F1 score is calculated using the harmonic mean.

F1=2 x (Precision x Recall)/(Precision+Recall) (Eq 3.).

ROC curve

The Receiver Operating Characteristic (ROC) curve is a tool employed for the evaluation of the performance of classification models. Its employment is most prevalent in the context of two-class classification problems. The analysis of the ROC curve facilitates the observation of the alterations in the true positive rate (TPR) and false positive rate (FPR) of the model across varying thresholds.

The following steps are to be taken in order to conduct ROC curve analysis:

The following definition of axes is provided:

Y-axis (TPR - True Positive Rate): The proportion of samples that are correctly classified as positive by the model. This is also referred to as 'sensitivity'.

TPR=(True Positive)/(True Positive+False Negative) (Eq 4.).

X-axis (FPR - False Positive Rate): The proportion of samples that the model incorrectly classifies as positive. Also known as “1 – specificity”.

FPR=(False Positive)/(False Positive+True Nagative) (Eq 5.).

Interpretation of ROC Curves

A ROC curve is a graphical representation of the relationship between true positive rate (TPR) and false positive rate (FPR) as a function of threshold. The curve is a representation of the model's capacity for discrimination. When the ROC curve is closer to the upper left quadrant, indicating a high TPR and low FPR, the model performs optimally.

The ideal model: The ROC curve for this model is closest to the point (0,1), i.e. the false positive rate is 0 and the true positive rate is 1.

Random forecasting model: The ROC curve for this model is a diagonal line with a 45° inclination, as the model makes random predictions for both classes.

Area Under Curve (AUC) Score

The AUC score is a metric used to assess the performance of a diagnostic test. The AUC is a metric that encapsulates the overall performance of a model by way of a single number. In the event of the AUC score approximating 0.5, the model is hypothesised to be predicting randomly. It is evident that the closer the AUC score is to 1, the superior the model's performance is deemed to be.

When the AUC is equal to 1.0, the model is considered to be a perfect model. If the AUC is 0.5, the model is considered to be a random prediction model. If the AUC is greater than 0.7, the model is considered to be significantly better than a random prediction. The model in question has been demonstrated to demonstrate high performance, with an AUC greater than 0.9.

The shape of the curve facilitates comprehension of the classifier model's efficacy at varying thresholds. Should the ROC curve demonstrate a rapid upward trend, remaining in close proximity to the left, it can be deduced that the model is accurately predicting positives and effectively minimising false positives.

RESULTS AND DISCUSSION

In this study, the performances of seven different CNN architectures in classifying Corn leaf diseases were analysed. In terms of accuracy, precision, F1 score and ROC-AUC metrics. Table 1 summarises the main performance measures obtained by all the models.

Table 1. Performance criteria demonstrated by models in the classification of leaf diseases of maize

Model Name	Accuracy (%)	Precision	F1 Score	AUC average Values
AlexNet	90.3	0.88	0.89	0.975
GoogLeNet	91.5	0.90	0.91	0.965
ResNet	93.2	0.92	0.93	0.967
DenseNet	95.1	0.94	0.95	0.970
EfficientNet	94.3	0.91	0.94	0.957
MobileNet	92.0	0.90	0.91	0.960
ConvNeXt	96.4	0.95	0.96	0.980

As illustrated in Figure 5, a performance comparison of these seven distinct CNN models is presented in terms of accuracy, precision, and the F1 score. In all models, the accuracy rate is consistently high, generally at the level of 1.0. However, it should be noted that the precision and F1 score values vary according to the model. The ConvNeXt model demonstrated the most optimal performance in terms of F1 score and precision. DenseNet and EfficientNet models also attracted attention with similarly elevated F1 scores. Conversely, the AlexNet, GoogLeNet and MobileNet models demonstrated the poorest performance in terms of F1 score. Conversely, the ResNet model exhibited a moderate degree of success. The results indicate that the classification accuracy of the models is generally high, but there are significant variations in the metrics.

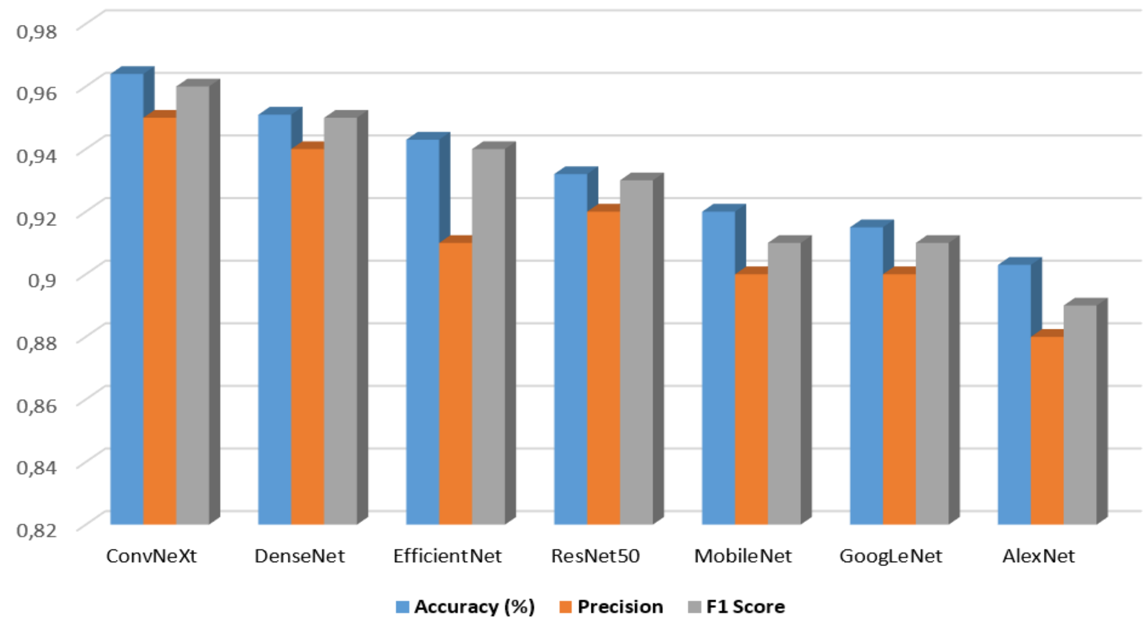

Figure 5. Comparison of CNN models in terms of accuracy, precision and F1-score.

As illustrated in Figure 6, the comparative AUC scores of seven distinct CNN models for four classes (Blight, Common Rust, Gray Leaf Spot, Healthy) are demonstrated. It was generally observed that all models attained the highest AUC scores in the 'Healthy' class. The EfficientNet, ConvNeXt and DenseNet models demonstrate superior overall performance and consistency between classes. The EfficientNet model has been observed to generate considerable interest, with the AUC scores demonstrating a high degree of proximity to 1.0, particularly within the 'Healthy' and 'Common Rust' classes. In contrast, the AlexNet and GoogLeNet models exhibited lower AUC values compared to other models, demonstrating substantial declines, particularly in the 'Gray Leaf Spot' category. In contrast, MobileNet and ResNet demonstrated moderate performance, yielding relatively higher AUC scores in the 'Blight' and 'Common Rust' classes. The results demonstrate that there are discrepancies in model performance between classes, with some models exhibiting superior prediction efficacy in specific classes.

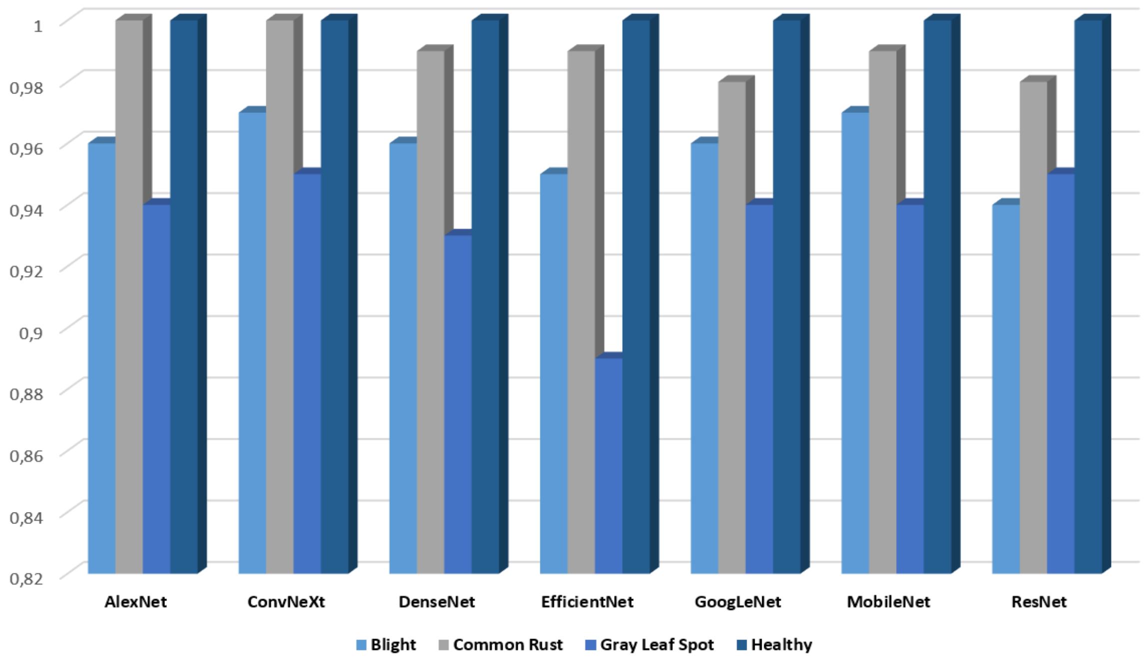

Figure 6. AUC values of CNN models on the ROC curve by class.

In this study, the performances of different CNN-based architectures in the classification of plant diseases were evaluated based on fundamental metrics such as accuracy, precision, F1 score and AUC. The results indicate that the ConvNeXt model demonstrated superior performance in both general success metrics and class-based AUC analysis when compared to other models. Specifically, it attained the highest values, with 96.4% accuracy, 0.95 precision, 0.96 F1 score and 0.980 AUC average. This result suggests that this architecture can learn visual differences in plant leaves with greater effectiveness. It is noteworthy that the DenseNet and EfficientNet models emerged as noteworthy alternatives, exhibiting a high degree of accuracy and an F1 score that is comparable to that of the top-performing models. However, the relatively lower AUC average in the EfficientNet model suggests that the discrimination between classes may have decreased in some classes.

In the class-based AUC analysis, it is noteworthy that the models generally demonstrated the highest success in the 'Healthy' class. This result suggests that healthy leaves exhibit clearer visual characteristics in comparison to diseased leaves. Conversely, the 'Gray Leaf Spot' category exhibited lower AUC values, particularly in relatively aged architectures such as AlexNet and GoogLeNet. This result suggests that the capacity of older networks to discern fine details may be constrained. Indeed, as Wei et al. (2022) have reported, architectures enhanced in depth have been shown to achieve higher levels of generalisation in the classification of plant diseases.

The observation that lightweight networks such as MobilNet and GoogLeNet can attain satisfactory accuracy and AUC values despite their modest hardware requirements indicates the potential suitability of these models for mobile applications or edge devices (Howard et al., 2017).. This finding underscores the necessity for decision support systems operating in the field to strike a balance between energy efficiency and accuracy.

In conclusion, this study demonstrates that the ConvNeXt architecture is a strong candidate for the automatic diagnosis of plant diseases due to its high accuracy and discriminative power between classes. However, it is imperative to consider factors such as class balance, application context (i.e. mobile or laboratory) and system resources when selecting a model. In future studies, the performance of these architectures in real-time applications and their generalisation capacity in different data sets can be investigated in detail. As illustrated in Figure 7, the learning curve of the ConvNextNet model is demonstrated, while Figure 8 presents the ROC curves.

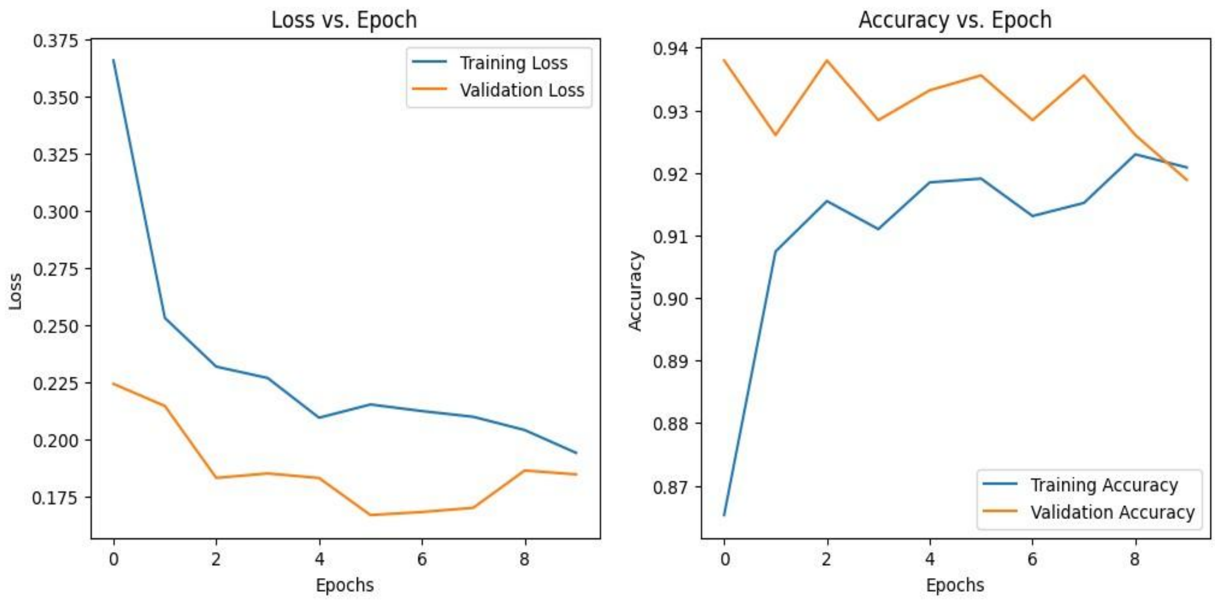

Figure 7. Learning curve of ConvNeXt Model

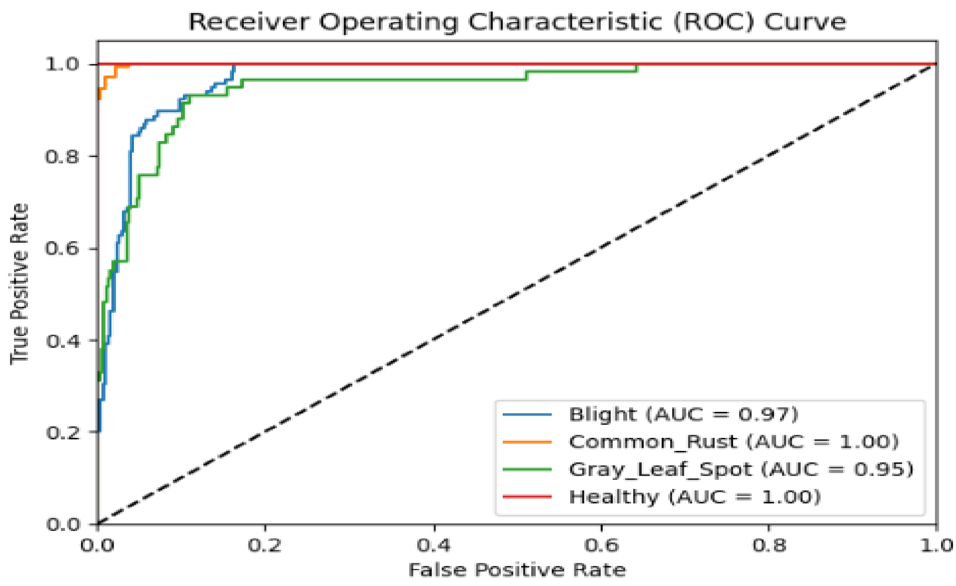

Figure 8. ROC curve of the ConvNeXt Model

CONCLUSION

This study presents a comparative evaluation of various deep learning-based convolutional neural network (CNN) architectures for classifying diseases observed in corn leaves. The models were evaluated using commonly accepted performance metrics such as accuracy, precision, F1 score, and AUC to determine their relative effectiveness in identifying plant diseases.

Among the various architectures examined, ConvNeXt achieved the highest performance across all evaluation criteria. It demonstrated notable advantages, particularly in class-based AUC analysis, by offering both strong overall accuracy and consistent differentiation between disease categories. However, other modern architectures such as DenseNet and EfficientNet also yielded competitive results, particularly in terms of F1 score and overall classification performance, highlighting their continued importance in such tasks.

On the other hand, lightweight models such as MobileNet and GoogLeNet, despite their computational efficiency and ability to achieve acceptable overall accuracy, have shown limitations in class-specific performance, particularly in the detection of diseases such as Grey Leaf Spot. Here, lower AUC values indicate weaker discriminative power. These results highlight the importance of considering both model efficiency and diagnostic accuracy when selecting architectures for use in resource-constrained environments.

In conclusion, the findings demonstrate the strong potential of deep learning-based CNN models for plant disease classification. In particular, ConvNeXt emerged as the most balanced and generalisable model among those tested and has become an attractive option for implementation in real-world decision support systems used in agricultural settings (Liu et al., 2022). However, the promising results of other architectures indicate that model selection should be tailored to the specific constraints and requirements of the intended application.

In future studies, these models should be tested on datasets composed of images captured under real-world field conditions to better evaluate their robustness and practical applicability. Additionally, ensemble or hybrid modelling approaches can be developed to address class imbalance. Such systems can contribute to improving both accuracy and stability (Ciran & Özbay, 2022).

All models were trained using fixed hyperparameters and a limited number of training epochs. While this approach does not aim to maximize the individual performance of each architecture, it provides a consistent foundation for fair comparison. The findings offer a valuable starting point for future research focused on model customization, dataset structure analysis, and real-world deployment readiness.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Author Contributions

Adnan GÖKTEN :Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; software; writing— original draft; writing—review and editing.

Erkut TEKELİ **:**Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; software; writing— original draft; writing—review and editing.

ORCID

Adnan GÖKTEN **http://orcid.org/ 0000-0001-8988-9720**
Erkut TEKELİ **http://orcid.org/ 0000-0001-9468-5378**

Article History	
Submission received:	08.05.2025
Revised:	10.06.2025
Accepted:	13.06.2025

REFERENCES

Bruns, H. A. (2017). Southern corn leaf blight: A story worth retelling. Agronomy Journal, 109(4), 1218-1224. https://doi.org/10.2134/agronj2017.01.0006

Chen, J., Wang, W., Zhang, D., Zeb, A., & Nanehkaran, Y. A. (2021). Attention embedded lightweight network for maize disease recognition. Plant Pathology, 70(3), 630-642. https://doi.org/10.1111/ppa.13322

Ciran, A., & Özbay, E. (2022). Ön-Eğitimli CNN Mimarilerinin Füzyonu ile Mısır Yaprağı Hastalıklarının Sınıflandırılması. Avrupa Bilim ve Teknoloji Dergisi, 44, 74-83. https://doi.org/10.31590/ejosat.1216356

Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K., & Prasanna, B. M. (2022). Global maize production, consumption and trade: trends and R&D implications. Food Security, 14(5), 1295-1319. https://doi.org/10.1007/s12571-022-01288-7

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6325-6334.

J, ARUN PANDIAN; GOPAL, GEETHARAMANI (2019), “Data for: Identification of Plant Leaf Diseases Using a 9-layer Deep Convolutional Neural Network”, Mendeley Data, V1, doi: 10.17632/tywbtsjrjv.1

Lenk, F., Bröring, S., Herzog, P., & Leker, J. (2007). On the usage of agricultural raw materials - Energy or food? An assessment from an economics perspective. Biotechnology Journal, 2(12), 1497-1504. https://doi.org/10.1002/biot.200700153

Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11976-11986.

Megersa, Z. M., Adege, A. B., & Rashid, F. (2023). Real-Time Common Rust Maize Leaf Disease Severity Identification and Pesticide Dose Recommendation Using Deep Neural Network. Knowledge, 4(4), 615-634.

Mehta, S., Kukreja, V., & Gupta, A. (2023). Revolutionizing Maize Disease Management with Federated Learning CNNs: A Decentralized and Privacy-Sensitive Approach. 2023 4th International Conference for Emerging Technology, INCET 2023, 1–6. https://doi.org/10.1109/INCET57972.2023.10170499

Saleem, A., Anwar, S., Nawaz, T., Fahad, S., Saud, S., Ur Rahman, T., Khan, M. N. R., & Nawaz, T. (2024). Securing a sustainable future: the climate change threat to agriculture, food security, and sustainable development goals. Journal of Umm Al-Qura University for Applied Sciences. https://doi.org/10.1007/s43994-024-00177-3

Singh D, Jain N, Jain P, Kayal P, Kumawat S, Batra N. PlantDoc: a dataset for visual plant disease detection. InProceedings of the 7th ACM IKDD CoDS and 25th COMAD 2020 Jan 5 (pp. 249-253).

Wei, K., Chen, B., Zhang, J., Fan, S., Wu, K., Liu, G., & Chen, D. (2022). Explainable deep learning study for leaf disease classification. Agronomy, 12(5), 1035. https://doi.org/10.3390/agronomy12051035

Zhang, X., Qiao, Y., Meng, F., Fan, C., & Zhang, M. (2018). Identification of maize leaf diseases using improved deep convolutional neural networks. IEEE Access, 6, 30370–30377. https://doi.org/10.1109/ACCESS.2018.2844405