IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 April 2025, accepted 4 May 2025, date of publication 7 May 2025, date of current version 22 May 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3567774

RESEARCH ARTICLE

A New Lightweight Hybrid Model for Pistachio Classification Using Transformers and EfficientNet

MUHAMMET ÇAKMAK
Department of Computer Engineering, Sinop University, 57100 Sinop, Türkiye
e-mail: mcakmak@sinop.edu.tr

ABSTRACT In recent years, Vision Transformers (ViTs) have gained prominence as a highly effective method for image classification, often outperforming traditional Convolutional Neural Networks (CNNs). However, their relatively slow processing speed limits their practical use, particularly in real-time applications. Conversely, CNN-based transfer learning models provide faster inference but may struggle with classification accuracy on complex datasets. To address these challenges, Temporal Coordinate Attention (TCA) modules have been introduced to optimize efficiency and performance. This study proposes a hybrid architecture combining EfficientNet, Vision Transformer, and Temporal Channel Attention modules that integrates the accuracy of ViTs, the computational efficiency of CNNs, and the enhancement capabilities of TCA modules. The model is designed to classify Siirt and Kirmizi pistachio varieties with high precision. It achieves outstanding results, including 99.07% accuracy, 99.12% recall, and a Cohen's Kappa score of 98.10%. These findings highlight the model's robustness, demonstrating its ability to perform reliable classifications with minimal bias, making it well-suited for real-world applications.

INDEX TERMS Classification, vision transformer, EfficientNet, TCA, machine learning.

NOMENCLATURE

ViTs	Vision Transformers.
CNN	Convolutional Neural Network.
TCA	Temporal Coordinate Attention.
ConV	Convolution.
k-NN	k-Nearest Neighbors.
SVM	Support Vector Machine.
DT	Decision Tree.
RF	Random Forest.
ANN	Artificial Neural Network.
GPU	Graphics Processing Unit.
AI	Artificial Intelligence.
PCA	Principal Component Analysis.
LDA	Linear Discriminant Analysis.
LSTM	Long Short-Term Memory.
HIS	Hue, Intensity, Saturation.
VGG16	Visual Geometry Group Network (16 layers).
VGG19	Visual Geometry Group Network (19 layers).
NAS	Neural Architecture Search.
DeiT	Data-efficient Image Transformer.
CaiT	Class-Attention in Image Transformers.
PiT	Pooling-based Vision Transformer.
FLOPs	Giga Floating Point Operations.
MLP	Multilayer Perceptron.
HSV	Hue, Saturation, Value (color space).
RGB	Red, Green, Blue (color space).
YOLOv7	You Only Look Once, version 7.
MNGNAS	Multi-Objective Neural Generator based on Neural Architecture Search.

The associate editor coordinating the review of this manuscript and approving it for publication was Maria Chiara Caschera.

I. INTRODUCTION

Image classification has been considered one of the leading tasks of computer vision and has been a subject that has been studied in depth due to the adoption of many applications such as object detection, medical imaging, and agriculture. Pistachios are a type of tree nut belonging to the Anacardiaceae family. It is a very popular food known for its flavors and health benefits. With their high protein and fiber content, they constitute a nutritious and healthy option and play an important role in nutrition. Research shows that nuts positively impact cardiovascular health, especially when consumed in moderate amounts [1]. It is also important to note that the quality of pistachios may vary depending on the variety, growing conditions, and processing methods [2]. One of the main functions of classification models is to classify pistachios to ensure the highest possible efficiency in the trading, marketing, and sales processes; This helps maintain consumer satisfaction [3].

Identifying the type of pistachio is crucial for trade purposes [4]. For instance, the classification of pistachios relates to the segmentation process of pistachios into various classes regarding various quality attributes such as size, shape, color, defects, and cracks. However, the currently available pistachio classification has been accomplished through manual means, which is time-consuming, results in subjective outcomes, and is error-prone. Automated pistachio classification methods offer a solution by providing fast, objective, and accurate classification results [5], [6].

Artificial Intelligence has shown remarkable results when combined with plant studies, particularly in the precise classification of various foods based on their visible distinguishing characteristics. Machine learning and deep learning methods are being widely applied in numerous fields, including plant recognition [7], categorization [8], and production [9], as well as disease and pest diagnosis [10], harvest prediction [11], plant phenotype and genotype determination [12], plant sensitivity analysis [13], plant record keeping [14], vegetation development [15], and diversity studies [15]. Various machine and deep learning approaches, such as k-NN, SVM, DT, RF, ANN, and boosting methods [16], [17], [18], [19] as well as CNN, RNN, LSTM [16], [20], [21], among others, have been utilized in various plant-related works in the aforementioned fields.

Over the past ten years, deep learning methods, especially convolutional neural networks (CNNs), have gained significant attention and popularity. CNNs possess an impressive ability to learn the most crucial features from vast datasets, given sufficient training, and they typically don't require complex feature extraction processes. Although traditional CNN architectures perform very well for classification, the layer structures in deep networks cause complexity and speed losses [22]. Newly developed attention-based systems achieve higher success in large-scale networks for multiple classification. Especially in recent years, attention-based and contextual information-based models such as ConvMixer, Swin Transformers, and Vision Transformers [23], [24], [25], [26] have achieved a remarkable success rate.

Researchers use transformer-based NLP models to strengthen image classification, segmentation and object detection systems [27]. The basic structure that distinguishes ViT-based systems from traditional computer vision systems is that they use the attention mechanism entirely. Using the attention mechanism, ViT systems enable relationships between each area of the image, allowing to study more relationships related to each pixel compared to learning per convolution layer. ViTs are particularly useful for detecting and classifying objects with distant contextual relationships [28]. ViTs have the advantage of being flexible to accommodate different input sizes. ViTs produce better results when trained on larger datasets and with the use of larger models. They can also be used in transfer learning using large pre-trained models. Thus, the same model can be used for different visual tasks. The attention mechanism can easily access the GPU to provide higher processing speed and time advantage during training and result generation [27]. However, ViTs also have some limitations and challenges. Especially in small datasets, when there are not enough examples, these models tend to overfit, which can negatively affect performance. In addition, the large amount of computational resources required by ViTs and their higher memory consumption restrict their use, especially on low-resource devices or in real-time applications [30]. The need for large datasets during the training process can make it more complex to effectively implement ViT models. In these aspects, despite the advantages brought by the attention mechanism, ViTs require optimization and resource management in usage scenarios.

As a result of the reviewed literature study, in recent years, ViTs have achieved superior performance in image classification compared to traditional CNNs. However, ViTs generally suffer from slower processing speeds, which is not as expected, especially in time-sensitive tasks. Although CNN-based transfer learning models achieve faster inference times, they may fall short in classification accuracy for complex datasets. This study presents a new hybrid framework that combines the accuracy of ViTs, the speed of CNN-based transfer learning models, and the efficiency of TCA modules. The proposed approach integrates the feature extraction capabilities of CNNs with the global attention mechanisms of ViTs, resulting in a model that provides both high accuracy and computational efficiency. Extensive experiments on datasets show that the hybrid model outperforms standalone ViT and CNN models in terms of accuracy while maintaining competitive inference speeds. This study provides a practical and scalable solution for image classification tasks that require both performance and efficiency, making it suitable for real-world applications.

In summary, the contributions of this article are as follows:

The proposed hybrid model combines the high-accuracy global attention mechanisms of ViTs with the fast inference capabilities of CNN-based transfer learning models, improving both performance and speed in classification tasks.

The hybrid model integrates the effective feature extraction capabilities of CNNs with the global contextual relationship learning capabilities of ViT, providing high classification accuracy and computational efficiency in complex datasets.

The competitive inference speeds of the hybrid model provide a practical solution especially for time-sensitive tasks, providing benefits not only in terms of accuracy but also in terms of fast result production.

By integrating TCA modules, the model reduces both computational costs and enables more efficient analysis of features.

Thanks to the flexible structure of the model, it has been proven to perform well in both large and complex datasets and has become a scalable solution for different image classification tasks.

By balancing both accuracy and speed requirements, the proposed model offers a practical and scalable method that can be used in a wide range of real-world applications.

The literature review is presented in Section II, the materials and methods are detailed in Section III, the results and discussion are provided in Section IV, and the conclusion is outlined in Section V.

II. RELATED WORK

Artificial Intelligence (AI) has shown promising results in accurately classifying different types of foods based on their visual features [31]. Machine learning and deep learning methods have shown great potential in the agricultural sector for various applications such as crop classification [32], [33], [34] disease detection [35], [36], [37], and yield prediction [38], [39], [40]. Deep learning models, a subset of AI, have shown great potential in accurately identifying different types of pistachios based on their visual features. Table 1 presents current literature on the use of AI in pistachio classification.

Rapidly developing technology has caused artificial intelligence applications to affect all sectors rapidly [42], [62], [63]. To date, machine learning and deep learning applications have been used in many fields, such as technology, health [64], economy [55], commerce [65], cyber security [66], [67] education, and agriculture [68]. Determining the disease classification through the observation and experience of an experienced expert is a process that takes a long time and lacks accurate prediction [69]. Liu et al. proposed a deep learning-based method for peanut classification. Deeplab v3+, Segnet, Unet, and Hypernet were used for the proposed model classification. According to the results obtained, they achieved better results by 0.61-1.15% through feature extraction and 0.43-4.96% by feature enhancement [70]. Pourmohammaddali et al. proposed a hybrid model combining genetic algorithm and ANN to classify peanut yield. Using the proposed GA-ANN model, they revealed the effect of Cu (34.6%), K (28.2%), and Fe (26.1%) minerals affecting peanut yield [71]. Vidyarthi et al. proposed a machine learning technique that uses image processing to accurately measure the size and mass of raw peanut kernels. The proposed method was used to identify peanut kernels and calculate their size by analyzing pixels. They used the Random Forest model to estimate peanut kernel mass. The proposed model achieved 95% accuracy [72]. Another study classified different peanut varieties using multi-element fingerprinting and pattern recognition. Five peanut varieties were classified by plasma optical emission measurements (ICP OES). Additionally, PCA and principal component analysis were used in the study. The resulting model classified all samples 99% correctly [44]. Gao and his colleagues used five different deep learning models: SqueezeNet, GoogLeNet, Xception, ResNet50, and AlexNet to classify the quality of peanuts. While SqueezeNet achieved the highest prediction accuracy rate of 99.02%, ResNet50 achieved the highest prediction accuracy rate of 99.96% [45]. Another study proposed a method using image patches to classify pistachio tree species. The proposed model used ORB features to detect and extract patches. EfficientNet-B1 was used on the extracted patches. The proposed model achieved an accuracy rate of 97.2% [46]. Bonifazi et al. used hyperspectral imaging (HSI) to detect contaminants in pistachios. HSI allows qualitative and quantitative analysis of both spectral and spatial information of the sample. According to the results obtained, the performances of four different classification models, prediction maps, and statistical parameters were evaluated [47]. Singh et al. used AlexNet, VGG16, and VGG19 methods to classify pistachios. According to the results obtained, AlexNet achieved 94.42%, VGG16 98.84% and VGG19 98.14% accuracy values [73]. Another study used the RetinaNet network to detect different types of hazelnuts. The proposed model achieved 94.75% accuracy on six videos (9486 frames) [74]. Hosseinpour and his team developed a one-dimensional Convolutional Neural Network (CNN) model to sort peanut nuts. The proposed model distinguishes open and closed-shell peanuts by analyzing the signals generated when peanuts are dropped onto a steel plate. The model achieved 98.75% accuracy. The results of the CNN model were compared with MLP and Random Forest. RF with manual time-domain features achieved an overall accuracy of only 50% and a mean square error (MSE) of 1.56. MLP achieved an OAC of 31.46% and an MSE of 2.57 [50]. Another study used the EfficientNet-B3 model to distinguish peanut species. According to the results, the EfficientNet-B3 model achieved an average precision, recall, and F1 score of 96.73%, 96.70%, and 96.67%, respectively [51]. An et al. classified varieties of chestnuts, hazelnuts, Brazil nuts, cashews, walnuts, pine nuts, macadamias, peanuts, pistachios, and walnuts using transfer learning and neural network modeling. Data augmentation was also applied to the dataset containing 200 images of all types. The proposed model achieved 99.55% accuracy [75]. In another study, six different peanut varieties were analyzed with machine learning algorithms according to their physical properties. The Gaussian processes algorithm achieved the highest correlation coefficient and lowest Root Mean Square Error (RMSE) values, surpassing the Multilayer Perceptron and Random Forest algorithms [76]. The researchers classified pistachios according to whether they were open or closed using AlexNet and Inception V3. According to the results, AlexNet and Inception V3 reached 96.13% and 96.54% test accuracy, respectively. Additionally, AlexNet achieved 100% testing accuracy [77]. Another study

TABLE 1. Summary of deep and machine learning articles used for the classification of pistachio species.

Reference Number	Year	Methods	Result and Contributions
[41]	2020	Deeplab v3+, Segnet, Unet ve Hypernet	The proposed study uses Deeplab v3+, Segnet, Unet, and Hypernet for peanut identification.
[11]	2020	Random Forest	Based on the patterns in the peanut pixels in the image, the mass of the peanut kernels was found with 95% accuracy using the RF model.
[12]	2021	k-NN	Using k-NN and principal component analysis on the 16-feature peanut dataset, 94.18% accuracy was achieved in classification.
[13]	2021	PCA and LDA	In the study, PCA and LDA models were used together to classify different varieties.
[14]	2021	Xception, ResNet50, SqueezeNet, GoogLeNet, and AlexNet	In the study, five different deep-learning models were used to determine the quality of peanuts: Xception, ResNet50, SqueezeNet, GoogLeNet, and AlexNet.
[15]	2021	LSTM	The proposed study uses ORB to identify different pistachio tree species. The resulting features were trained on EfficientNet-B1.
[47]	2021	HSI	The proposed study uses the hyperspectral imaging (HSI) method to detect contaminants in peanuts.
[48]	2022	AlexNet, VGG16 and VGG19	In the proposed study, pistachios were classified with AlexNet, VGG16, and VGG19 models with a success rate of 94.42%, 98.84%, and 98.14%, respectively.
[49]	2022	A hybrid Decision Tree, Random Forest, k-nearest Neighboring, Support Vector Machines, Naive Bayes, and Artificial Neural Network	The study proposed a hybrid machine learning model that was built using a Decision Tree, Random Forest, k-nearest neighbors, Support Vector Machines, Naive Bayes, and Artificial Neural Network algorithms to classify the level of adulteration.
[50], [51]	2022	CNN, Random Forest, and MLP	In the study, peanut species were classified with the Convolutional Neural Network (CNN) model according to their acoustic emission signals. The CNN model achieved 98.75% accuracy with raw time data.
[52]	2022	EfficientNet-B3	In the study, EfficientNet-B3 architecture was used to distinguish peanut species. In classifying four different types, the EfficientNet-B3 architecture achieved average precision, recall, and F1 scores of 96.73%, 96.70%, and 96.67%, respectively.
[53]	2023	NAS	The model uses an adaptive ensemble distillation algorithm The method obtained coefficients for the feature maps. The proposed MNGNAS algorithm outperforms 0.09% decrease compared with the PCDARTS algorithm.
[54]	2023	AlexNet and Inception V3	In the study, AlexNet and Inception V3 structures were used to detect pistachios according to whether they were open or closed.
[55]	2023	Attention Pyramid Transformer	Lin et. get. proposed an Attention Pyramid Transformer for the position in patches. The proposed mode was tested using ResNet101, DeitT-s, Swin-T, EAPT-s structures.
[56], [57]	2023	VGG16, VGG19	In the study, VGG16 and VGG19 deep learning architectures were used to classify pistachios.
[49]	2023	ResNet-50, VGG-19, and DenseNet201	The study focused on analyzing pistachio images using various color spaces. They used ResNet-50, VGG-19, and DenseNet201 for classification purposes.
[58]	2023	RF, CatBoost, LightGBM, and Xgboost	In the proposed study, feature and vector representations were used to classify pistachios. The proposed model achieved a success rate of 97.20% using Logistic Regression.
[59]	2024	Mixed attention-based multi-scale feature calibration network	Li et al. solved the uneven blur distribution, which is difficult to remove in low-quality blurry images, with a mixed attention-based multi-scale feature calibration network.
[60]	2024	YOLOV7	In the proposed study, the classification of Gentiana szechenyii Kanitz plants in Tibet was made using YOLOv7. According to the results obtained, classification was made with an accuracy value of 97%.
[61]	2025	Light MobileNetV2	A lightweight MobileNetV2 plant disease classifier was used in the study. The model achieved an accuracy value of 96.67.

proposed a new method to detect adulteration in pistachios using digital image processing and machine learning. In the study, a dataset was created from images of adulterated pistachios. A hybrid machine learning model was created using a Decision Tree, Random Forest, k-nearest Neighbour, Support Vector Machines, Naive Bayes, and Artificial Neural Network algorithms for classification. According to the results, the artificial neural network algorithm achieved the highest accuracy in direct classification, while the random forest showed the best performance in binary classification [54]. In another study, a non-destructive robotic sorting system for pistachios was developed using an object detection algorithm. The proposed model used a conveyor dual-camera system for the physical classification of peanuts. Detection accuracy for open-shelled and closed-shelled peanuts was obtained as 98% and 85%, respectively. The proposed model can separate closed-shell peanuts for storage, cracking, or processing [57]. Researchers used VGG16 and VGG19 deep learning architectures to classify pistachios. VGG16 achieved an accuracy value of 0.80 and an F1 measurement value of 0.83. VGG19 achieved an average accuracy value of 0.77 and an F1 value of 0.77 [56]. In another study, adulteration analysis of pistachio images was performed using HSV and RGB in different color spaces. DenseNet201, VGGNet-19, and ResNet-50 were used to classify the images. The results showed that VGGNet-19 achieved 100% accuracy in classifying the LAB color space. Additionally, HSV/VGGNet-19 and YUV/ResNet-50 models classified peanut adulteration with 98% accuracy [49]. In another study, pistachios were classified using ResNet architecture. As a result of experimental studies, an accuracy value of 0.86 was obtained [56]. Kumar et al. proposed a deep neural network model to classify pistachios. The proposed model was trained on feature vectors and classified by a machine learning algorithm. In the study, the combined use of the Logistic Regression and Painters network achieved a success rate of 97.20% [78].

Previous studies reviewed by the author in the literature have separately examined machine learning, transfer learning-based deep learning models, and transformer-based models and achieved localized success. However, the potential synergy that can be obtained by combining these models has not been fully explored. In this study, we present the proposed hybrid model as an architecture that can be used in other applications by combining the power and computational efficiency of the attention mechanism with the speed of deep learning models. This study aims to achieve more effective learning and generalization by combining the advantages of both ViTEfficientNet, and TCA modules.

III. MATERIALS AND METHODS

A. DATASET

For the proposed method, we used the publicly available Pistachio dataset [79] created by Özkan et. al. This data set contains 2148 images of two different pistachio varieties, Kirmizi and Siirt. Of the total images, 1232 are of Kırmızıpistachios, and 916 of them are of Siirt pistachios. Although the dataset contains 2,148 images, it is considered sufficient for the scope of this study due to the clear inter-class visual differences and almost balanced distribution between the two pistachio varieties. Similar dataset sizes have been successfully used in related deep-learning studies focusing on agricultural crop classification [56], [73], [78]. Sample images of the dataset are given in Figure 1. Images of the Kirimizi and Siirt classes in the dataset are shown.

FIGURE 1. Sample Kirmizi and Siit pistachio images from the Pistachio dataset.

Table 2 shows the training and testing data distribution in the pistachio dataset.

TABLE 2. The distribution of the pistachio dataset for training and testing.

	Types	Images	Total data	Training (80%)	Test (20%)
Pistachio dataset	Kirmizi	1232	2.148	1718	430
	Siirt	916			

B. PERFORMANCE METRICS

Pistachio dataset binary classification was used. The performance of the proposed hybrid model was measured using True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values obtained from the confusion matrix as shown in equations (1), (2), (3), (4), (5), (6) and (7). The models' effectiveness was compared based on several metrics, including F1-Score (F1Sco), Accuracy (Acc), Sensitivity (Sen), and Precision (Precision), all of which were computed using equations (1), (2), (3), (4), (5), (6), and (7).

For a class k,

$$Sen(k) = \frac{TP(k)}{TP\,(k) + FN(k)} \tag{1}$$

$$F1Sco(k) = \frac{2 * Pre\,(k) * Sen(k)}{Pre\,(k) + Sen(k)} \tag{2}$$

$$Acc\,(k) = \frac{TP\,(k) + TN(k)}{TP\,(k) + FN\,(k) + TN\,(k) + FP(k)} \tag{3}$$

$$Pre(k) = \frac{TP(k)}{TP\,(k) + FP(k)} \tag{4}$$

$$AverageSen = \frac{1}{classes} \sum_{k=1}^{classess} \text{Sen(k)} \tag{5}$$

$$AverageAcc = \frac{1}{classes} \sum_{k=1}^{classess} \text{Acc(k)} \tag{6}$$

$$AveragePre = \frac{1}{classes} \sum_{k=1}^{classess} \text{Pre(k)} \tag{7}$$

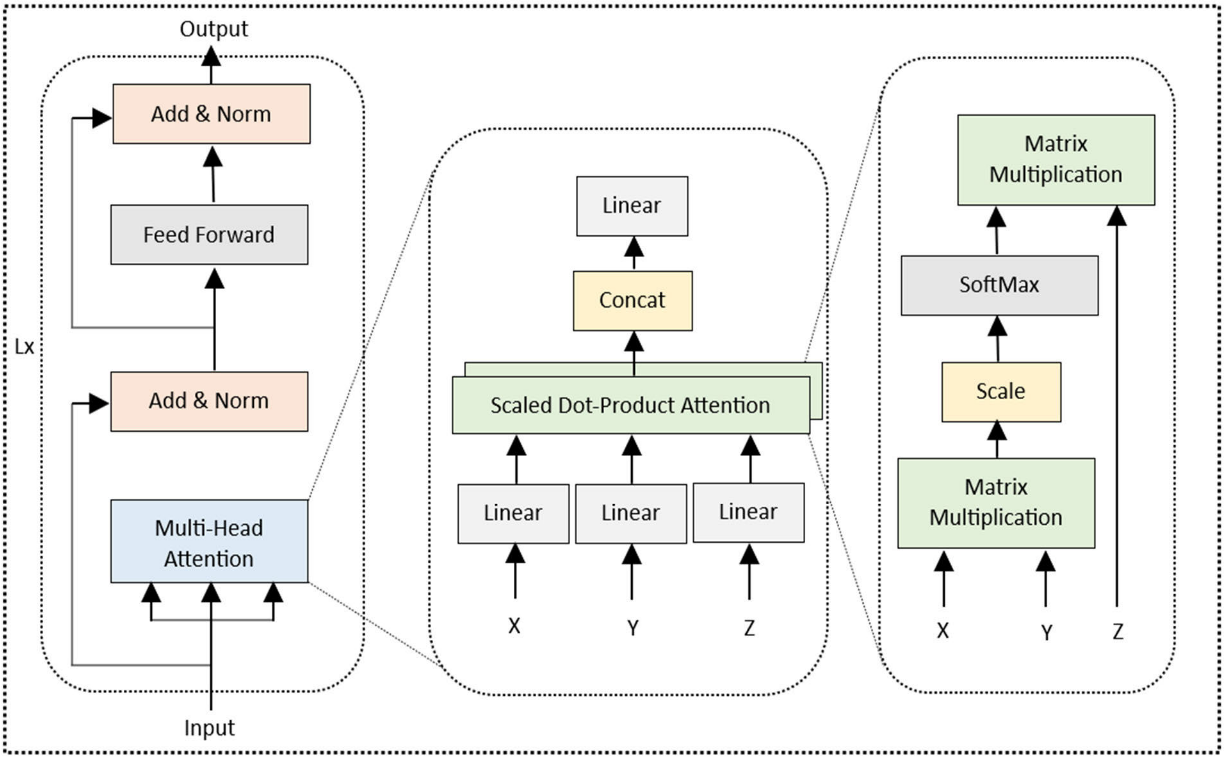

FIGURE 2. Structure of vision transformers.

C. PROPOSED METHODOLOGY

To illustrate the constituent steps of the proposed model, we first briefly introduce the Vision Transformer in Figure 2. In the second step, we show the TCA module in Figure 3 and the coupling of ViTs and TCA in Figure 4. Finally, in Figure 5, we introduce the EfficientTransformer model, a lightweight hybrid network model consisting of ViTs, TCA, and EfficentNet.

D. VISION TRANSFORMERS (ViTs)

CNNs are commonly used for capturing local information in images. However, their design often results in limited receptive fields, which makes them less effective at capturing global features. To overcome this limitation and utilize global information more efficiently, we incorporate the ViTs for holistic modelling. The ViTs act as the encoder component derived from the Transformer architecture. In the diagram below, the encoder consists of a series of L-consistent encoding layers placed in sequence. Each encoding layer consists of two primary sub-components: the initial sub-component uses a multi-head self-attention mechanism, while the subsequent sub-component utilizes a fully connected feedforward network. Figure 2 illustrates the structure of ViT and the multi-headed self-attention mechanism [80]. The mechanism is expressed mathematically with Eq. 7, Eq. 8, and Eq. 9.

$$Att\,(X, Y, Z) = softmax(\frac{XY^T}{\sqrt{d}_k}) \tag{8}$$

$$head_i = Att\left(XW_i^Q, YW_i^K, ZW_i^V\right) \tag{9}$$

$$MHead(X, Y, Z) = Concat\,(head_1, \ldots, head_i)\,W^o$$
$$\text{where } head_1 = Att(XW_i^Q, YW_i^K, ZW_i^V \tag{10}$$

d_k is a key vector of dimensions. W_i^Q, W_i^K, W_i^V express the transformation matrix, W^o is a linear transformation matrix and $head_i$ define the head of self-attention. Eq.7, Eq.8, and Eq.9 show encoder and decoder layers, and in addition to the attention sub-layers, there is a fully connected feed-forward network. This network works consistently across all positions, applying uniform transformations to each position. More specifically, it consists of two linear transformations with a ReLU activation function in between.

$$FFN\,(x) = max\,(0, xW_1+b_1)\,W_2+b_2 \tag{11}$$

Additionally, layer normalization is shown in Eq. 11 and Eq.12. Layer norms explain the normalization and sublayers show the fully connected feedforward and self-attention part.

$$y = Layernorm(x + sublayer(x)) \tag{12}$$

E. TEMPORAL-CHANNEL ATTENTION (TCA) MODULE

TCA refers to mechanisms emphasizing channel-specific attention over temporal dimensions in image or video data. This attention mechanism highlights the significance of channels (or feature maps) at distinct temporal instances or intervals within the data sequence. TCA is used as a mechanism to enhance inter-channel interaction after the fusion of ViT and EfficientNet outputs. These outputs, when combined, form a richer representation space in which dependencies across channels may reflect spatial, structural, or semantic hierarchies similar to temporal dependencies in sequence data. The attention mechanism benefits from the temporal abstraction capability to selectively focus on channel-wise patterns that are otherwise diffused or underrepresented in purely spatial modeling. The TCA approach is inspired by recent literature where 1D temporal evolutions and temporal growth attention modules [81], [82], [83] have been successfully introduced into static imaging tasks to enhance channel-wise reasoning. Figure 3 shows the structure of the TCA module.

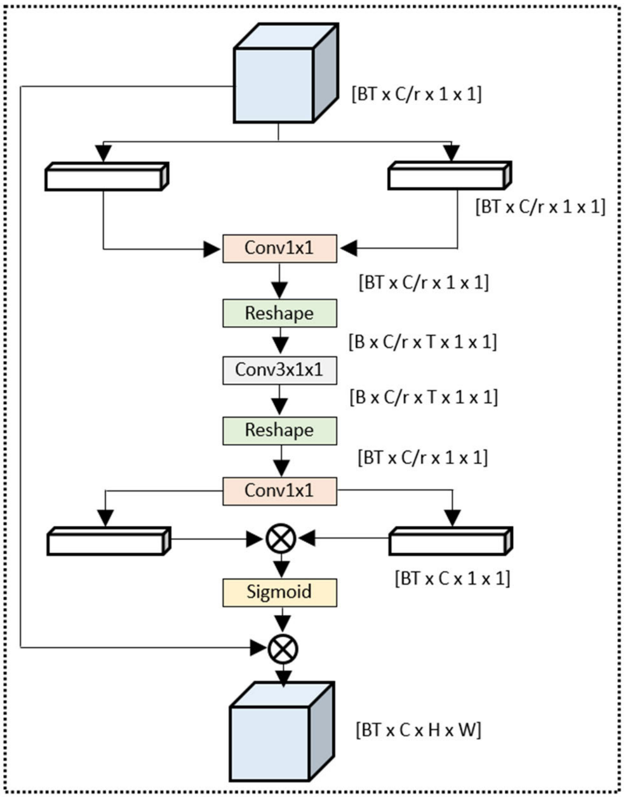

FIGURE 3. Structure of TCA module.

The module adopts a methodology that incorporates global maximum pooling and global average pooling techniques to extract the input feature map. $F \in R^{BTxCxHxW}$ refers to the 4-dimensional feature extraction map. $d_{tca-max}$ and $d_{tca-avg}$ denote input feature map (Eq.13 and Eq.14). The channel uses feeding 1×1 convolution. TCA convolution process benefits from reducing the number of channels, resulting in fewer parameters. This reduction is due to the interaction between dimensional transformation and temporal convolution. In the next step, the feature map size is loaded. Channels are remodelled with 1×1 convolution.

$$d_{tca-max} = MaxPool(F) \tag{13}$$

$$d_{tca-avg} = AvgPool(F) \tag{14}$$

The number of channels is also expanded to C. Temporal channel use $D_{tca-max}$ ve $D_{tca-max}$ attention weights. The feature map is weighted, and input features are feature-optimized across channel and time.

$$D_{tca-max} = Conv1 \times 1(relu\,(Conv3 \times 1\,x\,(relu\,(Conc1 \times 1\,(d_{tca-max})))) \tag{15}$$

$$D_{tca-avg} = Conv1 \times 1(relu\,(Conv3 \times 1\,x\,(relu\,(Conc1 \times 1\,(d_{tca-avg})))) \tag{16}$$

$$D_{tca} = sigmoid(D_{tca-max} + D_{tca-avg}) \tag{17}$$

F. PROPOSED HYBRID MODEL (EFFICIENTTRANSFORMERS)

The proposed model consists of three main structures and the combination of these structures. Running the EfficientNet model, taking the TCA input of the ViT model, and finally connecting the EfficientNet and ViT outputs to the TCA.

Algorithm of the proposed model:

Input resizing:

Initially, the input data $\boldsymbol{R^{BTxCxHxW}}$ *is resized to a format suitable for processing:* $\boldsymbol{R^{BTxCxHxW}} \rightarrow \boldsymbol{R^{BTxCxHxW}}$

Hybrid Block:

- *This block serves as the core component of the architecture. Let's denote the input tensor as X.*
- *Local feature extraction:*
  - $\boldsymbol{X_{local} = GroupConv(X, 3 \times 3 \times 3)}$
  - $\boldsymbol{X_{local} = GroupConv(X_{local}, 3 \times 1 \times 1)}$
- *ViT module for global information:*
  - $\boldsymbol{X_{global} = ViT\,(X)}$
  - $\boldsymbol{X_{global} = Conv(X_{global}, 3 \times 1 \times 1)}$
- *Future Vision:*
  - $\boldsymbol{X_{fusion} = Concatenate(X_{local}, X_{global})}$
  - $\boldsymbol{X_{fusion} = Conv(X_{fusion}, 3 \times 1 \times 1)}$
- *Residual Connection:*
  - $\boldsymbol{X_{residual} = X + X_{fusion}}$

TCA Module:

- *Let* $\boldsymbol{X_{residual}}$ *be the input to the TCA module.*
  - *The TCA module enhances temporal information interaction:* $\boldsymbol{X_{TCA} = TCA(X_{residual})}$

Feature Fusion and Residual Connection:

- *The output of the TCA module* $\boldsymbol{X_{TCA}}$ *is merged with the output of the Hybrid Block:* $\boldsymbol{X_{merged} = X_{residual} + X_{TCA}}$

Integration of EfficientNet Backbone:

- *The backbone network from EfficientNet is integrated into the Hybrid Block:*
- $\boldsymbol{X_{final} = HybridBlock(EfficientNet\,(X_{merged}))}$

Following this, the ViT module is utilized to acquire global information, and a $3 \times 1 \times 1$ convolution is employed for dimension reduction, thereby reducing computational complexity post-feature fusion. Subsequently, the local and global

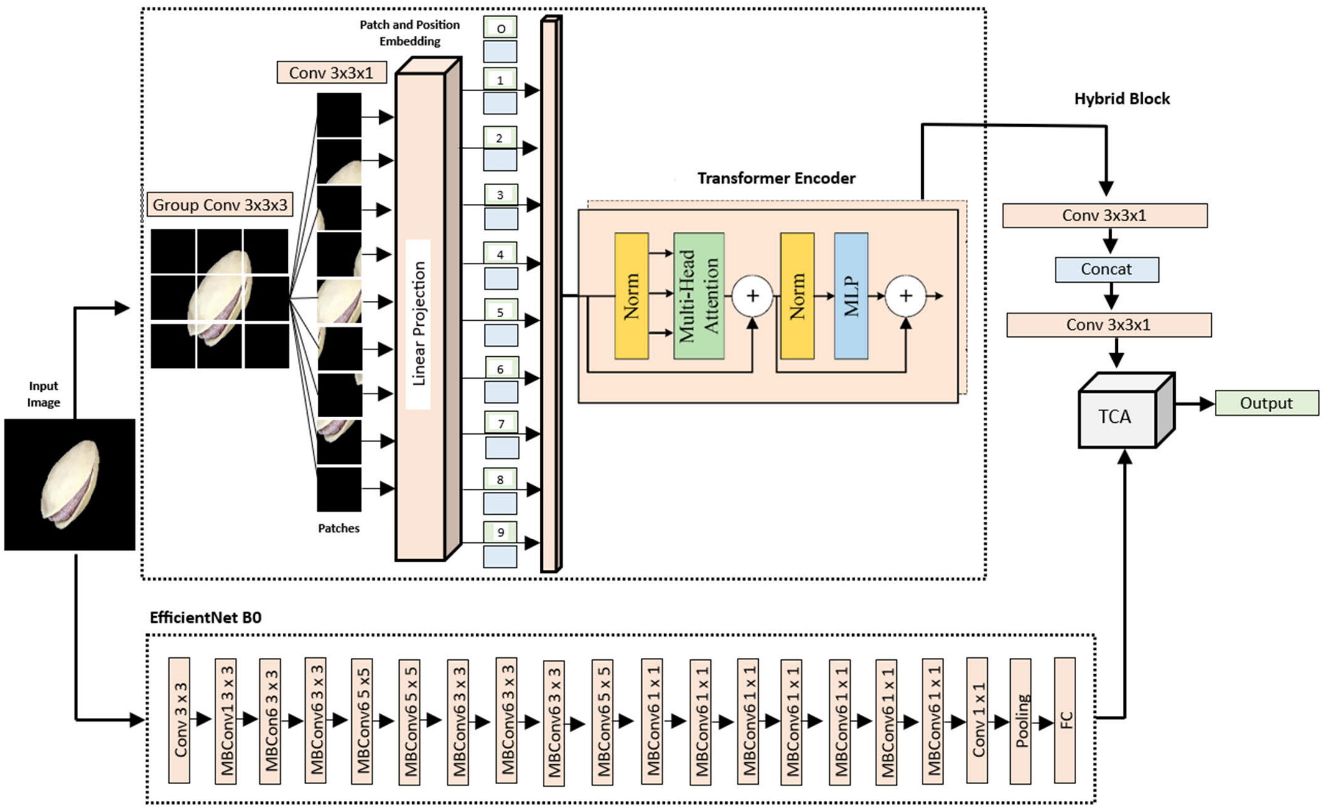

FIGURE 4. The proposed efficienttransformer model.

features are combined and integrated, with the channels being adjusted to match those of the input features through convolution. The TCA module is of utmost importance in enhancing the interaction of temporal information. Afterward, the input features are merged, and the model optimization is achieved by implementing a residual connection. This complex process enables the hybrid block to integrate information derived from global, local, and input features seamlessly, ultimately improving the overall model performance. The backbone network from EfficientNet is added to the hybrid block created by ViT. Thus, the TCA module combines the gains of both ViT and EfficientNet feature extraction. The TCA module simplifies the combined feature values. Additionally, the created hybrid block uses the ViT module, which models global information.

By following these steps, the model architecture effectively combines the features extracted by EfficientNet and ViT, enhanced by the TCA module, ultimately improving overall performance.

A simplified representation of the model is given in Figure 5. The architecture combines EfficientNetB0 for low-level feature extraction, a Transformer Encoder for capturing global dependencies, and a TCA module to enhance inter-channel relationships. The hybrid block merges output from both branches before classification, enabling robust and efficient image representation.

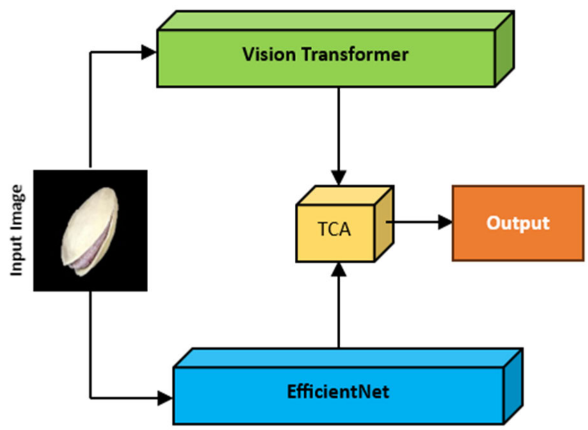

FIGURE 5. Simplified representation of the proposed model.

G. EXPERIMENTAL SETUP

The training for well-known deep learning models and the hybrid model was performed using an A5000 Tensor Core GPU with 64 GB memory. Python 3 was used for parallelization and transfer learning, making the methodology implementation easier. Pytorch was utilized as the training framework, which provides high-level APIs and tools for developing, fine-tuning, and evaluating complex neural network architectures. The proposed models were trained and optimized using a computer with an Intel i7 CPU, 64 GB

TABLE 3. Hyperparameter values of the compared models.

Hyperparameter	PiT	DeiT	ViT	CaiT	EfficientNet B0
Patch Size	16x16	16x16	16x16	16x16	-
Hidden Dimension	192	768	768	768	-
Number of Layers	12	12	12	24	-
Number of Attention Heads	3	12	12	8	-
MLP Ratio	4	4	4	-	-
Dropout Rate	0.1	0.1	0.1	0.1	0.2
Learning Rate	0.001 (Adam)	0.0005 (AdamW)	0.001 (AdamW)	0.0005 (AdamW)	0.001 (RMSProp)
Weight Decay	0.05	0.05	0.03	0.05	1e-5
Batch Size	32	32	32	32	32
Image Size	224x224	224x224	224x224	224x224	224x224
Optimizer	Adam	AdamW	AdamW	AdamW	RMSProp
Special Features	Pyramid Structure	Data-Efficient Architecture	Fully Attention-Based Mechanism	Class Attention Layers	Efficient Width and Depth Scaling

RAM, and 24GB GPU. Table 3 shows the hyperparameter values of the models.

The initial learning rate of the Adam optimizer used in the model was set to 0.01, and the minimum learning rate was set to 0.001. The proposed hybrid model was trained for 15 epochs. The images used in the training were resized to 224 × 224. To solve the problem of unbalanced class distribution and achieve smoother gradient descent [84], the binary cross-entropy (BCE) loss function was used. Eq 18. represents the working formula of binary cross entropy.

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(p_i\right) + (1-y_i)\log\left(1-p_i\right)\right] \quad (18)$$

IV. RESULTS AND DISCUSSIONS

A. STUDY ABLATION

In the first stage of the experimental study, an ablation analysis was performed for the proposed hybrid model. According to the analysis results shown in Table 4, the proposed Vision Transformer + EfficientNet + TCA hybrid model exhibited the best performance in terms of both accuracy and computational efficiency. While the EfficientNet model alone gave very successful results with 98.61% accuracy and 0.0139 MAE, it also provided a computationally efficient structure with an execution time of 293.31 seconds. On the other hand, the ViT model showed a lower success with 98.38% accuracy and 0.062 MAE and presented a very high computational cost with an execution time of 591.56 seconds. The hybrid structure, which combines the advantages of these two models and is enhanced with the TCA module, achieved the best result in all metrics with 99.07% accuracy, 0.0093 MAE, and 0.9810 Cohen's Kappa value, and was also the fastest model with an execution time of only 290.90 seconds. These findings show that the TCA module improves the learning quality with the channel-based attention mechanism, and the combination of EfficientNet and ViT achieves good results in terms of both accuracy and efficiency.

In the second stage of the experimental study, PiT, DeiT, CaiT, ViT, EfficientNet B0, and the proposed EfficientNet+Transformer+TCA methods were used to classify the pistachio image dataset, and the performances of the models were analyzed in detail. Performance evaluation was made according to accuracy, precision, recall, F1 score, Cohen Kappa, AUC-ROC, PR-ROC, and total running time. The comparative results of the models according to these metrics are given in Table 5.

TABLE 4. Ablation study of the compared models.

Model	Execution Time (seconds)	Accuracy (%)	MAE	Cohen's Kappa
EfficientNet	293.31	98.61	0.0139	0.9714
EfficientNet+TCA	283.01	98.84	0.0116	0.9763
ViT	591.56	98.38	0.062	0.9666
ViT+TCA				
ViT+EffNet+TCA	290.90	99.07	0.0093	0.9810

According to the experimental results, the proposed hybrid model achieved the highest classification success with 99.07% accuracy. The proposed model achieved the best result in classification stability with the lowest MAE value of 0.0093 and the Cohen Kappa value of 0.9810. The EfficientNet+Transformer+TCA model also produced consistent results in terms of PR-ROC and AUC-ROC. It has the shortest running time among all models with a running time of 290.90 seconds. In particular, the combination of the powerful feature extraction capacity of VisionTransformer and EfficientNet with the TCA module is the main element that distinguishes the model from other methods. The EfficientNet B0 model achieved the second-best results with 98.61% accuracy. It has the second-best fast running time (293.31 seconds). The feature extraction capability of EfficientNet B0 was effective in learning complex structures in the dataset, but it fell behind in critical metrics such as accuracy and F1 score compared to the proposed EfficientNet+Transformer+TCA model.

The CaiT model ranked third with 98.84% accuracy and 0.9881% F1 score. This model was successful in complex data structures by benefiting from the strong use of the

TABLE 5. Performance metrics of models.

Model	Execution Time (seconds)	Accuracy (%)	Precision	Recall	F1-Score	MAE	Cohen's Kappa	AUC-ROC	PR-AUC
DeiT	586.15	85.61	0.8707	0.8731	0.8561	0.1439	0.7174	0.5185	0.4790
PiT	526.46	96.98	0.9750	0.9647	0.9689	0.0302	0.9378	0.5255	0.4212
CaiT	604.95	98.84	0.9885	0.9878	0.9881	0.0116	0.9763	0.4872	0.4291
ViT	591.56	98.38	0.9862	0.9810	0.9833	0.062	0.9666	0.4847	0.4318
EfficientNet	293.31	98.61	0.9881	0.9837	0.9857	0.0139	0.9714	0.4532	0.5600
EfficientNet+Transformer+TCA	290.90	99.07	0.9912	0.9898	0.9905	0.0093	0.9810	0.4940	0.4476

attention mechanism. However, it was observed that it was less efficient than other models in terms of time with its running time (604.95 seconds). The PiT model showed a satisfactory performance with a 96.98% accuracy rate, but it was limited in Cohen Kappa (0.9378) and recall values. This situation shows that the model is less consistent in some classification cases. DeiT exhibited the lowest performance with an 85.61% accuracy rate. It achieved weak results, especially in metrics such as MAE value (0.1439) and Cohen Kappa (0.7174), and was insufficient in long-term classification tasks. The limited performance of this model may be due to the inability to learn distant contextual relations effectively. The proposed EfficientNet+Transformer+TCA model outperformed other models in terms of both high accuracy and time efficiency. This success was achieved by combining feature extraction and classification processes in an optimized hybrid framework. The classification success of the models used in the experimental study was analyzed via Confusion Matrix. For each model, correct classification (True Positives, True Negatives) and incorrect classification (False Positives, False Negatives) rates were calculated based on the confusion matrix and summarized in Table 6.

The EfficientNet+Transformer+TCA model has the highest correct classification rate (99.07% accuracy) among all models. The incorrect classification rates for Red and Siirt pistachios are quite low (False Positives: 1 Kirmizi, 3 Siirt). The model distinguished each of the pistachio types with a very high accuracy, which shows the efficiency-enhancing effect of the TCA module on the attention mechanism.

The EfficientNet B0 model produced no false positives in the Kirmizi pistachio class, while it made slightly more errors in the Siirt pistachio classes (False Negatives: 6). Nevertheless, it showed a strong performance with an overall accuracy rate of 98.61%. Especially fast processing times are one of the advantages of this model. The CaiT model has a low number of false positive and negative classifications (False Positives: 2 Kirmizi, 3 Siirt). The CaiT model showed a performance close to the EfficientNet B0 model with a 98.84% accuracy rate. However, long processing times limit the practicality of the model. Although the PiT model has a high true positive rate for the Kirmizi pistachio, it made more errors in the Siirt pistachio classes (False Negatives: 13). The DeiT model is the model with the lowest performance. It produced a high false positive rate (60), especially in the Kirmizi pistachio class. This shows that the model is inadequate in learning distant contextual relationships and thus causes its overall accuracy to decrease.

The training loss and accuracy values of the ViT, DeiT, CaiT, PiT, EfficientNet B0, and EfficientNet+Transformer+TCA hybrid models used in the experimental study were examined. The training and loss graphs obtained depending on the epoch number for each model are shown in Figure 6.

According to the experimental results, the proposed EfficientNet+Transformer+TCA model achieved the highest classification success with 99.0According to the performance analysis of the models, ViT showed a constant training loss decrease and stabilized after the 6th epoch and almost reached the best accuracy level by the 8th epoch. This shows that the model converged effectively and exhibited a balanced performance. Although the DeiT model showed a rapid decrease at the beginning, it exhibited significant fluctuations throughout the training process. Despite the sharp increases in training accuracy, there were decreases, suggesting that the model tended to overfit or had sensitivity problems to data distribution. The CaiT model stabilized with slight fluctuations after a rapid training loss decrease and achieved high accuracy in the early period. However, the stabilization process was slower compared to other models. The PiT model, on the other hand, showed a slow training loss decrease with sharp fluctuations and was subject to fluctuations even though it exhibited a regular increase in training accuracy. This situation shows that the model needs improvement in convergence efficiency. EfficientNet B0 stabilized rapidly with a smooth and consistent training loss decrease and achieved high training accuracy. The EfficientNet+Transformer+TCA hybrid model achieved the

TABLE 6. Performance metrics of models.

Model	True Positives (Kirmizi)	True Positives (Siirt)	False Positives (Kirmizi)	False Positives (Siirt)	False Negatives (Kirmizi)	False Negatives (Siirt)
EfficientNet+Transformer+TCA	246	181	1	3	1	3
ViT	247	177	0	7	0	7
EfficientNet B0	247	178	0	6	0	6
CaiT	245	181	2	3	2	3
PiT	247	171	0	13	0	13
DeiT	187	182	60	2	60	2

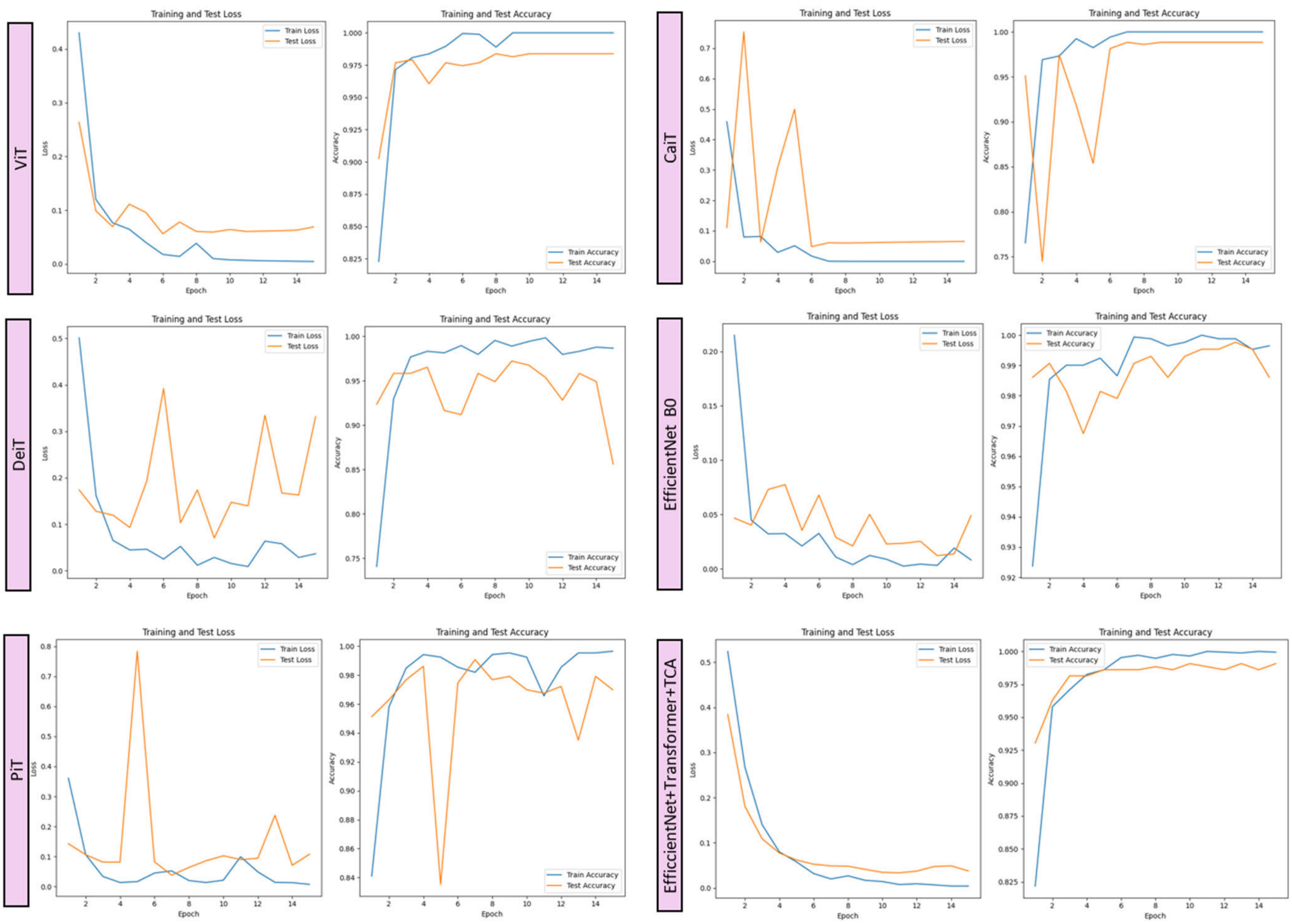

FIGURE 6. Training and testing loss and accuracy graphs of models.

lowest training loss and the highest accuracy among all models. This model achieved high stability and accuracy by integrating the lightweight structure of EfficientNet with Transformer-based attention mechanisms (TCA). The sharp decrease in training loss and early convergence in the accuracy trend significantly increased the learning dynamics and prediction performance of the model. Experimental studies have shown that the EfficientNet+Transformer+TCA hybrid model achieved the best performance metrics compared to other models. The proposed model has low training loss, high accuracy, and strong stabilization values. Combining the lightweight and efficient structure of EfficientNet used for hybridization with Transformer-based attention mechanisms increased both computational efficiency and classification accuracy.

The proposed EfficientNet+Transformer+TCA model has a few key features that make it effective. First, the TCA module allows for interaction between temporal information, prevents gradient disappearance, and reduces overfitting. The hybrid block starts by extracting local features using $3 \times 3 \times 3$ group convolution and $3 \times 1 \times 1$ convolution and then maps the tensor to a high-dimensional space. n order to obtain global information from the ViT module and reduce the computational complexity, $3 \times 1 \times 1$ dimension reduction convolution is applied. Then local and global features were combined. Channels are adjusted to match input features via convolution. The TCA module is used to improve temporal information interaction. Thus, information from input, local, and global features are combined. EfficientNet's backbone network is added to the hybrid block created by ViT. The TCA module combined the gains of both ViT and EfficientNet feature extraction. By simplifying the combined feature values, the TCA module enabled model optimization through residual linkage.

Figure 7 shows the comparison of ROC values of the models. According to the AUC-ROC analysis, when the discrimination power of the models between classes was evaluated, the PiT model showed the highest performance (0.5255), which shows that it can distinguish true positives from false positives well. Similarly, the DeiT model offers a competitive discrimination ability with an AUC-ROC score of 0.5185, while the overall accuracy rate remained lower. Despite having high accuracy, the ViT (0.4847) and CaiT (0.4872) models revealed that they are less flexible against decision threshold changes. EfficientNet, despite providing a high accuracy of 98.61%, has the lowest AUC-ROC score (0.4532), which shows that its discrimination between classes may be limited. The proposed hybrid model (EfficientNet+Transformer+TCA) exhibits an average discrimination performance with an AUC score of 0.4940, and provides the most successful results in general classification metrics; this shows that the model is sensitive to decision threshold optimization and that there is room for improvement in ROC-based evaluations.

The EfficientNet+Transformer+TCA hybrid model, which includes the ViT structure, can adapt to different input sizes. It can also achieve high performance when trained with larger data sets and models. ViT also has transfer learning capabilities. It allows the same model to be easily used for a variety of visual tasks. The integration of attention mechanisms in ViTs was effectively implemented on Graphics Processing Units (GPUs), providing a speed advantage in the training and inference processes. The combination of ViT's global modeling and EfficientNet's local perception resulted in high accuracy and robustness. ViT's attention mechanisms provided an alternative perspective to image processing tasks, while EfficientNet effectively balanced computational efficiency and model size. The proposed hybrid model enables

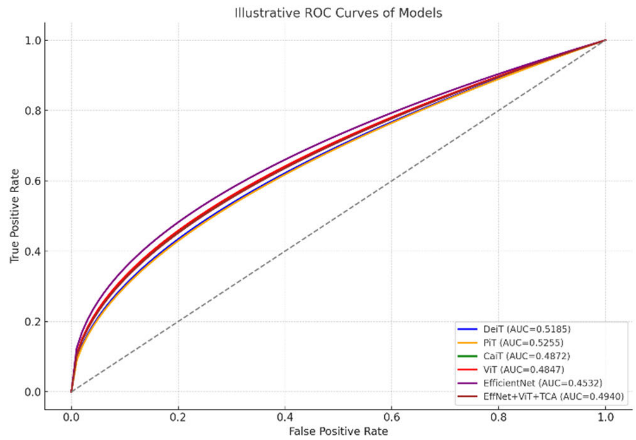

FIGURE 7. Comparison of ROC curves of models.

the interaction of temporal information from the Vision Transformer. The EfficientNet models with TCA help prevent the gradient from disappearing and reduce overfitting over time.

The comparative computational analysis in Table 7 reveals that the proposed hybrid model offers a balanced structure in terms of both accuracy and efficiency.

TABLE 7. Performance of model efficiency metrics.

Model Component	FLOPs (GFLops)	Parameters (Millions)
CaiT	17.3	46.5
DeiT	0.40	85.8
PiT	8.97	72.7
ViT	4.26	86.0
EfficientNetB0	0.39	4.0
TCA	0.12	2.4
EfficientNet+Transformer+TCA	4.77	92.4

EfficientNetB0 stands out as the model with the lowest computational cost with only 0.39 GFLOPs and 4.0 million parameters. On the other hand, although the ViT model provides higher accuracy with 4.26 GFLOPs and 86 million parameters, the computational load has increased significantly. CaiT has the highest computational complexity by far with 17.3 GFLOPs, and this high FLOP value is due to the detailed attention mechanisms and depth structure of the model. In terms of the number of parameters, it is more efficient than other transformer-based models with 46.5 million. PiT and DeiT operate with 8.97 and 0.40 GFLOPs, respectively. Although DeiT seems particularly efficient in terms of FLOPs, it is a fairly large model with 85.8 million parameters. The TCA module offers a lightweight structure with only 0.12 GFLOPs and 2.4 million parameters. Despite this, TCA significantly contributes to classification accuracy by increasing the channel-based attention power of the model. The EfficientNet+Transformer+TCA hybrid model, which

TABLE 8. Comparison of the proposed model with other models in the literature.

Study	Year	References	Methods	Accuracy (%)
Strawberry disease identification	2023	[85]	Spatial convolutional self-attention-based transformer	99.10
Small green apple detection	2023	[86]	FBoT-Net: Focal bottleneck transformer network	93.33
Segmentation of grape leaf diseases	2023	[87]	Local Reversible Transformer	98.99
Image detection method for grassland	2023	[88]	Transformer-based model “Am-mask”	95.71
Identification of cotton and corn plant	2023	[89]	Deep transformer encoder	95.00
Quality evaluation of postharvest nuts	2024	[90]	Deep modular combination optimization (DMCO) algorithm	94.50
Peanut Classification	2024	[91]	Dual-aspect attention spatial-spectral transformer	99.40
Detection of peanut kernels	2024	[92]	Multi-scale attention transformer	98.42
Plant disease detection and classification	2025	[61]	Light MobileNetV2	96.67
Our Study	2025		EfficientNet+Transformer+TCA	99.20

is a combination of all these components, exhibits a balanced structure with 4.77 GFLOPs and 92.4 million parameters, offering an optimal solution in terms of both accuracy and efficiency. Table 8 shows the comparison results of the proposed model with other models in the literature.

V. CONCLUSION

This study comprehensively compared the performance of different deep learning models ViT, DeiT, CaiT, PiT, EfficientNet B0, and the proposed EfficientNet+Transformer+TCA hybrid model on the classification task on the pistachio images dataset. The analysis of the models in terms of training loss, accuracy, Cohen Kappa, AUC-ROC, and PR-AUC metrics shows that the proposed hybrid model has superior performance compared to other models. The EfficientNet+Transformer+TCA model optimized the stability and overall classification success during the training process by providing both low training loss and high accuracy. Combining the lightweight and fast structure of EfficientNet with Transformer-based attention mechanisms (TCA), this model increased the efficiency of the learning process and was able to effectively model long-term relationships. Among other models, ViT showed satisfactory performance in terms of overall accuracy and stability. However, DeiT limited its potential to be a reliable classification model by showing instability and fluctuations during the training process. CaiT and PiT models provided competitive results in terms of accuracy, but fluctuations in training loss revealed the need for improvement in convergence processes. EfficientNet B0, on the other hand, exhibited a stable performance, but could not reach the accuracy and stability level of the hybrid model.

As a result, the proposed EfficientNet+Transformer+TCA hybrid model provides an effective solution for image classification tasks that require high accuracy and stability.

This model is suitable for use in both academic and industrial applications and exhibits a practical and scalable structure, especially in real-world applications. Future studies can investigate the applicability of this hybrid model in different data sets and problem areas and contribute to the further development of the model.

ACKNOWLEDGMENT

The authors declares that they have no competing interests or relationships that could potentially influence their work. This study was conducted using the data models and computational resources provided by “Sinop University Faculty of Engineering and Architecture Software and Artificial Intelligence Laboratory.” They extend their sincere gratitude to the laboratory for its invaluable support and contribution to the research. They confirm sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation. They give ethical and informed consent for the data used. During the preparation of this work, they used Grammarly and ChatGPT in order to grammar check improve readability and language. After using these tools, the author reviewed and edited the content as needed and took full responsibility for the content of the publication. Data codes can be publicly accessible on https://github.com/muhammetcakmak/efftransformers.git accounts.

REFERENCES

[1] R. Shahriaripour and A. Tajabadipour, “Zinc nutrition of pistachio: Interaction of zinc with other trace elements,” *Commun. Soil Sci. Plant Anal.*, vol. 41, no. 15, pp. 1885–1888, Jul. 2010, doi: 10.1080/00103624.2010.492445.

[2] E. Arjeh, H.-R. Akhavan, M. Barzegar, and Á. A. Carbonell-Barrachina, “Bio-active compounds and functional properties of pistachio hull: A review,” *Trends Food Sci. Technol.*, vol. 97, pp. 55–64, Mar. 2020, doi: 10.1016/j.tifs.2019.12.031.

[3] A. Rabadán and Á. Triguero, "Influence of food safety standards on trade: Evidence from the pistachio sector," *Agribusiness*, vol. 37, no. 3, pp. 489–514, Jul. 2021, doi: 10.1002/agr.21683.

[4] M. Arabameri, M. Mohammadi Moghadam, L. Monjazeb Marvdashti, S. M. Mehdinia, A. Abdolshahi, and A. Dezianian, "Pesticide residues in pistachio nut: A human risk assessment study," *Int. J. Environ. Anal. Chem.*, vol. 102, no. 16, pp. 3947–3960, Dec. 2022, doi: 10.1080/03067319.2020.1777289.

[5] E. Nezami-Alanagh, G.-A. Garoosi, M. Landín, and P. P. Gallego, "Combining DOE with neurofuzzy logic for healthy mineral nutrition of pistachio rootstocks in vitro culture," *Frontiers Plant Sci.*, vol. 871, p. 1474, Oct. 2018, doi: 10.3389/fpls.2018.01474.

[6] Z. Rahneshan, F. Nasibi, and A. A. Moghadam, "Effects of salinity stress on some growth, physiological, biochemical parameters and nutrients in two pistachio (Pistacia vera L.) rootstocks," *J. Plant Interact.*, vol. 13, no. 1, pp. 73–82, Jan. 2018, doi: 10.1080/17429145.2018.1424355.

[7] M. A. Hajam, T. Arif, A. M. U. D. Khanday, M. A. Wani, and M. Asim, "AI-driven pattern recognition in medicinal plants: A comprehensive review and comparative analysis," *Comput., Mater. Continua*, vol. 81, no. 2, pp. 2077–2131, Nov. 2024, doi: 10.32604/cmc.2024.057136.

[8] D. Barhate, S. Pathak, B. K. Singh, A. Jain, and A. K. Dubey, "A systematic review of machine learning and deep learning approaches in plant species detection," *Smart Agricult. Technol.*, vol. 9, Dec. 2024, Art. no. 100605, doi: 10.1016/j.atech.2024.100605.

[9] F. Zapf and T. Wallek, "Case-study of a flowsheet simulation using deep-learning process models for multi-objective optimization of petrochemical production plants," *Comput. Chem. Eng.*, vol. 162, Jun. 2022, Art. no. 107823, doi: 10.1016/j.compchemeng.2022.107823.

[10] P. Hari and M. P. Singh, "Adaptive knowledge transfer using federated deep learning for plant disease detection," *Comput. Electron. Agricult.*, vol. 229, Feb. 2025, Art. no. 109720, doi: 10.1016/j.compag.2024.109720.

[11] M. Mhamed, Z. Zhang, W. Hua, L. Yang, M. Huang, X. Li, T. Bai, H. Li, and M. Zhang, "Apple varieties and growth prediction with time series classification based on deep learning to impact the harvesting decisions," *Comput. Ind.*, vol. 164, Jan. 2025, Art. no. 104191, doi: 10.1016/j.compind.2024.104191.

[12] Z. Lin, S. Yoshikawa, M. Hamasaki, K. Kikuchi, and S. Hosoya, "Automated phenotyping empowered by deep learning for genomic prediction of body size in the tiger pufferfish, takifugu rubripes," *Aquaculture*, vol. 595, Jan. 2025, Art. no. 741491, doi: 10.1016/j.aquaculture.2024.741491.

[13] I. Tahyudin, H. A. A. Rozaq, H. Nambo, R. Rozak, and A. Tikaningsih, "Deep learning algorithm for room temperature detection using bioelectric potential of plant data," *Biomed. Signal Process. Control*, vol. 101, Mar. 2025, Art. no. 107214, doi: 10.1016/j.bspc.2024.107214.

[14] F. B. Tilahun, "Fuzzy-based predictive deep reinforcement learning for robust and constrained optimal control of industrial solar thermal plants," *Appl. Soft Comput.*, vol. 159, Jul. 2024, Art. no. 111432, doi: 10.1016/j.asoc.2024.111432.

[15] Z. Zhang, M. Yang, Q. Pan, X. Jin, G. Wang, Y. Zhao, and Y. Hu, "Identification of tea plant cultivars based on canopy images using deep learning methods," *Scientia Horticulturae*, vol. 339, Jan. 2025, Art. no. 113908, doi: 10.1016/j.scienta.2024.113908.

[16] R. Sujatha, J. M. Chatterjee, N. Jhanjhi, and S. N. Brohi, "Performance of deep learning vs machine learning in plant leaf disease detection," *Microprocessors Microsyst.*, vol. 80, Feb. 2021, Art. no. 103615, doi: 10.1016/j.micpro.2020.103615.

[17] C. Jackulin and S. Murugavalli, "A comprehensive review on detection of plant disease using machine learning and deep learning approaches," *Meas., Sensors*, vol. 24, Dec. 2022, Art. no. 100441, doi: 10.1016/j.measen.2022.100441.

[18] I. Ahmed and P. K. Yadav, "A systematic analysis of machine learning and deep learning based approaches for identifying and diagnosing plant diseases," *Sustain. Oper. Comput.*, vol. 4, pp. 96–104, Jan. 2023, doi: 10.1016/j.susoc.2023.03.001.

[19] O. A. Talab and I. Avci, "Energy management in microgrids using model-free deep reinforcement learning approach," *IEEE Access*, vol. 13, pp. 5871–5891, 2025, doi: 10.1109/ACCESS.2025.3525843.

[20] J. P. Vásconez, I. N. Vásconez, V. Moya, M. J. Calderón-Díaz, M. Valenzuela, X. Besoain, M. Seeger, and F. A. Cheein, "Deep learning-based classification of visual symptoms of bacterial wilt disease caused by ralstonia solanacearum in tomato plants," *Comput. Electron. Agricult.*, vol. 227, Dec. 2024, Art. no. 109617, doi: 10.1016/j.compag.2024.109617.

[21] M. A. Kiflie, D. P. Sharma, and M. A. Haile, "Deep learning for Ethiopian indigenous medicinal plant species identification and classification," *J. Ayurveda Integrative Med.*, vol. 15, no. 6, Nov. 2024, Art. no. 100987, doi: 10.1016/j.jaim.2024.100987.

[22] D. Ma, H. Fang, N. Wang, H. Lu, J. Matthews, and C. Zhang, "Transformer-optimized generation, detection, and tracking network for images with drainage pipeline defects," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 38, no. 15, pp. 2109–2127, Oct. 2023, doi: 10.1111/mice.12970.

[23] D. T. N. Nhut, T. D. Tan, T. N. Quoc, and V. T. Hoang, "Medicinal plant recognition based on vision transformer and BEiT," *Proc. Comput. Sci.*, vol. 234, pp. 188–195, Jan. 2024, doi: 10.1016/j.procs.2024.02.165.

[24] A. K. Singh, A. Rao, P. Chattopadhyay, R. Maurya, and L. Singh, "Effective plant disease diagnosis using vision transformer trained with leafy-generative adversarial network-generated images," *Expert Syst. Appl.*, vol. 254, Nov. 2024, Art. no. 124387, doi: 10.1016/j.eswa.2024.124387.

[25] S. Vallabhajosyula, V. Sistla, and V. K. K. Kolli, "A novel hierarchical framework for plant leaf disease detection using residual vision transformer," *Heliyon*, vol. 10, no. 9, May 2024, Art. no. e29912, doi: 10.1016/j.heliyon.2024.e29912.

[26] W. Xia, L. Pu, X. Zou, P. Shilane, S. Li, H. Zhang, and X. Wang, "The design of fast and lightweight resemblance detection for efficient post-deduplication delta compression," *ACM Trans. Storage*, vol. 19, no. 3, pp. 1–30, Aug. 2023, doi: 10.1145/3584663.

[27] J. Xu, S. H. Park, and X. Zhang, "A temporally irreversible visual attention model inspired by motion sensitive neurons," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 595–605, Jan. 2020, doi: 10.1109/TII.2019.2934144.

[28] A. G. Alharthi and S. M. Alzahrani, "Do it the transformer way: A comprehensive review of brain and vision transformers for autism spectrum disorder diagnosis and classification," *Comput. Biol. Med.*, vol. 167, Dec. 2023, Art. no. 107667, doi: 10.1016/j.compbiomed.2023.107667.

[29] D. P. C. Le, D. Wang, and V.-T. Le, "A comprehensive survey of recent transformers in image, video and diffusion models," *Comput., Mater. Continua*, vol. 80, no. 1, pp. 37–60, Jul. 2024, doi: 10.32604/cmc.2024.050790.

[30] S. Islam, H. Elmekki, A. Elsebai, J. Bentahar, N. Drawel, G. Rjoub, and W. Pedrycz, "A comprehensive survey on applications of transformers for deep learning tasks," *Expert Syst. Appl.*, vol. 241, May 2024, Art. no. 122666, doi: 10.1016/j.eswa.2023.122666.

[31] E. Nezami and P. P. Gallego, "History, phylogeny, biodiversity, and new computer-based tools for efficient micropropagation and conservation of pistachio (Pistacia spp.) germplasm," *Plants*, vol. 12, no. 2, p. 323, Jan. 2023, doi: 10.3390/plants12020323.

[32] L. Pierre Pott, T. Jorge Carneiro Amado, R. Augusto Schwalbert, G. Mateus Corassa, and I. Antonio Ciampitti, "Crop type classification in Southern Brazil: Integrating remote sensing, crop modeling and machine learning," *Comput. Electron. Agricult.*, vol. 201, Oct. 2022, Art. no. 107320, doi: 10.1016/j.compag.2022.107320.

[33] Z. Gao, D. Guo, D. Ryu, and A. W. Western, "Training sample selection for robust multi-year within-season crop classification using machine learning," *Comput. Electron. Agricult.*, vol. 210, Jul. 2023, Art. no. 107927, doi: 10.1016/j.compag.2023.107927.

[34] H. Wang, Z. Ye, Y. Wang, X. Liu, X. Zhang, Y. Zhao, S. Li, Z. Liu, and X. Zhang, "Improving the crop classification performance by unlabeled remote sensing data," *Expert Syst. Appl.*, vol. 236, Feb. 2024, Art. no. 121283, doi: 10.1016/j.eswa.2023.121283.

[35] Y. Zhang, G. Zhou, A. Chen, M. He, J. Li, and Y. Hu, "A precise apple leaf diseases detection using BCTNet under unconstrained environments," *Comput. Electron. Agricult.*, vol. 212, Sep. 2023, Art. no. 108132, doi: 10.1016/j.compag.2023.108132.

[36] J. Li, Y. Qiao, S. Liu, J. Zhang, Z. Yang, and M. Wang, "An improved YOLOv5-based vegetable disease detection method," *Comput. Electron. Agricult.*, vol. 202, Nov. 2022, Art. no. 107345, doi: 10.1016/j.compag.2022.107345.

[37] R. Abbasi, P. Martinez, and R. Ahmad, "Crop diagnostic system: A robust disease detection and management system for leafy green crops grown in an aquaponics facility," *Artif. Intell. Agricult.*, vol. 10, pp. 1–12, Dec. 2023, doi: 10.1016/j.aiia.2023.09.001.

[38] R. Gauthier, C. Largouët, and J.-Y. Dourmad, "Prediction of litter performance in lactating sows using machine learning, for precision livestock farming," *Comput. Electron. Agricult.*, vol. 196, May 2022, Art. no. 106876, doi: 10.1016/j.compag.2022.106876.

[39] F. Yang, D. Zhang, Y. Zhang, Y. Zhang, Y. Han, Q. Zhang, Q. Zhang, C. Zhang, Z. Liu, and K. Wang, "Prediction of corn variety yield with attribute-missing data via graph neural network," *Comput. Electron. Agricult.*, vol. 211, Aug. 2023, Art. no. 108046, doi: 10.1016/j.compag.2023.108046.

[40] M. Hayat, "Squeeze & excitation joint with combined channel and spatial attention for pathology image super-resolution," *Franklin Open*, vol. 8, Sep. 2024, Art. no. 100170, doi: 10.1016/j.fraope.2024.100170.

[41] D. Pacheco-Paramo and L. Tello-Oquendo, "Delay-aware dynamic access control for mMTC in wireless networks using deep reinforcement learning," *Comput. Netw.*, vol. 182, Dec. 2020, Art. no. 107493, doi: 10.1016/j.comnet.2020.107493.

[42] Z. Chen, T. Gao, B. Sheng, P. Li, and C. L. P. Chen, "Outdoor shadow estimating using multiclass geometric decomposition based on BLS," *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 2152–2165, May 2020, doi: 10.1109/TCYB.2018.2875983.

[43] M. Koklu and I. A. Ozkan, "Multiclass classification of dry beans using computer vision and machine learning techniques," *Comput. Electron. Agricult.*, vol. 174, Jul. 2020, Art. no. 105507, doi: 10.1016/j.compag.2020.105507.

[44] M. Esteki, E. Heydari, J. Simal-Gandara, Z. Shahsavari, and M. Mohammadlou, "Discrimination of pistachio cultivars based on multi-elemental fingerprinting by pattern recognition methods," *Food Control*, vol. 124, Jun. 2021, Art. no. 107889, doi: 10.1016/j.foodcont.2021.107889.

[45] J. Gao, J. Ni, H. Yang, and Z. Han, "Pistachio visual detection based on data balance and deep learning," *Nongye Jixie Xuebao/Trans. Chin. Soc. Agricult. Machinery*, vol. 52, no. 7, pp. 367–372, 2021, doi: 10.6041/j.issn.1000-1298.2021.07.040.

[46] A. Heidary-Sharifabad, M. S. Zarchi, S. Emadi, and G. Zarei, "An efficient deep learning model for cultivar identification of a pistachio tree," *Brit. Food J.*, vol. 123, no. 11, pp. 3592–3609, Oct. 2021, doi: 10.1108/bfj-12-2020-1100.

[47] G. Bonifazi, G. Capobianco, R. Gasbarrone, and S. Serranti, "Contaminant detection in pistachio nuts by different classification methods applied to short-wave infrared hyperspectral images," *Food Control*, vol. 130, Dec. 2021, Art. no. 108202, doi: 10.1016/j.foodcont.2021.108202.

[48] A. Chug, A. Bhatia, A. P. Singh, and D. Singh, "A novel framework for image-based plant disease detection using hybrid deep learning approach," *Soft Comput.*, vol. 27, no. 18, pp. 13613–13638, Sep. 2023, doi: 10.1007/s00500-022-07177-7.

[49] G. Çinarer, N. Doğan, K. Kılıç, and C. Doğan, "Rapid detection of adulteration in pistachio based on deep learning methodologies and affordable system," *Multimedia Tools Appl.*, vol. 83, no. 5, pp. 14797–14820, Jul. 2023, doi: 10.1007/s11042-023-16172-5.

[50] M. Hosseinpour-Zarnaq, M. Omid, A. Taheri-Garavand, A. Nasiri, and A. Mahmoudi, "Acoustic signal-based deep learning approach for smart sorting of pistachio nuts," *Postharvest Biol. Technol.*, vol. 185, Mar. 2022, Art. no. 111778, doi: 10.1016/j.postharvbio.2021.111778.

[51] A. Soleimanipour, M. Azadbakht, and A. Rezaei Asl, "Cultivar identification of pistachio nuts in bulk mode through EfficientNet deep learning model," *J. Food Meas. Characterization*, vol. 16, no. 4, pp. 2545–2555, Aug. 2022, doi: 10.1007/s11694-022-01367-5.

[52] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using EfficientNet deep learning model," *Ecological Informat.*, vol. 61, Mar. 2021, Art. no. 101182, doi: 10.1016/j.ecoinf.2020.101182.

[53] Z. Chen, G. Qiu, P. Li, L. Zhu, X. Yang, and B. Sheng, "MNGNAS: Distilling adaptive combination of multiple searched networks for one-shot neural architecture search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13489–13508, Nov. 2023, doi: 10.1109/TPAMI.2023.3293885.

[54] C. Doğan, E. Şehirli, N. Doğan, and İ. Buran, "Non-targeted approach to detect pistachio authenticity based on digital image processing and hybrid machine learning model," *J. Food Meas. Characterization*, vol. 17, no. 2, pp. 1693–1702, Apr. 2023, doi: 10.1007/s11694-022-01671-0.

[55] X. Lin, S. Sun, W. Huang, B. Sheng, P. Li, and D. D. Feng, "EAPT: Efficient attention pyramid transformer for image processing," *IEEE Trans. Multimedia*, vol. 25, pp. 50–61, 2023, doi: 10.1109/TMM.2021.3120873.

[56] E. Avuçlu, "Classification of pistachio images with the ResNet deep learning model," *Selcuk J. Agricult. Food Sci.*, vol. 37, no. 2, pp. 291–300, Aug. 2023, doi: 10.15316/sjafs.2023.029.

[57] A. E. Karadağ and A. Kılıç, "Non-destructive robotic sorting of cracked pistachio using deep learning," *Postharvest Biol. Technol.*, vol. 198, Apr. 2023, Art. no. 112229, doi: 10.1016/j.postharvbio.2022.112229.

[58] G. Dhanush, N. Khatri, S. Kumar, and P. K. Shukla, "A comprehensive review of machine vision systems and artificial intelligence algorithms for the detection and harvesting of agricultural produce," *Sci. Afr.*, vol. 21, Sep. 2023, Art. no. e01798, doi: 10.1016/j.sciaf.2023.e01798.

[59] L. Li, Z. Chen, L. Dai, R. Li, and B. Sheng, "MA-MFCNet: Mixed attention-based multi-scale feature calibration network for image dehazing," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 8, no. 5, pp. 3408–3421, Oct. 2024, doi: 10.1109/tetci.2024.3382233.

[60] B. Chang, Y. Wang, X. Zhao, G. Li, and P. Yuan, "A general-purpose edge-feature guidance module to enhance vision transformers for plant disease identification," *Expert Syst. Appl.*, vol. 237, Mar. 2024, Art. no. 121638, doi: 10.1016/j.eswa.2023.121638.

[61] S. Duhan, P. Gulia, N. S. Gill, and E. Narwal, "RTR_Lite_MobileNetV2: A lightweight and efficient model for plant disease detection and classification," *Current Plant Biol.*, vol. 42, Jun. 2025, Art. no. 100459, doi: 10.1016/j.cpb.2025.100459.

[62] P. Porwal et al., "IDRiD: Diabetic retinopathy–segmentation and grading challenge," *Med. Image Anal.*, vol. 59, Jan. 2020, Art. no. 101561, doi: 10.1016/j.media.2019.101561.

[63] M. Hayat, N. Ahmad, A. Nasir, and Z. Ahmad Tariq, "Hybrid deep learning EfficientNetV2 and vision transformer (EffNetV2-ViT) model for breast cancer histopathological image classification," *IEEE Access*, vol. 12, pp. 184119–184131, 2024, doi: 10.1109/access.2024.3503413.

[64] S. B. A. Kartbak, M. B. Özel, D. N. C. Kocakaya, M. Çakmak, and E. A. Sinanoğlu, "Classification of intraoral photographs with deep learning algorithms trained according to cephalometric measurements," *Diagnostics*, vol. 15, no. 9, p. 1059, Apr. 2025, doi: 10.3390/diagnostics15091059.

[65] M. Çakmak and Z. Albayrak, "AFCC-r: Adaptive feedback congestion control algorithm to avoid queue overflow in LTE networks," *Mobile Netw. Appl.*, vol. 27, no. 5, pp. 2138–2152, Oct. 2022, doi: 10.1007/s11036-022-02011-8.

[66] S. Ibrahim, A. M. Youssef, M. Shoman, and S. Taha, "Intelligent SDN to enhance security in IoT networks," *Egyptian Informat. J.*, vol. 28, Dec. 2024, Art. no. 100564, doi: 10.1016/j.eij.2024.100564.

[67] H. C. Altunay and Z. Albayrak, "SMS spam detection system based on deep learning architectures for Turkish and English messages," *Appl. Sci.*, vol. 14, no. 24, p. 11804, Dec. 2024, doi: 10.3390/app142411804.

[68] B. Sheng, P. Li, Y. Zhang, L. Mao, and C. L. P. Chen, "GreenSea: Visual soccer analysis using broad learning system," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1463–1477, Mar. 2021, doi: 10.1109/TCYB.2020.2988792.

[69] M. Vimala, S. Palanisamy, S. Guizani, and H. Hamam, "Efficient GDD feature approximation based brain tumour classification and survival analysis model using deep learning," *Egyptian Informat. J.*, vol. 28, Dec. 2024, Art. no. 100577, doi: 10.1016/j.eij.2024.100577.

[70] Z. Liu, J. Jiang, X. Qiao, X. Qi, Y. Pan, and X. Pan, "Using convolution neural network and hyperspectral image to identify moldy peanut kernels," *LWT*, vol. 132, Oct. 2020, Art. no. 109815, doi: 10.1016/j.lwt.2020.109815.

[71] B. Pourmohammadali, M. H. Salehi, S. J. Hosseinifard, I. Esfandiarpour Boroujeni, and H. Shirani, "Studying the relationships between nutrients in pistachio leaves and its yield using hybrid GA-ANN model-based feature selection," *Comput. Electron. Agricult.*, vol. 172, May 2020, Art. no. 105352, doi: 10.1016/j.compag.2020.105352.

[72] S. K. Vidyarthi, R. Tiwari, S. K. Singh, and H.-W. Xiao, "Prediction of size and mass of pistachio kernels using random forest machine learning," *J. Food Process Eng.*, vol. 43, no. 9, p. 13473, Sep. 2020, doi: 10.1111/jfpe.13473.

[73] D. Singh, Y. S. Taspinar, R. Kursun, I. Cinar, M. Koklu, I. A. Ozkan, and H.-N. Lee, "Classification and analysis of pistachio species with pre-trained deep learning models," *Electronics*, vol. 11, no. 7, p. 981, Mar. 2022, doi: 10.3390/electronics11070981.

[74] M. Rahimzadeh and A. Attar, "Detecting and counting pistachios based on deep learning," *Iran J. Comput. Sci.*, vol. 5, no. 1, pp. 69–81, Mar. 2022, doi: 10.1007/s42044-021-00090-6.

[75] R. An, J. Perez-Cruet, and J. Wang, "We got nuts! Use deep neural networks to classify images of common edible nuts," *Nutrition Health*, vol. 30, no. 2, pp. 301–307, Jul. 2024, doi: 10.1177/02601060221113928.

[76] C. Saglam and N. Cetin, "Prediction of pistachio (Pistacia vera L.) mass based on shape and size attributes by using machine learning algorithms," *Food Anal. Methods*, vol. 15, no. 3, pp. 739–750, Mar. 2022, doi: 10.1007/s12161-021-02154-6.

[77] H. Aktaş, T. Kızıldeniz, and Z. Ünal, "Classification of pistachios with deep learning and assessing the effect of various datasets on accuracy," *J. Food Meas. Characterization*, vol. 16, no. 3, pp. 1983–1996, Jun. 2022, doi: 10.1007/s11694-022-01313-5.

[78] S. S. Kumar, A. Sigappi, G. A. S. Thomas, Y. H. Robinson, and S. P. Raja, "Classification and analysis of pistachio species through neural embedding-based feature extraction and small-scale machine learning techniques," *Int. J. Image Graph.*, vol. 24, no. 3, May 2024, Art. no. 2450032, doi: 10.1142/s0219467824500323.

[79] I. A. Özkan, M. Köklü, and R. Saraçoğlu, "Classification of pistachio species using improved k-NN classifier," *Prog. Nutrition*, vol. 23, no. 2, p. 9686, Feb. 2021, doi: 10.23751/pn.v23i2.9686.

[80] K. Chen, Z. Bi, X. Song, Q. Niu, J. Liu, B. Peng, S. Zhang, M. Liu, X. Li, T. Wang, X. Pan, J. Xu, C. Jiang, J. Wang, and P. Feng, "Transformer: Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2024, pp. 1–11.

[81] Y. Liu, X. Huang, L. Xiong, R. Chang, W. Wang, and L. Chen, "Stock price prediction with attentive temporal convolution-based generative adversarial network," *Array*, vol. 25, Mar. 2025, Art. no. 100374, doi: 10.1016/j.array.2025.100374.

[82] H. Zhu, L. Yang, W. Ding, and Z. Han, "Pixel-level rapid detection of aflatoxin B1 based on 1D-modified temporal convolutional network and hyperspectral imaging," *Microchemical J.*, vol. 183, Dec. 2022, Art. no. 108020, doi: 10.1016/j.microc.2022.108020.

[83] M. He, L. Zhu, and L. Bai, "Jointly leveraging 1D and 2D convolution on diachronic entity embedding for temporal knowledge graph completion," *Appl. Soft Comput.*, vol. 176, May 2025, Art. no. 113144, doi: 10.1016/j.asoc.2025.113144.

[84] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, in Lecture Notes in Computer Science, Sep. 2017, pp. 240–248, doi: 10.1007/978-3-319-67558-9_28.

[85] G. Li, L. Jiao, P. Chen, K. Liu, R. Wang, S. Dong, and C. Kang, "Spatial convolutional self-attention-based transformer module for strawberry disease identification under complex background," *Comput. Electron. Agricult.*, vol. 212, Sep. 2023, Art. no. 108121, doi: 10.1016/j.compag.2023.108121.

[86] M. Sun, R. Zhao, X. Yin, L. Xu, C. Ruan, and W. Jia, "FBoT-net: Focal bottleneck transformer network for small green apple detection," *Comput. Electron. Agricult.*, vol. 205, Feb. 2023, Art. no. 107609, doi: 10.1016/j.compag.2022.107609.

[87] X. Zhang, F. Li, H. Jin, and W. Mu, "Local reversible transformer for semantic segmentation of grape leaf diseases," *Appl. Soft Comput.*, vol. 143, Aug. 2023, Art. no. 110392, doi: 10.1016/j.asoc.2023.110392.

[88] Y. Zhang, T. Wang, Y. You, D. Wang, J. Gao, and T. Liang, "A transformer-based image detection method for grassland situation of Alpine meadows," *Comput. Electron. Agricult.*, vol. 210, Jul. 2023, Art. no. 107919, doi: 10.1016/j.compag.2023.107919.

[89] R. Ş. Bağcı, E. Acar, and Ö. Türk, "Identification of cotton and corn plant areas by employing deep transformer encoder approach and different time series satellite images: A case study in diyarbakir, Turkey," *Comput. Electron. Agricult.*, vol. 209, Jun. 2023, Art. no. 107838, doi: 10.1016/j.compag.2023.107838.

[90] Z. Guo, H. Wang, H. Dong, L. Xia, I. A. Darwish, Y. Guo, and X. Sun, "Optimizing aflatoxin B1 detection in peanut kernels through deep modular combination optimization algorithm: A deep learning approach to quality evaluation of postharvest nuts," *Postharvest Biol. Technol.*, vol. 220, Feb. 2025, Art. no. 113293, doi: 10.1016/j.postharvbio.2024.113293.

[91] Z. Guo, J. Zhang, H. Wang, S. Li, X. Shao, H. Dong, J. Sun, L. Geng, Q. Zhang, Y. Guo, X. Sun, L. Xia, and I. A. Darwish, "Dual-aspect attention spatial–spectral transformer and hyperspectral imaging: A novel approach to detecting aspergillus flavus contamination in peanut kernels," *Postharvest Biol. Technol.*, vol. 213, Jul. 2024, Art. no. 112960, doi: 10.1016/j.postharvbio.2024.112960.

[92] Z. Guo, J. Zhang, H. Wang, H. Dong, S. Li, X. Shao, J. Huang, X. Yin, Q. Zhang, Y. Guo, X. Sun, and I. Darwish, "Enhanced detection of aspergillus flavus in peanut kernels using a multi-scale attention transformer (MSAT): Advancements in food safety and contamination analysis," *Int. J. Food Microbiol.*, vol. 423, Oct. 2024, Art. no. 110831, doi: 10.1016/j.ijfoodmicro.2024.110831.

MUHAMMET ÇAKMAK received the master's and Ph.D. degrees with the Department of Computer Engineering. He is currently holding a position as an Assistant Professor with the Department of Computer Engineering, Sinop University. His research interests include network optimization, machine learning, deep learning, and natural language.