Volume 1 Issue 2 2023
Pages: 139-148
ISSN: 2980-3152

CURRENT TRENDS IN COMPUTING

http://ctc.karabuk.edu.tr
RESEARCH ARTICLE

AUTOMATIC CLASSIFICATION OF WALNUT LEAF IMAGES WITH GRADCAM AND DEEP LEARNING

ALPER TALHA KARADENİZ[1], ERDAL BAŞARAN[2] AND YÜKSEL ÇELİK[3]

[1]*Digital Transformation and Software Office, Trabzon University, 61000, Trabzon, Türkiye*
[2]*Department of Computer Technology, Agri Ibrahim Cecen University, 04000, Agri, Türkiye*
[3]*Computer Engineering Department, Karabük University, 78050, Karabük, Türkiye*

ABSTRACT: Walnut leaves' similar color and formation distinguishing between varieties is considerably challenging for individuals. Examining and categorizing such plant leaves one by one can be a time-consuming and costly process. Hence, experimental studies are conducted in the laboratory to classify walnut varieties. Within the scope of this study, an original dataset consisting of 1751 walnut leaf images obtained from 18 different walnut varieties was prepared. Various preprocessing techniques were applied to the original dataset, and additionally, data augmentation methods were employed to obtain an expanded dataset. Both datasets were trained using deep learning models. Among these models, the Vgg16 CNN model which has shown the most superior performance. The proposed model, trained with Vgg16 on the augmented dataset, produced Gradcam images and was further classified using the Vgg16 CNN algorithm. According to experimental results, the proposed model achieved a success rate of 77.11%. This study demonstrates the successful utilization of deep learning techniques for classifying walnut varieties from walnut leaf images.

1. INTRODUCTION

Plants hold a critical position among the world's resources, demanding the imperative sustainability of these resources for the future [1]. This could enhance productivity and effectiveness. The precise identification of different types of walnuts is of utmost significance to agricultural engineers and producers within the walnut industry.

Analyzing plant leaves requires meticulous inspection of each leaf. It can be arduous for the human eye to differentiate between plant species that share similar color and morphology. The manual examination and classification of plant leaves across diverse categories prove to be inefficient in both time and cost. Consequently, computer-based automated techniques for diagnosing and categorizing leaf images present substantial benefits in terms of time and expenses [2].

In the realm of walnut cultivation, the precise selection of the appropriate walnut seedling holds immense importance. Once an orchard is established, a significant duration, approximately 3-5 years, is required before it starts yielding fruits. During this period, dealing with unwanted walnut varieties or wild trees might be necessary. Additionally, tasks like irrigation, fertilization, pruning, and disease management demand additional cost and time. Deal with such unfavorable circumstances, a producer may opt for a variety change or reorganize

E-mail address: alperkaradeniz@trabzon.edu.tr (*)
Key words and phrases. Walnut dataset, GradCam, Deep learning, Classification.

the orchard. Furthermore, in orchards established with unnamed or mixed varieties, fruits cannot be standardized for the market since they are harvested mixed. This complicates consumers' ability to obtain products according to their desired quality, yield, or fruit weight. Hence, selecting the right variety and proper seedling holds paramount importance in establishing a walnut orchard capable of producing for at least 700 years. During this phase, walnut seedlings are typically grafted at the end of August or the beginning of September; the graft union merges with the rootstock, but it doesn't develop. Identifying variety from leaves emerging from subsequent graft buds in late April to early May provides crucial information on variety confusion or whether the seedlings are wild, preventing potential legal issues in advance [3]. Throughout the history of walnut cultivation, numerous studies exist in the literature regarding the differentiation and standardization of walnut varieties [4].

Functions performed by the human eye can be executed faster and more accurately through artificial intelligence techniques [5].

The innovative aspects of this study and its contributions to the literature are summarized as follows.

- It is one of the pioneering studies in literature to classify walnut varieties from leaf images.
- The original dataset consisting of 18 different species and 1751 walnut leaf images was introduced.
- Grad-cam and Vgg16 CNN models were examined in detail.
- The feature extraction process from CNN models is explained.

This paper proceeds as follows: Section 2 presents recent studies in the literature related to this study are mentioned. Dataset, applied methods, and the proposed model are introduced in Section 3. Section 4 presents the results obtained from the experimental studies. Finally, the paper concludes in Section 5.

2. RELATED WORKS

Feature extraction constitutes a pivotal phase within the machine learning process, exerting a profound influence on classification performance. Notably, the enhanced capabilities of Central Processing Units (CPUs) and Graphics Processing Units (GPUs) due to technological progressions have bolstered speed and capacity. These advancements have significantly augmented data processing capabilities and facilitated the evolution of deep learning architectures [6]. Simultaneously, research within the domain of deep learning has been dedicated to exploring its potential in detecting and classifying plant leaf diseases [7].

Vasif Nabiyev et al. introduced a methodology employing Convolutional Neural Networks (CNNs) and Transfer Learning for plant identification. Their approach involved fine-tuning from the ImageNet domain for transfer learning in the Oxford Flowers Dataset study. The study utilized the pre-trained MobileNetV2 network from the ImageNet database, achieving an impressive success rate of 0.9897. Furthermore, the development of a mobile application yielded positive outcomes [8].

Ibtesam M. Dheir et al. classified a dataset comprising five distinct walnut types from a collection of 2868 images. The model architecture incorporated 4 convolutional layers utilizing the Rectified Linear Unit (ReLU) activation function. The analysis conducted by Ibtesam M. Dheir et al. on walnut leaf classification involved a model that consisted of sequential layers, including a Max Pooling layer followed by a flattening layer. Within the subsequent layers, there were 512 hidden units, resulting in a total of 2,603,205 trainable parameters in the network. Their approach achieved a notable success rate of 0.98 [9].

Yixue Liu et al. delved into the classification of 21 distinct grape leaves by implementing preprocessing steps and leveraging CNN algorithms. Their study introduced the Grad-CAM algorithm to assess the impact of complementary image preprocessing on classification outcomes. Employing the

Googlenet model led to a commendable test success rate of 0.974 [10].

Daniel Nkemelu et al. aimed to classify 12 different plant seedlings. Their investigation encompassed tests involving diverse methods such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and a self-developed CNN model. By incorporating various preprocessing techniques and CNN algorithms, their study achieved the highest accuracy rate of 0.926 [11].

In an initial study using the dataset, a novel proposal was presented: a model based on the ResNet architecture. This proposed approach integrated the ResNet architecture during the feature extraction phase, utilized the Atom Search Optimization algorithm for feature selection, and employed SVM (Support Vector Machine) for classification. Experimental tests resulted in a success rate of 81.77% [12].

Another study concerning the same dataset involved the individual training and testing of nine commonly used CNN models. The investigation concluded that the Vgg16 model attained the highest success rate [13].

Producers must exercise great care in selecting suitable seedlings for walnut cultivation. Distinguishing the correct walnut variety from the leaves is a challenging process even for experts. Fruit-bearing from a walnut tree usually takes between 3 to 5 years. During this time, a producer might spend on an undesirable walnut variety or fail to yield under suitable regional conditions. In such cases, the producer typically attempts variety changes or a complete redevelopment of the walnut orchard. Legal disputes have arisen in numerous places due to uncertain seedlings, leading to issues between producers and nurseries. Every year, there is a series of legal cases between producers who purchased incorrect walnut variety seedlings and nurseries.

This study employed CNN (Convolutional Neural Network) models to identify walnut varieties. Preprocessing methods were applied to the dataset before presenting it to CNN models, and experimental test results were compared. Data augmentation was applied to enhance the model's performance, repeating experimental tests and comparing the results. It aims to prevent unwanted issues such as acquiring and planting wrong seedlings, losses until fruit yield is achieved, losses from seedlings not planted in suitable climatic conditions, and potential legal proceedings. In this manner, more effective establishment and completion of walnut orchards could be possible.

3. MATERIALS AND METHODS

3.1 Dataset

Our dataset comprises of 1751 leaf images sourced from 18 distinct walnut varieties. These images were procured from walnut trees situated in the Yalova Atatürk Horticultural Application Garden. To capture these images, walnut leaves were carefully removed from identified tree branches and positioned against a white background for imaging purposes. The Canon EOS 100D model camera was utilized to capture these images, employing daylight conditions and close-up shots. The camera settings were configured on white balance mode automatically, and the image resolution stood at 18 megapixels.

All the leaves were harvested from trees of recognized varieties on the same day, ensuring consistency in the dataset collection. The image acquisition process was efficiently completed within a few hours, employing a standardized procedure for imaging each leaf. Nevertheless, slight variations might exist due to factors such as sunlight angles, shadows, and the position of the individual taking the images. Figure 1 presents samples showcasing the 18 walnut species encompassed within the dataset. During the creation of the walnut dataset, special emphasis was placed on capturing images that highlighted more prominent textural and morphological features of the leaves. This emphasis was crucial for subsequent feature extraction from the leaf images. In leaf-image-based classification, the distinctiveness of morphological traits such as size, texture, shape, and vein patterns hold paramount importance [14].

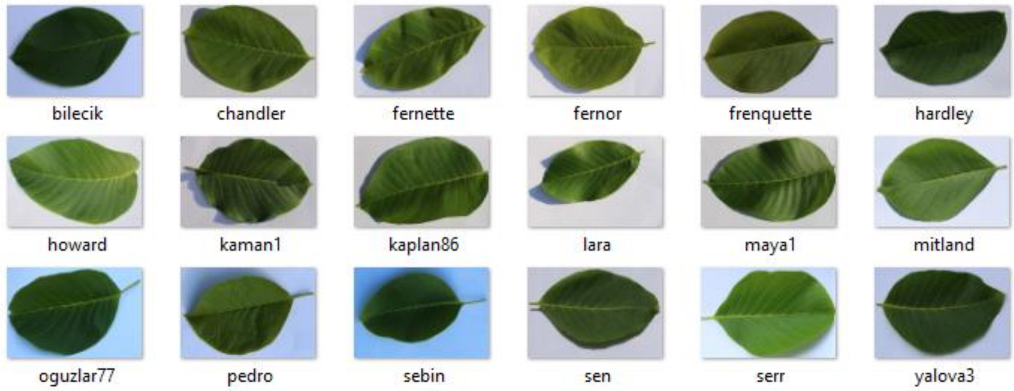

FİGURE 1. Samples of 18 Different Walnut Leaf Varieties in the Collected Dataset.

In various research studies, data augmentation techniques have been commonly employed to bolster a model's ability to generalize using deep learning and machine learning methods [15]. In this particular investigation, data augmentation was utilized as a strategy to enhance the performance metrics of the proposed model. By implementing operations like rotation, brightness adjustment, shifting, zooming, and flipping, the number of images in the dataset increased from 1751 to 6606. Tables 1 and 2 present the original walnut varieties along with their corresponding leaf counts in the original dataset and the augmented leaf counts for each walnut variety in the expanded dataset, respectively.

TABLE 1. The numbers corresponding to the 18 Walnut varieties in the collected dataset.

Variety Name	**Number of Images**	**Variety Name**	**Number of Images**
Bilecik	96	Lara	63
Chandler	82	Maya1	74
Fernette	89	Mitland	147
Fernor	104	Oguzlar77	59
Frenquette	126	Pedro	77
Hardley	95	Sebin	88
Howard	85	Sen	157
Kaman1	98	Ser	119
Kaplan86	84	Yalova3	108
		Total	1751

TABLE 2. The numbers corresponding to the 18 walnut varieties in the augmented dataset.

Variety Name	**Number of Images**	**Variety Name**	**Number of Images**
Bilecik	378	Lara	249
Chandler	324	Maya1	291
Fernette	351	Mitland	418
Fernor	406	Oguzlar77	232
Frenquette	488	Pedro	303
Hardley	375	Sebin	344
Howard	332	Sen	528
Kaman1	389	Ser	441
Kaplan86	330	Yalova3	427
		Total	6606

3.2 Gradient-Weighted Class Activation Mapping (Grad-Cam)

Grad-Cam is utilized to understand which regions of an image are most crucial for classification based on the convolutional features identified by the network, using the gradient of the classification score. It is a method that employs heatmaps to highlight distinctive features of an image when describing a given image. To rank the importance of a feature, the average of all elements of the gradient can be taken. While Grad-Cam is generally an effective method, it may not be suitable for sensitive data and models [16].

Grad-Cam can be used in the layers of any deep learning model. The mathematical formula is as follows in Equation 1 [17].

$$w_k^c = \frac{1}{z} \sum_{i=0}^{n} \sum_{j=1}^{h} \frac{\partial y^c}{\partial A_{ij}^k} \qquad (1)$$

Where y^c denotes the score of class c and the dimension of Ak is $W \times H$. By differential operation of y^c with respect to Ak, the weight of the map Ak for class c is calculated and denoted as w_k^c. Z is the normalization factor. The architecture of the Grad-Cam method is shown in Figure 2.

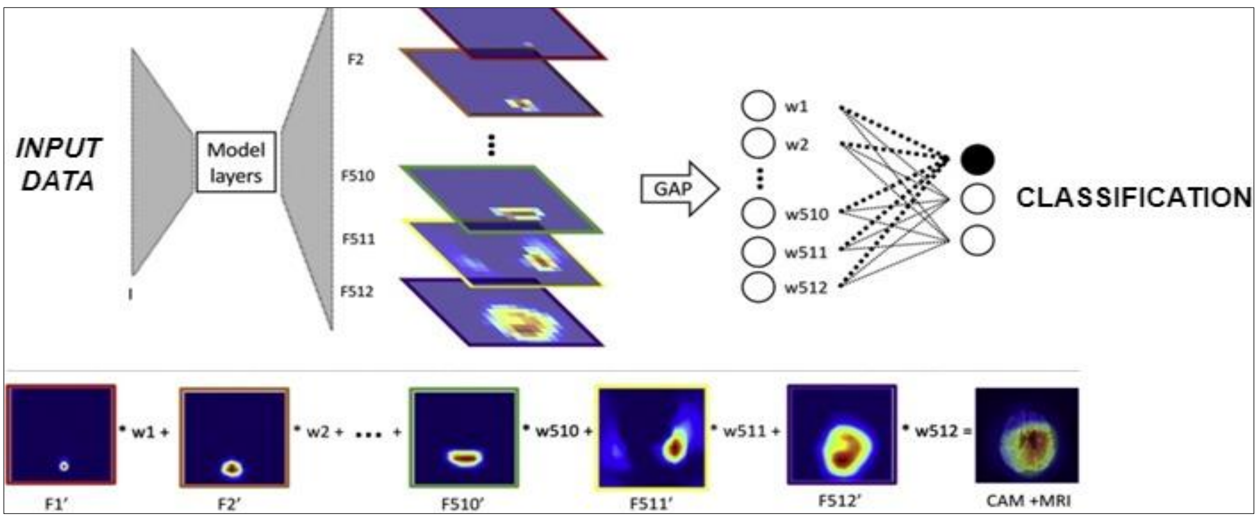

FIGURE 2. The architecture of GRADCAM [18].

3.3 VGG16

The Vgg16 (Visual Geometry Group 16) is a Convolutional Neural Network (CNN) architecture that was introduced in 2014, encompassing approximately 138 million parameters. This architecture diverges from using an extensive array of hyper-parameters and instead employs $3 x 3$ filters combined with $2 x 2$ pooling operations at each stage. Within the fully connected layer, there are a total of 3 layers, with the first two utilizing the Rectified Linear Unit (ReLU) activation function and the last one employing Softmax. Vgg16 comprises 16 layers, and its input layer is specifically designed to process images sized at 224×224 pixels [19]. The schematic representation of the general architecture of Vgg16 is depicted in Figure 3.

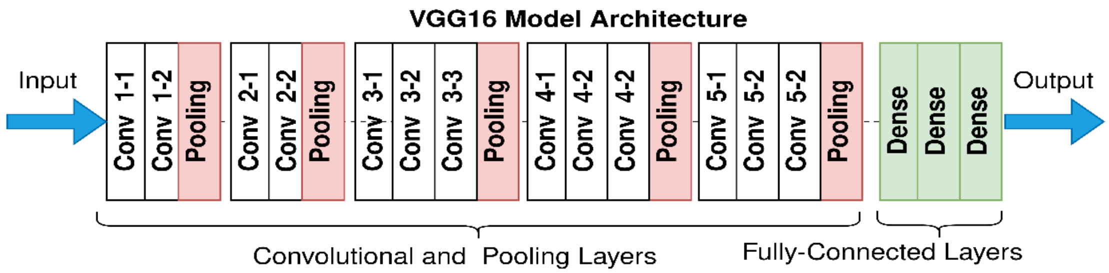

FIGURE 3. The architecture of VGG16.

3.4 Proposed Model

In the proposed model, Vgg16 CNN model was used to obtain Grad-Cam images to select salient features of the images. Because it is the CNN model that achieved the highest success in the study conducted with the data set used in the study [13].

Grad-Cam highlights distinctive features by coloring salient features in images with heat maps [16]. When obtaining Grad-Cam images with Vgg16, the most prominent features can be detected from many layers of Vgg16. The layer commonly used in the literature is the last convolution layer, usually known as conv5_3. In this study, features obtained from the same layer were used. Grad-Cam image examples created by applying pre-processing and data augmentation are shown in Figure 4.

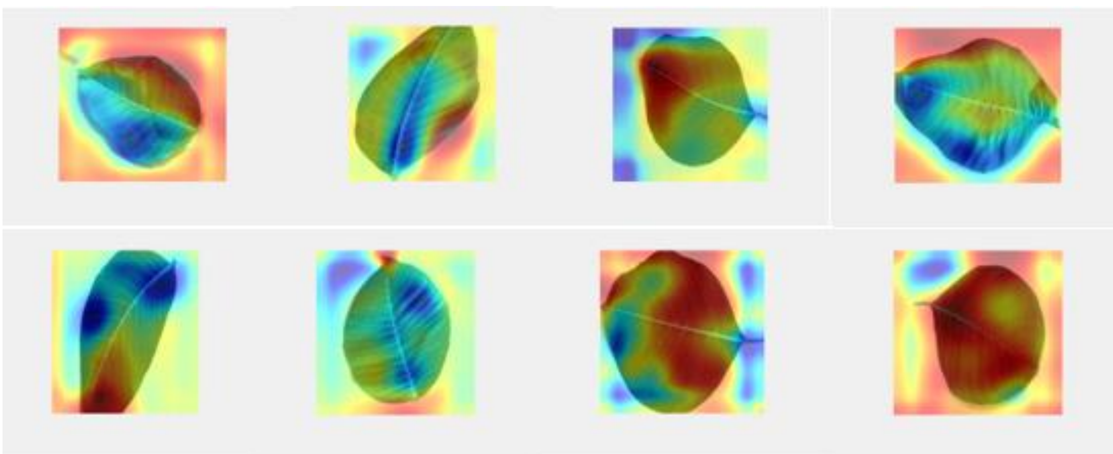

FIGURE 4. Grad-Cam image examples of walnut leaves.

For the classification of the generated walnut images, the Vgg16 CNN model, which exhibited the highest performance in the original walnut dataset, was employed. The flowchart depicting the GRADCAM-VGG16 method is illustrated in Figure 5.

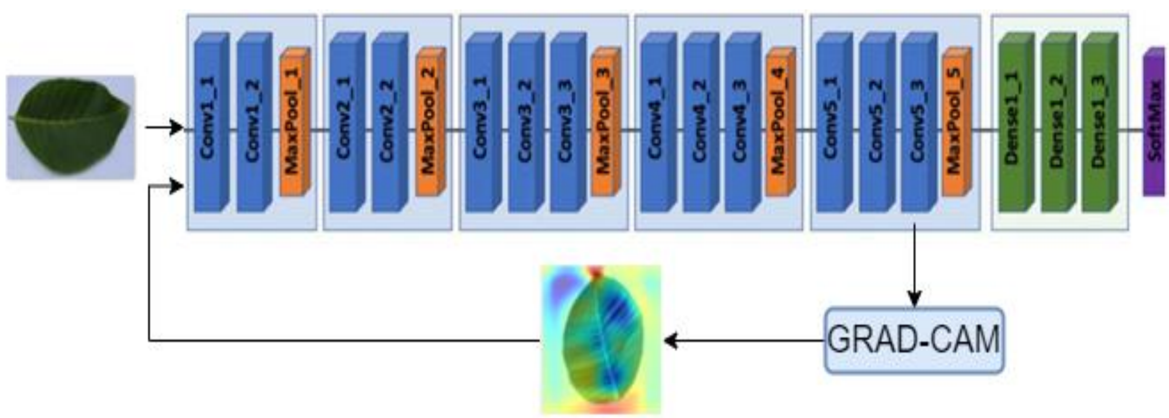

FIGURE 5. The architecture of GRADCAM-VGG16 model.

4. EXPERIMENTAL TEST AND RESULTS

Multi-class categorization is necessary since the walnut data set contains 18 distinct categories. The True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) metrics that are derived from the confusion matrix are used to evaluate performance in multi-class classification.

The total number of correctly categorized samples for all varieties other than the related species is represented by TN, whereas the number of correctly classified samples for each variety is represented by TP. The variables "FN" and "FP" denote the quantity of incorrectly identified samples of the respective variety.

The following performance metrics are employed in our study:

- Sensitivity, defined as the proportion of all genuine positives to accurately anticipated positives.
- Specificity, defined as the ratio of all genuine negatives to accurately anticipated negatives.
- Accuracy, which displays the proportion of cases across all samples that are correctly classified.
- Precision is defined as the ratio of all positive predictions to accurately predicted positives.
- Recall is a measure that indicates the proportion of processes that were expected to be positive and that came to pass. The formula for recall and sensitivity are the same.
- The harmonic means of the precision and recall values are given to us by the F-score.
- The following equations [12] provide the necessary computations for the performance calculation metrics given.

$$\text{Accuracy (Acc)} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{2}$$

$$\text{Sensitivity (Se)} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

$$\text{Specificty (Sp)} = \frac{\text{TN}}{\text{TN+FP}} \tag{4}$$

$$\text{Precision (Prec)} = \frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

$$\text{F-score } (\text{F} - \text{Sc}) = \frac{\text{2TP}}{\text{2TP+FP+FN}} \tag{6}$$

For experimental studies, Intel Core i7 processor, 8 GB RAM and 4 GB Nvidia graphics processor card were used. Matlab2021b was chosen as the software environment. During the training phase of CNN models, hyperparameters need to be adjusted in order to obtain more effective results. By experimenting with literature information, the hyperparameters that gave the best results were selected. The mini-batch size was determined as 16, the epoch was 32, and the learning rate was $1\text{x}10{-}4$.

The dataset was trained with the GRADCAM model, and a new data set was created with the colored images obtained as a result of training the walnut leaves with the model. After obtaining the GRADCAM images of walnut leaves and training them with the VGG16 model, the accuracy rate was determined to be 77.11%. The results of the experimental tests are given in Table 3. The train/epoch graph of model training is given in Figure 6. The confusion matrix obtained for calculating the experimental test results is demonstrated in Figure 7. Figure 6 shows the convergence value was reached approximately after the 15th epoch.

TABLE 3. Experimental test results of the proposed model

Model	Accuracy	Sensitivity	Specificity	Precision	F-score
Vgg16-GradCam	77.11	71.82	98.62	71.49	76.82

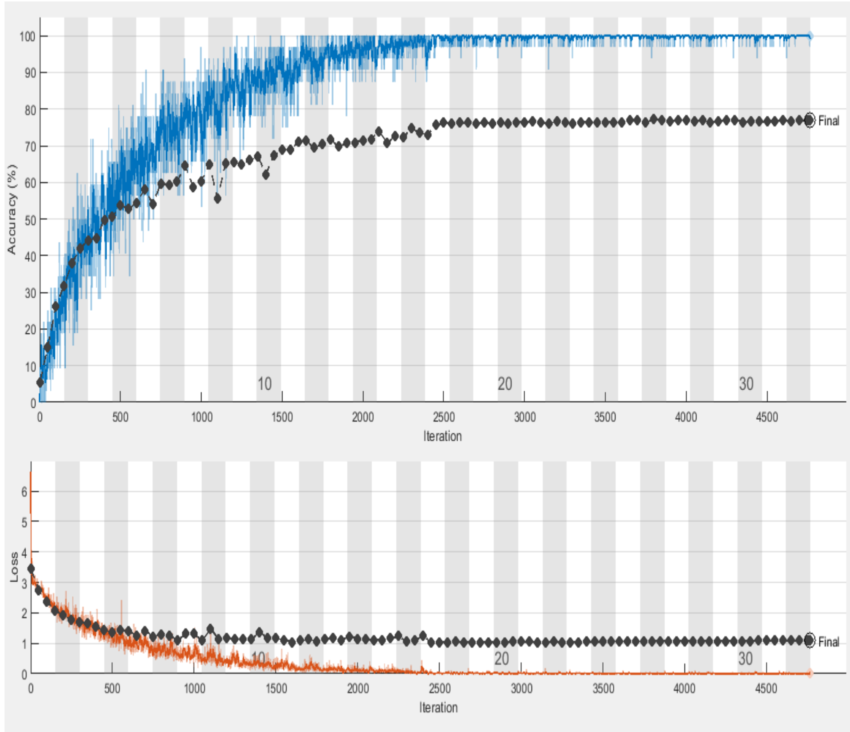

FIGURE 6. The training and loss graph obtained from the GRADCAM-VGG16 model.

	Frenquette	Bilecik	Chandler	Fernette	Fernor	Hardley	Howard	Kaman1	Kaplan86	Lara	Maya1	Mithland	Oguzlar77	Pedro	Sebin	Sen	Serr	Yalova3
Frenquette	135	1	1	2	1	0	1	2	1	0	0	2	0	0	0	0	1	0
Bilecik	4	82	0	1	1	4	0	0	0	0	0	0	2	2	0	6	4	7
Chandler	2	0	65	3	3	2	1	5	1	3	2	1	0	5	0	2	1	0
Fernette	3	0	0	90	2	1	0	2	0	3	0	2	0	0	1	1	0	0
Fernor	2	0	3	9	93	0	2	1	1	5	1	1	0	2	0	0	1	2
Hardley	2	3	4	0	0	95	0	0	0	1	0	0	0	4	0	3	0	0
Howard	0	0	1	5	9	0	68	2	3	3	1	0	1	3	1	1	1	1
Kaman1	2	1	0	1	5	1	3	91	3	0	4	0	0	1	0	0	3	0
Kaplan86	0	0	1	2	3	2	3	6	54	0	25	0	0	2	0	0	1	0
Lara	0	1	5	5	2	0	6	3	0	47	1	0	1	2	0	1	0	1
Maya1	0	0	3	1	0	3	4	1	28	1	44	0	0	0	0	0	3	0
Mithland	1	0	0	0	0	0	0	0	0	0	0	166	0	0	0	1	1	1
Oguzlar77	0	1	0	1	1	0	1	0	0	0	2	0	54	0	8	0	0	2
Pedro	1	3	4	0	4	1	0	2	1	1	0	0	1	69	2	1	1	0
Sebin	0	2	0	0	0	1	0	0	0	1	0	0	6	0	85	4	2	3
Sen	1	12	0	2	0	2	0	0	0	0	0	1	1	1	1	151	6	3
Serr	2	5	3	2	1	2	0	2	0	0	1	2	7	1	2	13	91	4
Yalova3	2	5	0	0	1	1	0	0	0	0	0	1	2	0	1	8	2	103

FIGURE 7. The confusion-matrix of GRADCAM-VGG16 model

5. CONCLUSION

The accurate differentiation of walnut varieties based on their leaves during the sapling stage is critically important in walnut cultivation. This provides significant advantages to the growers in terms of time and cost, given the extended duration for walnut trees to bear fruit, thereby avoiding substantial damages incurred by planting incorrect varieties. In this study, original dataset comprising 1751 walnut leaf images from 18 distinct walnut varieties was introduced, obtained with the assistance and permissions from experts at the Yalova Atatürk Garden Cultures Central Research Institute's Experimental Gardens.

Initially, preprocessing and data augmentation techniques were applied to the original dataset. Preprocessing rectified errors stemming from image capture on the leaf images. Subsequently, operations such as rotation, brightness adjustment, shearing, zooming, and flipping were employed to augment the dataset, resulting in a fourfold increase in the image count. All experimental assessments were conducted on both the original and augmented datasets. The data augmentation process notably enhanced the performance outcomes as observed through various experimental tests.

A new dataset was generated by applying the Grad-Cam method to the augmented dataset images. These Grad-Cam images were subsequently classified using the Vgg16 model, yielding an accuracy rate of 77.11%. While this accuracy achievement is competitive, it falls short of the desired level.

In the study, features were extracted in the conv5_3 layer of the Vgg16 CNN model. This layer was chosen because it is frequently used in the literature. However, it was observed that the desired result could not be achieved by extracting features from a single CNN layer. Additionally, Obviously Grad-cam does not provide successful results in classifying walnut leaf images since Grad-cam features selected from very similar leaf images were not sufficient for classification.

For future studies, the pre-processing step will first be developed with different image processing techniques. In addition, model success will be increased with the proposed original deep learning models and algorithms. Furthermore, the model with the highest success will be available for users via mobile application.

Data Statement

The original data set used in the study can be accessed via the link;
https://www.kaggle.com/datasets/alpertalhakaradeniz/walnut-leaf-dataset.

Conflict of Interest

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

REFERENCES

[1] K. Kulkarni, Z. Zhang, L. Chang, J. Yang, P. C. A. da Fonseca, and D. Barford, "Building a pseudo-atomic model of the anaphase-promoting complex," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 69, no. 11, pp. 2236–2243, 2013.

[2] F. Doğan and İ. Türkoğlu, "Derin öğrenme algoritmalarının yaprak sınıflandırma başarımlarının karşılaştırılması," *Sak. Univ. J. Comput. Inf. Sci.*, vol. 1, no. 1, pp. 10–21, 2018.

[3] T. Karadeniz and E. Güler, "Cumhuriyetin İlk Yillarindan Günümüze Ceviz Yetiştiriciliği," *Bahçe*, 2017.

[4] T. Karadeniz, "Ordu Yöresinde yetiştirilen ceviz genotiplerinin (Juglans regia L.) seleksiyonu," *Ordu Üniversitesi Bilim ve Teknol. Derg.*, vol. 1, no. 1, pp. 65–74, 2011.

[5] S. Solak and U. Altinişik, "Görüntü işleme teknikleri ve kümeleme yöntemleri kullanılarak fındık meyvesinin tespit ve sınıflandırılması," *Sak. Univ. J. Sci.*, vol. 22, no. 1, pp. 56–65, 2018.

[6] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] M. H. Saleem, J. Potgieter, and K. M. Arif, “Plant disease detection and classification by deep learning,” *Plants*, vol. 8, no. 11, p. 468, 2019.

[8] T. Karahan and V. Nabiyev, “Plant identification with convolutional neural networks and transfer learning,” *Pamukkale Üniversitesi Mühendislik Bilim. Derg.*, vol. 27, no. 5, pp. 638–645, 2021.

[9] I. M. Dheir, A. Soliman, A. Mettleq, and A. A. Elsharif, “Nuts Types Classification Using Deep learning,” *Int. J. Acad. Inf. Syst. Res.*, vol. 3, no. 12, pp. 12–17, 2019.

[10] Y. Liu, J. Su, G. Xu, Y. Fang, F. Liu, and B. Su, “Identification of grapevine (vitis vinifera l.) cultivars by vine leaf image via deep learning and mobile devices,” 2020.

[11] D. K. Nkemelu, D. Omeiza, and N. Lubalo, “Deep convolutional neural network for plant seedlings classification,” *arXiv Prepr. arXiv1811.08404*, 2018.

[12] A. T. Karadeniz, Y. Çelik, and E. Başaran, “Classification of walnut varieties obtained from walnut leaf images by the recommended residual block based CNN model,” *Eur. Food Res. Technol.*, pp. 1–12, 2022.

[13] A. T. Karadeniz, E. Başaran, and Y. Celik, “Identification Of Walnut Variety From The Leaves Using Deep Learning Algorithms,” *Bitlis Eren Üniversitesi Fen Bilim. Derg.*, vol. 12, no. 2, pp. 531–543, 2023.

[14] A. Beikmohammadi, K. Faez, and A. Motallebi, “SWP-LeafNET: A novel multistage approach for plant leaf identification based on deep CNN,” *Expert Syst. Appl.*, vol. 202, p. 117470, 2022.

[15] A. Dobrescu, M. V. Giuffrida, and S. A. Tsaftaris, “Doing more with less: a multitask deep learning approach in plant phenotyping,” *Front. Plant Sci.*, vol. 11, p. 141, 2020.

[16] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[17] H. Jiang *et al.*, “A multi-label deep learning model with interpretable grad-CAM for diabetic retinopathy classification,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 1560–1563.

[18] Y. Zhang, D. Hong, D. McClement, O. Oladosu, G. Pridham, and G. Slaney, “Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging,” *J. Neurosci. Methods*, vol. 353, p. 109098, 2021.

[19] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv Prepr. arXiv1409.1556*, 2014.