Cumhuriyet Sci. J., 45(4) (2024) 701-706
DOI: https://doi.org/10.17776/csj.1548189

Cumhuriyet Science Journal

| csj.cumhuriyet.edu.tr | Founded: 2002 ISSN: 2587-2680 e-ISSN: 2587-246X Publisher: Sivas Cumhuriyet University

Classification of Grapevine Leaf Types with Vision Transformer Architecture

Esra Kavalcı Yılmaz[1,a,*], Hatice Aktaş[1,b] , Kemal Adem[2,c]

[1] *Department of Computer Engineering, 'Sivas University of Science and Technology, Sivas, Türkiye.*
[2] *Department of Computer Engineering, Sivas Cumhuriyet University, Sivas, Türkiye.*
**Corresponding author*

Research Article

History
Received: 11/09/2024
Accepted: 13/12/2024

ABSTRACT

Viticulture plays an important role in agriculture. Farmers prefer grapevine cultivation because not only its fruit but also its leaves are used in various fields. Both the use and trade of grapevine leaves within the country is an important source of income. Grapevine leaves, which are grown in almost all countries and used as edible, vary in terms of species. Determining and cultivating the species according to their suitability in terms of productivity is important. In this study, artificial intelligence methods were used to classify grapevine leaf species. The dataset consisting of five different classes, including 100 grapevine leaf images for each class, totalling 500 images, was classified using ViT, VGG19 and MobileNet methods. When the methods used in this study to help increase productivity in production are evaluated, ViT method has the best accuracy rate with 94%.

Keywords: Grapevine leaf classification, Deep learning, ViT, Artificial intelligence.

[a] *esra.kavalcı@sivas.edu.tr* *https://orcid.org/0000-0003-1314-4495*
[b] *haticeaktas@sivas.edu.tr* *https://orcid.org/0000-0002-1104-9307*
[c] *kemaladem@cumhuriyet.edu.tr* *https://orcid.org/0000-0002-3752-7354*

Introduction

Agriculture is one of the most important sources of livelihood in our country. Viticulture has a great place in agriculture, which has many varieties and branches. It is often preferred by farmers both for the livelihood it provides and for its diversity of use. Grapevine is one of the wide range of products provided by viticulture. In our country, the fruit obtained from the grapevine is used in different areas. In addition, grapevine leaves are in demand as much as the fruit [1]. Grapevine leaves, which are frequently consumed in every region of our country, is a product with economic return. The trade of grapevine leaves, which are made durable by using different storage methods, has also increased in recent years. It is important to cultivate, collect, and store different types of grapevine leaves suitable for edible use in different regions. The grapevine leaves, which have different characteristics according to their thickness, shape and hairiness, are carefully selected during cultivation and collection.

The high diversity of grapevine leaves grown in almost every region and the manual determination of the species can negatively affect the classification process and efficiency of the appropriate grapevine leaves to be selected for consumption. The use of systems established with artificial intelligence in the field of agriculture, as in every field, is quite common today. [2]. In the literature, there are many studies involving artificial intelligence for the detection of diseases, identification and classification of species in the field of agriculture [3, 4]. Artificial intelligence has been used in some previous studies on grapevine and grapevine leaves. In these studies, diseases such as mould, grapevine yellowness, leaf blight, black measles, and esca seen in grapevine and grapevine leaves were tried to be detected by artificial intelligence [5-10]. In the presented studies, in addition to methods such as XGBoost, SVM, DNN, CNN, AlexNet, GoogleNet, and ResNet-50, artificial intelligence methods were also used in the image processing phaseIn another study [11], nitrogen concentration estimation of grapevine leaves was also performed using artificial intelligence. Some artificial intelligence methods used to classify the diversity of grapevine leaves were also presented in [12, 13]. In the study [12], where data augmentation was first performed using a dataset containing images of grapevine leaves, Support Vector Machines were used for classification and an accuracy rate of 96.14% was obtained. In the classification study [13], in which both CNN model and Support Vector Machines were used together, grapevive leaf image data set containing 5 different classes was used and 97.6% success rate was obtained in classification.

In addition to agriculture, studies in the literature show that the ViT model yields successful results in different fields. The best example of this can be given as studies in the field of health. Zang and Wen presented a transformer-based framework for automatic diagnosis of COVID-19 from chest CT scans. In this framework, lung segmentation was performed with UNet and 3D volume level features were extracted using the Swin transformer. The study was performed on the MIA-COV19D dataset and they obtained an F1 score of 0.935 and an accuracy of 94.3% with the Swin transformer [14]. Tyagi et al. used CNN, VGG16 and Vision Transformer (ViT) models to automatically detect pneumonia from chest X-rays. The ViT model showed the best performance in pneumonia detection with an accuracy of 96.45% and it was stated that it can be used to recognize other lung-related

diseases [15]. Dai et al. proposed the TransMed model, which combines the advantages of CNN and transducers for multimodal medical image classification. The model aims to establish long-range dependencies between modalities while extracting low-level features of images. It was tested on two datasets, classifying parotid gland tumors and knee injuries. For parotid gland tumors, 88.9% success was achieved with the proposed TransMed model [16]. In another study, Kamran et al. proposed an approach that converts fundus images into Fluorescence Angiography (FA) images to diagnose retinal abnormalities. Fundus images are synthesized into FA images using a Generative Adversarial Network (GAN) and these images are given to a transform model. The proposed model achieved 85.7% accuracy, 83.3% sensitivity and 90% specificity in classifying normal and abnormal retina, respectively [17]. Zeid et al. investigated the effectiveness of Vision Transformers in the classification of Colorectal Cancer (CRC) histological images. In their study, they used a publicly available CRC histology dataset and performed experiments with a total of 5000 images with eight different tissue categories. By training Vision Transformer and Compact Convolutional Transformer models, they achieved 93.3% and 95% accuracy rates, respectively [18]. In another study, Xu et al. proposed a model that combines transfer learning with attention mechanism for glaucoma detection. This method strengthened the distinction between general and specific features by identifying the regions containing information in images, and also allowed visualization of these regions thanks to the attention mechanism. The model was evaluated on two different datasets and achieved impressive results with 85.77% accuracy, 84.9% sensitivity, 86.9% specificity and 0.929 AUC values[19].

Vision Transformer has been used in different fields with its minimization of data loss, self-attention mechanism and highlighting of important features and it has been seen that good results have been obtained. In this study, considering the advantages of the ViT method, grapevine leaf classification was performed according to their types. The Vit model is trained with image dataset containing grapevine varieties for multiclass classification. Grapevine leaf images were processed with artificial intelligence and classified according to 5 different previously determined species classes. With this study, it is aimed to determine the species of grapevine leaves, which are of great interest both in the domestic and foreign markets, more effectively and to contribute to increasing the cultivation efficiency.

Materials and Methods

Data Set

The 'Grapevine Leaf Image Dataset' used in the study was created by Köklü et al. [13] and shared on the Kaggle website. The dataset consists of images of 5 different types of grapevine leaves, namely Ak, Ala Idris, Büzgülü, Dimnit, and Nazli. In the dataset, there are 100 images of each class in 512x512 size, totaling 500 images in total [20]. Sample images of each class are presented in Figure 1. In this study, the data set was tripled by using data augmentation methods such as random rotation, horizontal and vertical translation, color changes and random cropping. The dataset, which initially consisted of 400 training and 100 test images, reached 1200 training and 300 test images as a result of these processes.

Figure 1. Sample images of classes

Deep Learning Methods

Deep learning methods are a subset of machine learning involving neural networks with multiple layers to model complex patterns in data. These methods are especially effectively used in tasks such as image and speech recognition, natural language processing, and autonomous driving. Deep learning algorithms can learn from large amounts of data by simulating the structure of the human brain, making them powerful tools for predictive analytics and decision making [21,22]. It is widely used for disease diagnosis in healthcare, fraud detection in finance, plant species and disease diagnosis in agriculture, and content recommendations in entertainment [23-29]. The deep learning methods used in this study are briefly mentioned below.

- Vision Transformer (ViT): Vision Transformer (ViT), a deep learning model commonly used in computer vision, was introduced by Dosovitskiy et al. in 2020 [30]. Unlike traditional convolutional neural networks (CNNs), ViT uses a transformer architecture, which has become a popular alternative for image processing tasks. ViT utilizes a set of attention mechanisms instead of convolutional layers during the training of data. It helps to reduce image size by splitting images into smaller patches, allowing higher-dimensional images to be processed compared to previous models. By converting images into patches and applying an attention mechanism, ViT can more efficiently

capture complex patterns in the data. Thanks to this approach, higher success can be achieved with less data in studies using ViT. This model has proven its effectiveness in tasks such as image classification and object detection, showing significant improvements over CNNs in various benchmarks. As a result, ViT has become a powerful tool for advancing the state of the art in computer vision [31]. The main reason for choosing the ViT (Vision Transformer) model in this study is to avoid the loss of information caused by the pooling layers used in traditional CNN models. While CNN models are successful in capturing local features, they are limited in modeling long-range relationships and can lead to loss of details during pooling operations. The ViT model divides the image into fixed-size patches, enriches the features of each patch with position information, and can globally model the relationships between all patches in the image with a self-attention mechanism. In this way, it provides superior performance, especially in applications that require detail precision and work on limited data sets [32].

VGG19: VGG19, proposed by Simonyan and Zisserman [33] in 2015, is a convolutional neural network model designed for image classification tasks. VGG19 consists of 19 layers, 16 convolutional layers and 3 fully connected layers, making it a deep and powerful architecture. VGG19 provides a robust framework for complex image recognition tasks. The architecture of the network is characterised by its simplicity; it consistently uses small 3x3 convolutional filters throughout the layers, which allows it to effectively capture complex features in images. Despite its depth, VGG19 is relatively easy to implement, making it a popular choice for researchers and developers alike. One of the greatest strengths of VGG19 is its ability to generalise well across a wide range of visual tasks, making it widely used in computer vision studies [34].

MobileNet: MobileNet, developed by Google, is a convolutional neural network model specifically designed for efficient performance on mobile and embedded devices. One of its key features is the fast convergence of the model during training. In addition, MobileNet is ideal for environments with limited resources due to its low memory consumption [35]. It also reduces the overall computational load by requiring fewer computational operations compared to traditional models. This efficiency is achieved through the use of deeply separable convolution layers that significantly reduce the number of required parameters and computations. Its lightweight architecture ensures high accuracy while minimising latency, even when running on hardware with limited capabilities. This makes MobileNet a popular choice for applications such as image recognition, object detection, and other machine learning tasks that need to be performed on the go. Its flexibility also allows developers to easily scale the model to meet the specific needs of their applications, ensuring a balance between accuracy and efficiency [36]

Results and Discussion

In this study, three different deep learning models, namely ViT, VGG19, and MobileNet, were used to successfully classify grapevine leaves according to their species. The studies were carried out in Python language using the Tensorflow library. In the study, image sizes were set to 224x224x3, 'batch' size = 16 and 'epoch' = 100. The precision, recall, and f1 score values obtained as a result of the classification processes of ViT, VGG19, and MobileNet models for each class are presented in Table 1, Table 2, and Table 3, respectively.

Table 1. ViT results

Class	Precision	Recall	f1-Score
Ak	0.98	1.00	0.99
Ala Idris	0.94	0.98	0.96
Büzgülü	0.97	0.93	0.95
Dimnit	0.95	0.93	0.94
Nazli	0.97	0.95	0.96

Table 2. VGG19 results

Class	Precision	Recall	f1- Score
Ak	0.84	0.91	0.88
Ala Idris	0.78	0.79	0.81
Büzgülü	0.80	0.67	0.70
Dimnit	0.79	0.78	0.75
Nazli	0.83	0.85	0.86

Table 3. MobileNet results

Class	Precision	Recall	f1- Score
Ak	0.87	0.98	0.92
Ala Idris	0.86	0.80	0.83
Büzgülü	0.81	0.72	0.76
Dimnit	0.84	0.80	0.82
Nazli	0.83	0.92	0.87

When the results of the Vit Model (Table 1) are analyzed, it is seen that the most successful classification process among the classes is performed in Dimnit and Ala Idris classes. The most unsuccessful results were obtained in the images belonging to the Büzgülü class.

Table 4 shows the best results obtained by using the hyper parameters "activation_function='gelu', learning_rate =1e-4, optimization_algorithm=RectifiedAdam, loss_function =CategoricalCrossentropy". Accuracy values, average presicion- recall-F1-socre values, kappa values and loss values are listed for the models used

Tablo 4. ViT, VGG19, MobileNet doğruluk değerleri

Model	Accuracy	Prec Avg	Recall Avg	F1-Score Avg	Kappa	Loss
ViT	0.96	0.96	0.95	0.96	0.95	0.87
VGG19	0.81	0.81	0.80	0.80	0.74	0.98
MobileNet	0.84	0.84	0.84	0.84	0.80	0.94

When Table 4 is analyzed, it is seen that the most successful classification process is provided by the ViT Model (0.96). Also, when the loss values are analyzed, it can be said that the ViT model provides a more stable and efficient learning compared to the others with a lower loss value (0.87). The ViT model provides superiority in capturing the global context thanks to its attention mechanism, enabling effective learning of long-range dependencies in images. In addition, it allows the image to be processed by dividing it into patches. This patch-based input method makes the application easier by reducing the need for extensive preprocessing. Thanks to these advantages, it can be said that the ViT model offers better generalization and more successful results after the training process compared to other models. The box plot of ViT, VGG19, and MobileNet models organized according to the classification results is presented in Figure 2.

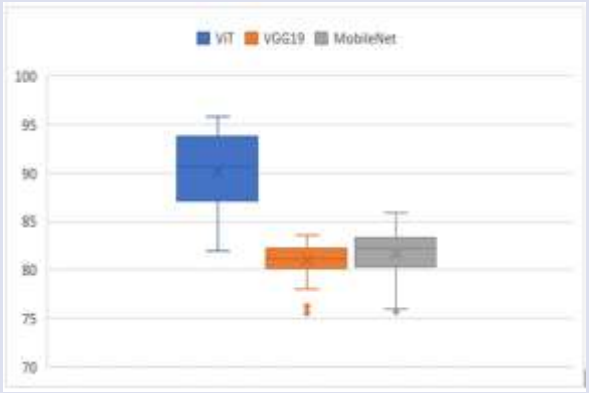

Figure 2. ViT, VGG19 and MobileNet boxplots.

Figure 2 shows that ViT offers the widest range of performance with an accuracy value ranging between 86 and 94 and a median value of around 91. According to these results, ViT indicates high but variable performance. The VGG19 model, on the other hand, has a closer range, the accuracy is mostly between 80 and 83 and the median value is about 82. Therefore, it can be said that it is more consistent than the ViT model but indicates a slightly lower performance. Finally, the MobileNet model shows the closest range with an accuracy between 80 and 84 and a median of around 83. According to these results, it can be said that it shows the lowest and most consistent performance among the three models. ViT has potentially the best performance but shows higher variability, while VGG19 offers a stable performance. MobileNet lags behind in both range and average accuracy. Overall, ViT stands out for its high but fluctuating accuracy, VGG19 for its consistency and MobileNet for its low but stable performance.

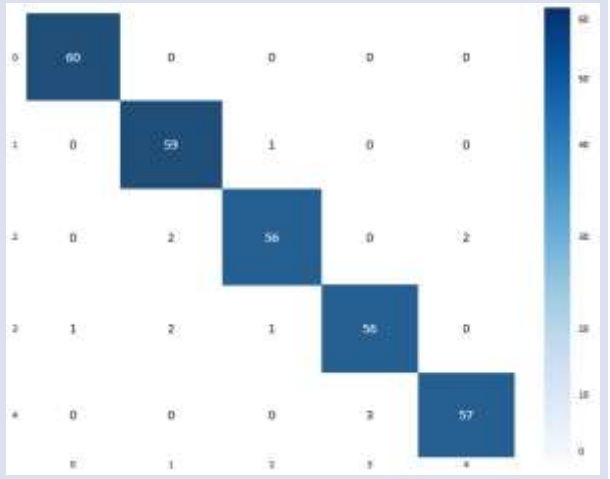

Figure 3. Confusion matrix of the ViT model.

Conclusion and Recommendations

Correct classification of agricultural products increases productivity in production. When the classification process is carried out by means of artificial intelligence methods, the process will be independent of human beings, the expert performing the classification process will not be affected by external factors such as physical, mental conditions and experience level, and the classification process will be concluded more successfully. In addition, the experts performing the classification process will use their time more efficiently by completing more complex tasks instead of routine tasks during this period. In line with these reasons, in this study, classification processes were performed using deep learning methods in order to increase the classification success of the grapevine leaf. Three different deep learning methods, namely ViT, VGG19, and MobileNet, were used in the study and the results were evaluated comparatively. Considering the results obtained, it is seen that the most successful result is obtained with 94% accuracy rate with ViT, a current deep learning model. In future studies, it is thought that more accurate results can be obtained by using various data augmentation methods, increasing and diversifying the data set. In addition, the classification success can be improved by hyperparameter optimization.

Conflict of interest

There are no conflicts of interest in this work.

References

[1] Gülcü M., Torçuk A. İ., *Yemeklik Asma Yaprağı Üretimi ve Pazarlamasında Kalite Parametreleri*, *Meyve Bilimi*, c. 1, ss. 75-79, (2016)

[2] Adem K., Yılmaz E. K., Ölmez F., Çelik K., Bakır H., *A Comparative Analysis of Deep Learning Parameters for Enhanced Detection of Yellow Rust in Wheat*, *UMAG*, 16(2) (2024)

[3] Yılmaz E. K., Oğuz T., Adem K., *A CNN-Based Hybrid Approach to Classification of Raisin Grains*, 1st International Conference on Frontiers in Academic Research, (2023).

[4] Yılmaz E. K., Adem K., Kılıçarslan S., Aydın H. A., *Classification of lemon quality using hybrid model based on Stacked AutoEncoder and convolutional neural network*, *Eur Food Res Technol*, 249(6) (2023) 1655-1667.

[5] Hernández I., Gutiérrez S., Ceballos S., Iñíguez R., Barrio I., Tardaguila J., *Artificial Intelligence and Novel Sensing Technologies for Assessing Downy Mildew in Grapevine*, *Horticulturae*, 7(5) (2021)

[6] Cruz A. *vd.*, *Detection of grapevine yellows symptoms in Vitis vinifera L. with artificial intelligence*, *Computers and Electronics in Agriculture*, 157 (2019) 63-76.

[7] Poblete-Echeverría C., Hernández I., Gutiérrez S., Iñiguez R., Barrio I., Tardaguila J., *Using artificial intelligence (AI) for grapevine disease detection based on images*, *BIO Web Conf.*, 68 (2023) 01021.

[8] Nagi R. Tripathy S. S., *Deep convolutional neural network based disease identification in grapevine leaf images*, *Multimed Tools Appl*, 81 (18) (2022) 24995-25006.

[9] Alessandrini M., Calero Fuentes Rivera R., Falaschetti L., Pau D., Tomaselli V., ve Turchetti C., *A grapevine leaves dataset for early detection and classification of esca disease in vineyards through machine learning*, *Data in Brief*, 35 (2021) 106809.

[10] Jaisakthi S. M., Mirunalini P., Thenmozhi D., Vatsala, *Grape Leaf Disease Identification using Machine Learning Techniques*, *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, (2019) 1-6.

[11] Moghimi A., Pourreza A., Zuniga-Ramirez G., Williams L. E., Fidelibus M. W., *A Novel Machine Learning Approach to Estimate Grapevine Leaf Nitrogen Concentration Using Aerial Multispectral Imagery*, *Remote Sensing*, 12(21) (2020) 3515.

[12] İmak A., Doğan G., Şengür A., ve Ergen B., *Asma Yaprağı Türünün Sınıflandırılması için Doğal ve Sentetik Verilerden Derin Öznitelikler Çıkarma, Birleştirme ve Seçmeye Dayalı Yeni Bir Yöntem*, *International Journal of Pure and Applied Sciences*, 9(1) (2023) 46-55.

[13] Koklu M., Unlersen M. F., Ozkan I. A., Aslan M. F., Sabanci K., *A CNN-SVM study based on selected deep features for grapevine leaves classification*, *Measurement*, 188 (2022) 110425.

[14] Zhang, L., Wen, Y. A transformer-based framework for automatic COVID19 diagnosis in chest CTs. In Proceedings of the IEEE/CVF international conference on computer vision, (2021) 513-518.

[15] Tyagi, K., Pathak, G., Nijhawan, R., & Mittal, A. Detecting pneumonia using vision transformer and comparing with other techniques. In 2021 5th international conference on electronics, communication and aerospace technology (ICECA), (2021) 12-16.

[16] Dai, Y., Gao, Y., Liu, F., Transmed: Transformers advance multi-modal medical image classification, *Diagnostics*, 11(8) (2021) 1384.

[17] Kamran, S. A., Hossain, K. F., Tavakkoli, A., Zuckerbrod, S. L., Baker, S. A., Vtgan: Semi-supervised retinal image synthesis and disease prediction using vision transformers. In Proceedings of the IEEE/CVF international conference on computer vision, (2021)3235-3245.

[18] Zeid, M. A. E., El-Bahnasy, K., Abo-Youssef, S. E., Multiclass colorectal cancer histology images classification using vision transformers. In 2021 tenth international conference on intelligent computing and information systems (ICICIS) (2021) 224-230.

[19] Xu, X., Guan, Y., Li, J., Ma, Z., Zhang, L., Li, L., Automatic glaucoma detection based on transfer induced attention network, *BioMedical Engineering OnLine*, 20(1) (2021) 39.

[20] *Grapevine Leaves Image Dataset,* https://www.kaggle.com/datasets/muratkokludataset/grapevine-leaves-image-dataset

[21] Alaca Y., Emin B., Akgul A., A *comparative study of deep learning models and classification algorithms for chemical compound identification and Tox21 prediction*, *Computers & Chemical Engineering*, 189 (2024) 108805,

[22] Assim O. M., Mahmood A. F., *A novel Universal Deep Learning Approach for Accurate Detection of Epilepsy*, *Medical Engineering & Physics*, (2024) 104219.

[23] Közkurt C., Diker A., Elen A., Kılıçarslan S., Dönmez E., ve Demir F. B., *Trish: an efficient activation function for CNN models and analysis of its effectiveness with optimizers in diagnosing glaucoma*, *J Supercomput*, 80 (11) (2024) 15485-15516.

[24] Kılıçarslan S., Diker A., Közkurt C., Dönmez E., Demir F. B., ve Elen A., *Identification of multiclass tympanic membranes by using deep feature transfer learning and hyperparameter optimization*, *Measurement*, c. 229, s. 114488, (2024)

[25] Daza A., González Rueda N. D., Aguilar Sánchez M. S., Robles Espíritu W. F., Chauca Quiñones M. E., *Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis, Systematic Literature Review, Challenges and Future Works*, *International Journal of Information Management Data Insights*, 4 (2) (2024) 100267.

[26] Altaş Z., Ozguven M., Adem K., *Bazı Bağ Hastalıklarının Faster R-CNN Modeli ile Otomatik Tespit Edilmesi ve Sınıflandırılması*, *Turkish Journal of Agriculture - Food Science and Technology*, 11(97) 97-103.

[27] Adem K., Ozguven M. M., ve Altas Z., *A sugar beet leaf disease classification method based on image processing and deep learning*, *Multimed Tools Appl*, 82(8) (2023) 12577-12594,

[28] Alnasyan B., Basheri M., Alassafi M., *The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review*, *Computers and Education: Artificial Intelligence*, 6 (2024) 100231.

[29] Dönmez E., Kılıçarslan S., ve Diker A., *Classification of hazelnut varieties based on bigtransfer deep learning model*, *Eur Food Res Technol*, c. 250, sy 5, ss. 1433-1442, (2024)

[30] Dosovitskiy A. *vd.*, *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*, International Conference on Learning Representations, (2020)

[31] Dümen S., Yılmaz E. K., Adem K., ve Avaroglu E., *Performance of vision transformer and swin transformer models for lemon quality classification in fruit juice factories*, *European Food Research and Technology*, ss. 1-12, (2024)

[32] Heo, J., Seo, S., Kang, P., Exploring the differences in adversarial robustness between ViT-and CNN-based models using novel metrics, Computer Vision and Image Understanding, 235 (2023) 103800.

[33] Simonyan K. Zisserman A., *Very Deep Convolutional Networks for Large-Scale Image Recognition*, (2015) *arXiv*: arXiv:1409.1556.

[34] Bansal M., Kumar M., Sachdeva M., Mittal A., *Transfer learning for image classification using VGG19: Caltech-101 image data set*, *J Ambient Intell Human Comput*, 14(4) (2023) 3609-3620.

[35] Shome N., Kashyap R., Laskar R. H., *Detection of tuberculosis using customized MobileNet and transfer learning from chest X-ray image*, *Image and Vision Computing*, 147 (2024) 105063.

[36] Kılıçarslan S., Aydın H. A., Adem K., Yılmaz E. K., *Impact of optimizers functions on detection of Melanoma using transfer learning architectures*, *Multimed Tools Appl* (2024).