Journal of Agricultural Sciences
(Tarim Bilimleri Dergisi)

J Agr Sci-Tarim Bili
e-ISSN: 2148-9297
jas.ankara.edu.tr

JAS

Hybrid Content-Based Image Retrieval System for a Comprised 27-Class Euphorbia Seed Dataset Using Deep Feature Fusion

Murat Kursat[a], **Mucahit Karaduman**[b], **Hursit Burak Mutlu**[c], **Irfan Emre**[d], **Muhammed Yildirim** [c*]

[a]*Faculty of Science and Arts, Department of Biology, Bitlis Eren University, Bitlis, TURKEY*
[b]*Department of Software Engineering, Malatya Turgut Ozal University, Malatya, TURKEY*
[c]*Department of Computer Engineering, Malatya Turgut Ozal University, Malatya, TURKEY*
[d]*Department of Basic Education, Faculty of Education, Firat University, Elazig, TURKEY*

ARTICLE INFO
Research Article
Corresponding Author: Muhammed Yildirim, E-mail: muhammed.yildirim@ozal.edu.tr
Received: 20 March 2025 / Revised: 20 April 2025 / Accepted: 08 May 2025 / Online: 30 September 2025

Cite this article
Kursat M, Karaduman M, Mutlu H B, Emre I, Yildirim M (2025). Hybrid Content-Based Image Retrieval System for a Comprised 27-Class Euphorbia Seed Dataset Using Deep Feature Fusion. *Journal of Agricultural Sciences (Tarim Bilimleri Dergisi),* 31(4):931-940. DOI: 10.15832/ankutbd.1661676

ABSTRACT
Content-based Image Retrieval (CBIR) systems have been used frequently in recent years, along with developing technology. Especially in large datasets, retrieval-based systems produce more successful results. This study created a dataset consisting of 27 different *Euphorbia* seed types belonging to the same genus. It is difficult for Convolutional Neural Network (CNN) architectures to produce successful results in the created dataset. In addition, the high computational and memory requirements of CNN architectures have further increased the need for CBIR systems in large datasets. Therefore, a hybrid retrieval system was developed to make inferences from 27 different seed images. In the developed system, feature extraction was performed using Darknet53, Xception, and Densenet201 architectures. These extracted features were concatenated to bring together different features of the same image. Then, unnecessary features were eliminated from the combined features with the Neighborhood Component Analysis (NCA) method. The cosine similarity measurement metric was used to measure the similarity between the query image and other images. Precision-recall curves and Average Precision (AP) metrics were used to measure the performance of the proposed retrieval-based system. In the study, an average AP value of 0.96809 was obtained. The morphology of the seeds is a critical characteristic of *Euphorbia*, and this work has validated the artificial intelligence methodology.

Keywords: CBIR, Retrieval, Seed, Classification, Deep Learning

1. Introduction

The genus *Euphorbia* L., which includes over 2000 species worldwide, is mostly found in temperate and subtropical climates. It grows in a variety of plant forms, including trees, shrubs, and herbs (Fayed & Hassan 2007). The genus is the second biggest genus of flowering plants after *Astragalus* L. with distinct macro- and micro-morphological seed features (Frodin 2004; Pahlevani & Akhani 2011). Morphological and anatomical traits, such as leaf form, floral structure, and seed and fruit features, have been used in plant taxonomy (Ibrahim et al. 2023). For the purpose of identifying and classifying the plant, the morphological traits of the *Euphorbia* seed are crucial and analysis of seed images is now crucial to maintaining biodiversity (Da Silva et al. 2016; Loddo et al. 2021). Using a light and scanning electron microscope, the color, length, width, shape, caruncle size, caruncle shape, and lipid granule shape of *Euphorbia* seeds were traditionally measured and examined (Kursat et al. 2023). Traditionally, seeds have been categorized using attributes including color, shape, and texture. This is often done by experts visually examining each sample, which is an extremely laborious and time-consuming process (Gulzar et al. 2020). However, automating the classification of seed qualities based on picture pixel values, minimizing distortions from natural light and microscopes, and speeding up the procedure are only a few advantages of image analysis techniques over manual examination (Loddo et al. 2021).

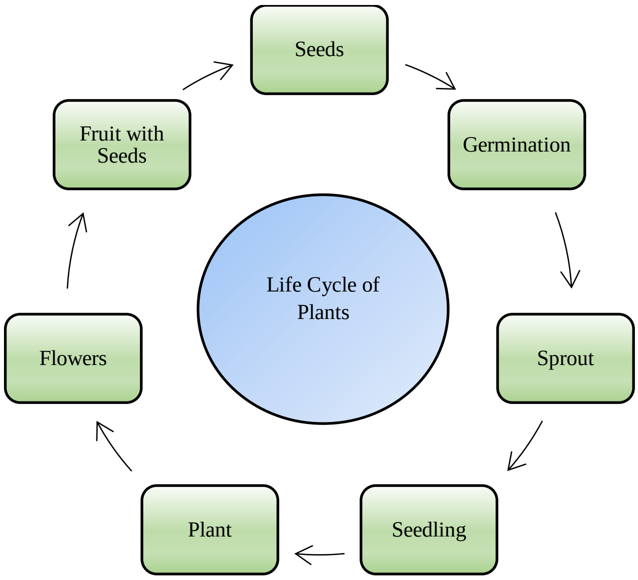

Figure 1- Life Cycle of Plants

Figure 1 shows the plant life cycle, which starts from a seed and then turns into a seed again. This life cycle's healthy and uninterrupted continuity is very important for humans and nature.

Advances in computer vision and image processing technologies have led to significant advancements in image-based agricultural product recognition techniques (Huang et al. 2022). The separation of objects from the background is the initial stage of image processing and is essential to the precision of all measurements (Dayrell et al. 2023). One of the best methods for retrieving visual data is CBIR (Muneesawang & Guan 2004). The primary concept of CBIR is the analysis of picture information using low-level features and, is a useful technique for retrieving images based on their visual contents, including color, form, texture, and so forth (Yue et al. 2011; Karthikeyan et al. 2023). It is used in many computer vision and image information systems which includes medical, criminology and agriculture for identification of leaf, stem and fruit diseases (Patil & Kumar 2017). It is one of ideal methods for retrieving a picture it minimizes texture-based issues and image indexing (Pandey et al. 2013). Searching and retrieving digital images from a vast database is possible with CBIR, which makes use of image content features and similarity metrics and feature representation play a critical role in determining how well a content-based image retrieval system retrieves images (Saritha et al. 2019). The goal of this study is to extract characteristics from seed images using deep learning techniques, including seed color, texture, shape etc. The classification of plant species has greatly improved recently thanks to deep learning approaches (Haupt et al. 2018). Deep learning and image-based methods facilitate the identification of plant species while also expediting the process and enabling non-experts to participate actively (Kumar et al. 2023). Machine learning's subject of deep learning is a well-known and extensively used method that has been used in a variety of fields, including speech recognition, computer vision, biology including plant classification, and medicine (Gawli & Gaikwad 2020).

With the development of technology, the amount of data kept in databases is increasing daily. The processing of this data kept in databases is of great importance. Deep learning-based methods cannot produce successful results in large datasets, and these models take a long time to work and put extra load on the memory. Therefore, a dataset of 27 classes of the same type was created. Recently, the retrieval method, which has become popular, was developed to retrieve the images in the created dataset. The developed method retrieved images similar to the query image from the dataset, and high-performance values were achieved.

Deep learning algorithms were used to generate it from plant photos in the 2019 study by Gyires-Tóth, Osváth, Papp, and Szűcs. CNN architectures were utilized for feature learning, and fully connected layers with logsoftmax output were employed for classification (Gyires-Tóth et al. 2019). Additionally, they maximized MAP (Mean Average Precision), which is essential to the efficiency of picture retrieval, by optimizing the length of the returned lists. As a result, they were able to enhance MAP in the test set by more than 50% over the baseline. Another research by (Loddo et al. 2021) employed deep learning algorithms to classify two plant seed datasets based on their species or family. The accuracy values for the first and second findings they received were 95.65% and 97.47%, respectively. Furthermore, they reported that the deep learning approach addressed the retrieval problem as well, yielding satisfactory results. (Loddo et al. 2021) suggested that the findings achieved from both challenges are seen to be a great beginning point for the development of a comprehensive system for the recognition, categorization, and retrieval of seeds, which will provide significant assistance in the domains of botany and agriculture. The done by (Picek et al. 2022) presents a novel retrieval-based approach for recognition via closest neighbor classification in a deep embedding space, and examines and compares machine learning techniques for automatic image-based plant species detection. They showed that retrieval technique performed better in all assessed scenarios (0.28%, 4.13%, 10.25%, respectively) on ExpertLifeCLEF 2018, PlantCLEF 2017, and iNat2018–Plantae (Picek et al. 2022). In a study done by Ibrahim et al. (2022), a deep learning analysis dataset was constructed using fruit photos for 52 species that are members of four distinct families:

Apiaceae, Brassicaceae, Asteraceae, and Apocynaceae. They suggested a novel design for a Convolution Neural Network (CNN) model, which would extract the fruit features, classify each image with its family, and use the trained model to predict that the new fruits belong to their four families. The maximum accuracy for the training and testing module was 99.82%. Also, (Ibrahim et al. 2023) used two distinct datasets including photographs of wild plant seeds that were collected from Egypt. They indicated that plant taxonomy can benefit greatly from deep learning as well because it can automate the process of classifying different plant species according to their available characteristics. In addition, (Ibrahim et al. 2023) found that the proposed approaches outperformed the traditional methods with the greatest accuracy of 93%, F1-score, and area under the curve (AUC) of 95%, respectively, according to extensive experiments conducted with the transfer learning method (DenseNet201). Also, (Putzu et al. 2020) contend that, given CNNs remarkable performance in image representation and classification tasks, CNNs could potentially be used by Content-Based Image Retrieval (CBIR) systems, particularly when Relevance Feedback (RF) methods are used. They suggested that the efficiency of the suggested strategies in enhancing the CNN's representation power with regard to the user idea of picture similarity is demonstrated by experimental findings on various datasets (Lorenzo et al. 2020). According to Hussein, Mashohor, and Saripan's (2011) research, which used content-based image retrieval systems to analyze 280 plant leaf images in their dataset, the suggested approach had a high accuracy of 92%. In addition, Apriyanti, Arymurth and Handoko (2013) suggested that their strategy has the potential to improve content-based floral image retrieval accuracy for Orchid species by up to ± 14%, as demonstrated by the results (Apriyanti et al. 2013). Also, The CBIR system was created to retrieve damaged soybean leaves utilizing the color, shape, and texture properties of the leaf, as demonstrated in a study by Patil and Kumar (Patil & Kumar 2017). According to their findings, soybean leaves afflicted by the septoria brown spot disease, mosaic virus, and pod mottle disease attain retrieval efficiencies of roughly 96%, 68%, and 76%, respectively. Moreover, they showed that combining characteristics yields an average retrieval efficiency of 72% for the top 10 retrievals and 80% for the top 5 retrievals. Furthermore, Karthikeyan and Raja's (2023) study develops a unique DTLDN-CBIRA model for agricultural plant disease picture retrieval (Karthikeyan et al. 2023). Also, Saritha et al. developed a deep learning framework for CBIR by training large-scale Deep Belief Networks to produce effective feature representations of images. Additonally, they carried out a wide range of empirical investigations to provide thorough assessments of Deep Belief Networks, which are used to acquire features for a range of CBIR tasks in various contexts (Saritha et al. 2019). It has been suggested that the performance of the models is adversely affected by the fact that plant species belong to the same family and that there are many classes in the research. However, they found that the accuracy value obtained in the developed Plant21 model using Euphorbia plants in the flowering stages is quite high. Also, it has been indicated the developed Plant21 model is successful in classifying *Euphorbia* species (Kursat et al. 2023).

In the study, an Euphorbia seed dataset of the same type with 27 classes was created. Pre-trained Darknet53 (Toğaçar 2022), Xception (Sutaji & Yıldız 2022), and Densenet201 (Auni & Sugiharti 2024) architectures were used as the base to extract the features of the images in this dataset. The 2700 x 1000 feature maps obtained from each architecture were combined, and then unnecessary features were eliminated with the NCA (Ding et al. 2024) method. As a result, a 2700 x 900 feature map was obtained. It is aimed that the proposed model works faster thanks to the optimized feature map. The aim is to combine different features of the same image with the feature fusion step. In this way, the aim is to produce more successful results in the proposed model. In our retrieval-based system, the cosine similarity measurement metric was used, and competitive results were obtained in the proposed model. In conclusion, Artificial intelligence techniques were used to classify seed photos of Euphorbia taxa, offering a novel way for plant recognition and classification that will benefit both plant identification professionals and non-experts.

2. Background

In this section, the dataset we created from 27 different *Euphorbia* seed types of the same type used in the study is introduced. Then, the feature extraction, feature fusion and feature selection steps used in the study are examined. In addition, in this section, the CBIR system and the three proposed models are examined.

2.1. Dataset

The 27 *Euphorbia* taxa included in the seed samples utilized in this study were collected in their native habitats between April and August of the 2019–2021 year. The plant materials were identified by Prof. Dr. M. Kürşat according to Flora of Turkey and East Aegean Islands (Davis 1965). Voucher specimens were deposited in the Bitlis Eren University Herbarium. *Euphorbia* seed images were carried out under a stereomicroscope (Leica S8APO) with an incorporated camera (Leica DFC295). The dataset contains 27 classes, each containing 100 images. The dataset consists of a total of 2700 images in .jpg format. The class names in the dataset are presented in Table 1.

Table 1- Class numbers and names

No	*Name*	*No*	*Name*	*No*	*Name*
1	Euphorbia chamaesyce L.	**10**	Euphorbia gaillardotii Boiss. & Blanche	**19**	Euphorbia macroclada Boiss.
2	Euphorbia rhytidosperma Boiss. & Balansa.	**11**	Euphorbia eriophora Boiss.	**20**	Euphorbia cheiradenia Boiss. & Hohen.
3	Euphorbia grisophylla M.L.S.Khan.	**12**	Euphorbia rhabdotosperma Radcl.-Sm.	**21**	Euphorbia seguieriana Neck. subsp. seguieriana
4	Euphorbia macrocarpa Boiss. & Buhse.	**13**	Euphorbia helioscopia L.	**22**	Euphorbia heteradena Jaub. & Spach.
5	Euphorbia orientalis L.	**14**	Euphorbia aleppica L.	**23**	Euphorbia esula subsp. tommasiniana (Bertol.) Kuzmanov
6	Euphorbia altissima var. Altissima.	**15**	Euphorbia szovitsii var. kharputensis Azn. ex M.S.Khan.	**24**	Euphorbia sanasunitensis Hand.-Mazz.
7	Euphorbia altissima var. glabrescens Boiss. ex M.S.Khan.	**16**	Euphorbia falcata subsp. Falcata**.**	**25**	Euphorbia iberica Boiss.
8	Euphorbia stricta L.	**17**	Euphorbia denticulata Lam.	**26**	Euphorbia oblongifolia (K.Koch) K.Koch**.**
9	Euphorbia microsphaera Boiss.	**18**	Euphorbia craspedia Boiss.	**27**	Euphorbia erubescens Boiss

In the rest of the article, class numbers are used instead of class names. Figure 2 provides examples of the plant species that were employed in the study.

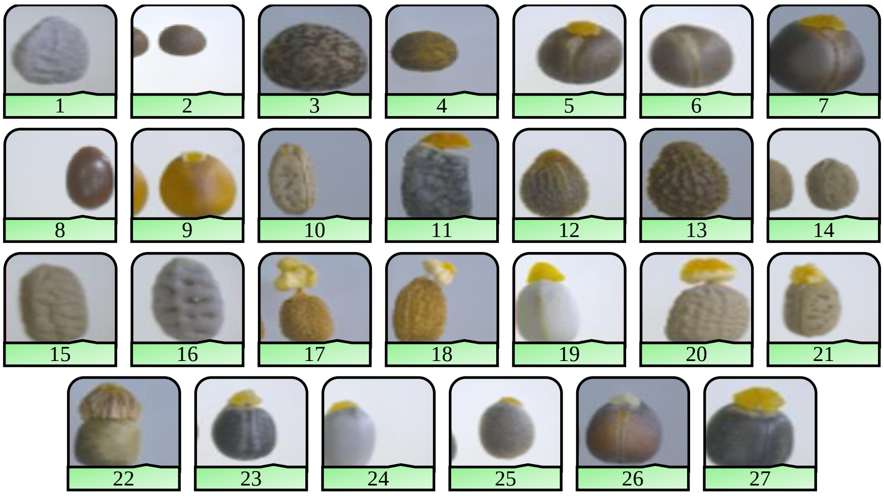

Figure 2- Sample Images in the Twenty-Seven Class Dataset

2.2. Content-based image retrieval (CBIR) systems

CBIR systems provide easy searching in large datasets and effectively find images with similar features. CBIR systems achieve successful results in large datasets. The high computational cost of deep learning architectures, high processing power requirements, and limited interpretability have made retrieval systems popular, especially in large datasets. Retrieval systems are lighter in terms of computation. Therefore, retrieval systems provide advantages for fast retrieval and real-time applications. Since retrieval systems generally use lower-dimensional feature vectors, memory requirements are reduced, which allows retrieval systems to work more effectively with large datasets. In addition, retrieval systems are based on simple feature extraction and similarity measures, which provide more interpretable results. In the proposed model, feature maps of the images were extracted using a combination of Darknet-53, Xception, and DenseNet-201 architectures and NCA. The images whose feature maps were extracted with Darknet53, Xception, and Densenet201 architectures; feature vectors were combined to obtain a feature map of 2700x3000 size for 2700 images. The NCA optimization method was used to remove non-distinctive features from the obtained feature vector size and to improve working performance. With the NCA dimension reduction method, the feature vector, which was originally 2700x3000, was optimized to 2700x900. The block diagram of the proposed model for the dataset used in CBIR and the feature map extraction for the queried image is shown in Figure 3.

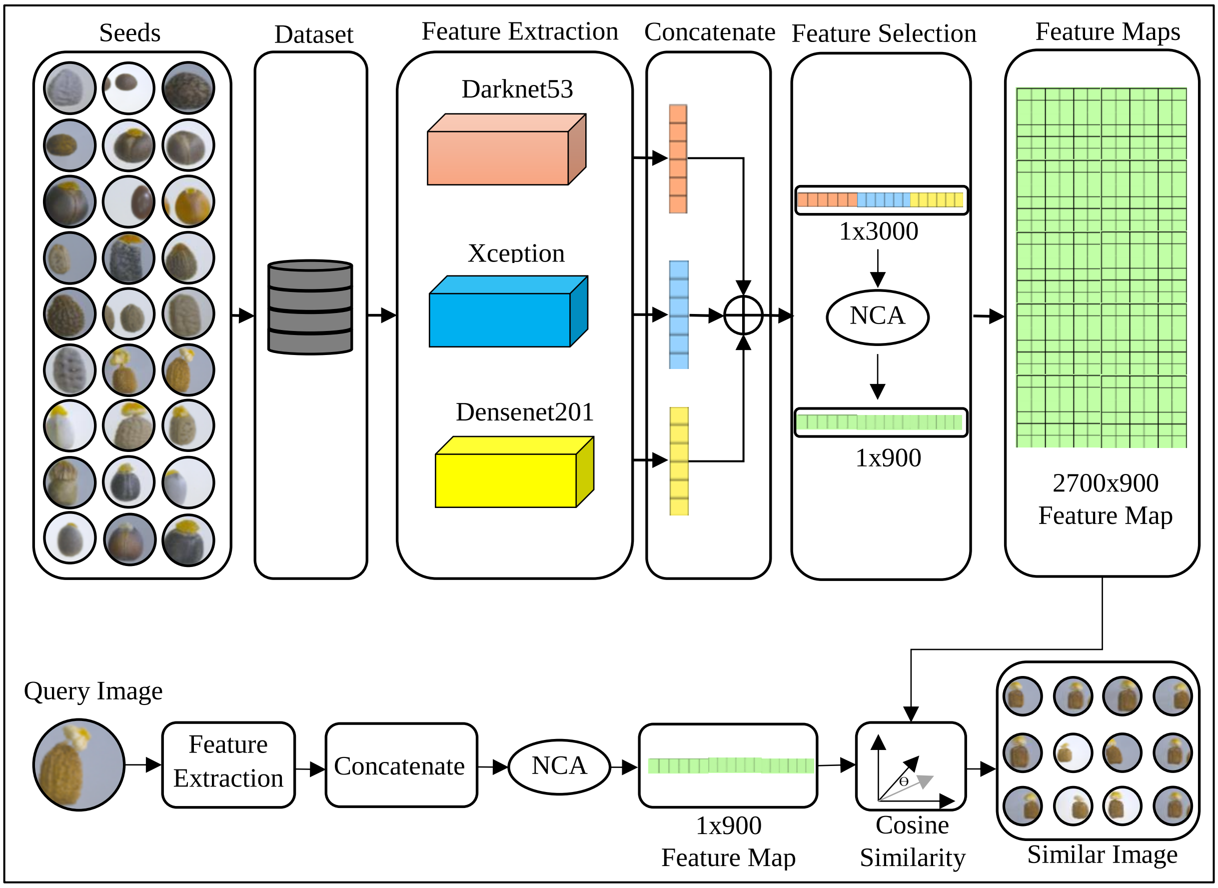

Figure 3- Suggested CBIR System

The feature map acquired via CBIR was used to conduct an image search. The feature map of the image to be searched was extracted in the model created as CBIR, much like in the suggested model, and CNN architectures on the feature map were compared. By evaluating the P-R curve with an 11-point sequential access curve, the comparison with CNN architectures utilizing the Cosine similarity measurement method was assessed. Darknet53, Xception, Densenet201, Squeezenet, Efficientnetb0 (Tan et al. 2019), and MobilenetV2 (Sandler et al. 2018) were compared with the proposed model in CNN architectures. In the evaluation made on the 11-point sequential access P-R curve, high performance was achieved with the NCA (Darknet53 + Xception + Densenet201) proposed model with the cosine method.

2.2.1. Cosine similarity measure

The ratio of the computation derived from the product of the cosine angle of these two vectors is the cosine similarity determined for two vectors. Two vectors are considered comparable when their cosine similarity value is 1. The cosine similarity is calculated using Equation 1 (Lahitani et al. 2016).

$$\text{Cos}\,\alpha = \frac{A.B}{|A|.|B|} = \frac{\sum_{i=1}^{n} A_i.B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2}.\sqrt{\sum_{i=1}^{n}(B_i)^2}} \qquad (1)$$

The weights of each feature in vector A and vector B are denoted by A and B, respectively, in Equation 1. The cosine similarity rule states that the degree of similarity between two vectors increases with decreasing angle (Lahitani et al. 2016).

Twenty related images in the dataset are accessed by the suggested CBIR model, as seen in Figure 4. The dataset's 20 comparable images are displayed in order after a randomly chosen image is chosen.

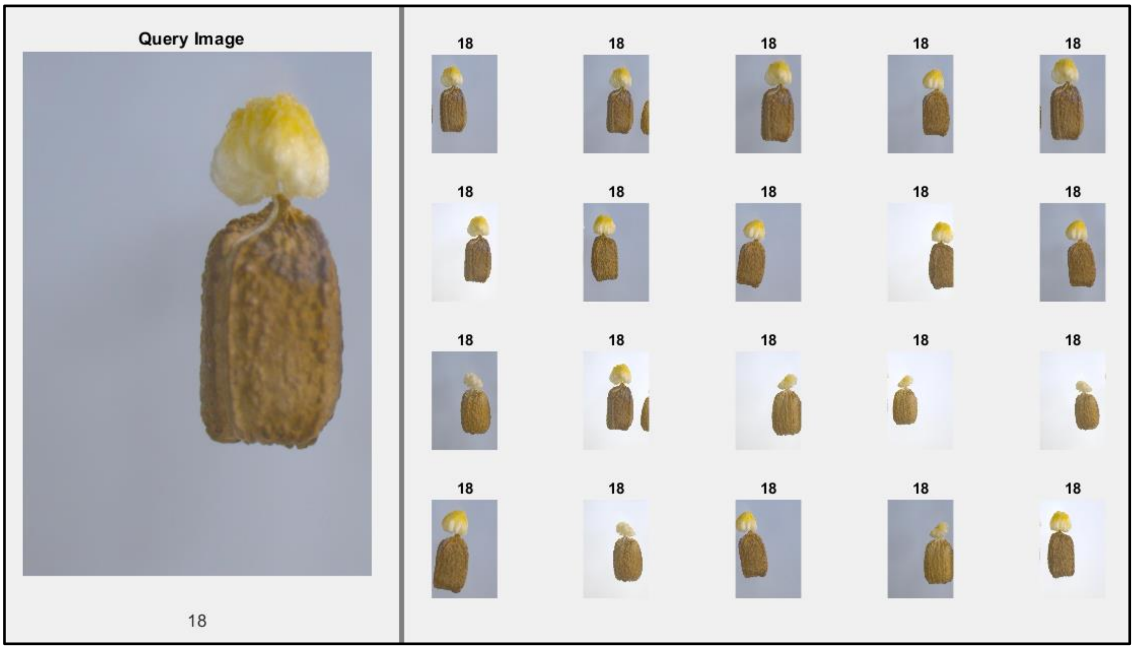

Figure 4- Twenty query images accessed by the suggested model

Figure 4 shows the sequential access to 20 images from the dataset that were similar to the query image of the 18th class provided for the query. Twenty images that are comparable to the one accessed in an image from the 18th class are in the same class as a consequence of the query.

3. Experimental Results

In this investigation, the deep learning and textural models' parameters were kept unchanged. The MATLAB 2024b was used to get the results of these investigations on a PC running Windows 10 64-bit, with an Intel i5 processor and 8 GB of RAM. Precision and Recall graphs and Average Precession (AP) metrics were used for performance evaluation.

3.1. Results of deep and textural based models

In this study, six pre-trained deep models were used. Feature extraction was performed using these models, and the successful models were used as the basis for the proposed model. There are studies in the literature using CNN-based models (Yildirim 2024; Sunnetci et al. 2022; Yucel & Yildirim 2024). In CBIR systems, the order of the accessed image is essential. Standard 11-point P-R graphs are used to evaluate such CBIR systems. A P-R graph is provided for 20 randomly selected photos in each class in the CBIR system that we created. In the P-R graph, the performance of Darknet53, Densenet201, Xception, Squeezenet (Theerthagiri et al. 2024), MobilenetV2, and Efficientnetb0, and the proposed model is evaluated by using Cosine, one of the similarity measurement methods.

The interpolated 11-point sequential access was assessed in this study using the AP value and the P-R curve. The data collection has twenty-one classes. The average P-R curve was derived by querying each of the twenty-one classes' photos separately. Twenty photographs were retrieved from the CBIR system after each class's images were queried independently. By analyzing the photos accessed and queried in each class, the average P-R curve was produced. Following a separate evaluation of the P-R curves for the 21 classes, 2,700 photos from the dataset were queried, and the average P-R curve of the 20 images retrieved in each query was computed and assessed.

An 11-point sequential access P-R graph has been used to analyze each of the 2,700 images in 27 classes in Figure 5 in order to access 20 images in the CBIR system. 54,000 picture accesses were compared because each of the 2,700 photos in the dataset had 20 accesses. The performance of the architectures was evaluated in this comparison. When the performances in the classes are examined in Figure 5, it is evaluated that the most successful architecture is our proposed model NCA (Darknet53 + Densenet201 + Xception). The effect of the performance in the classes of the tested algorithms was observed. It showed the highest success compared to other measurement metrics with AP=0.84137. Histogram Intersection showed the lowest success with AP=0.73312.

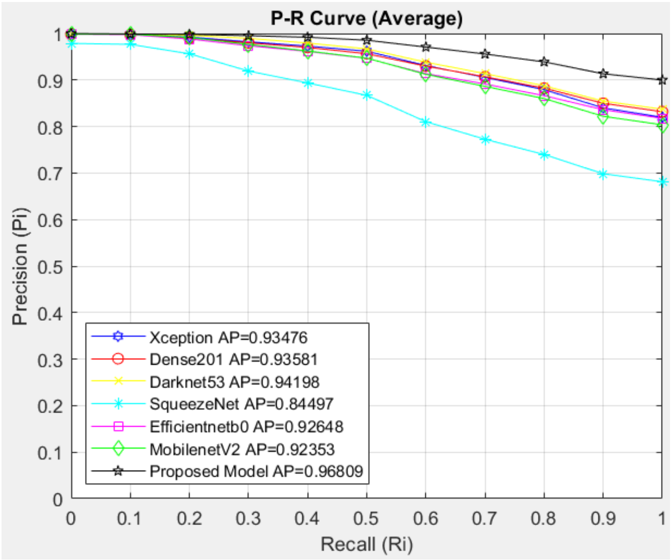

Figure 5- Overall Performance P-R Graph

When the general performance of the similarity methods in the proposed CBIR system is examined in Figure 5, our proposed model achieves the highest performance with the Cosine similarity measurement method with AP= AP=0.96809. Other similarity measurement methods are examined as Xception 0.93476, Dense201 0.93581, Darknet53 0.94198, SqueezeNet 0.84497, MobilenetV2 0.92353, and Efficientnetb0 0.92648, respectively. The SqueezeNet architecture provides the lowest performance.

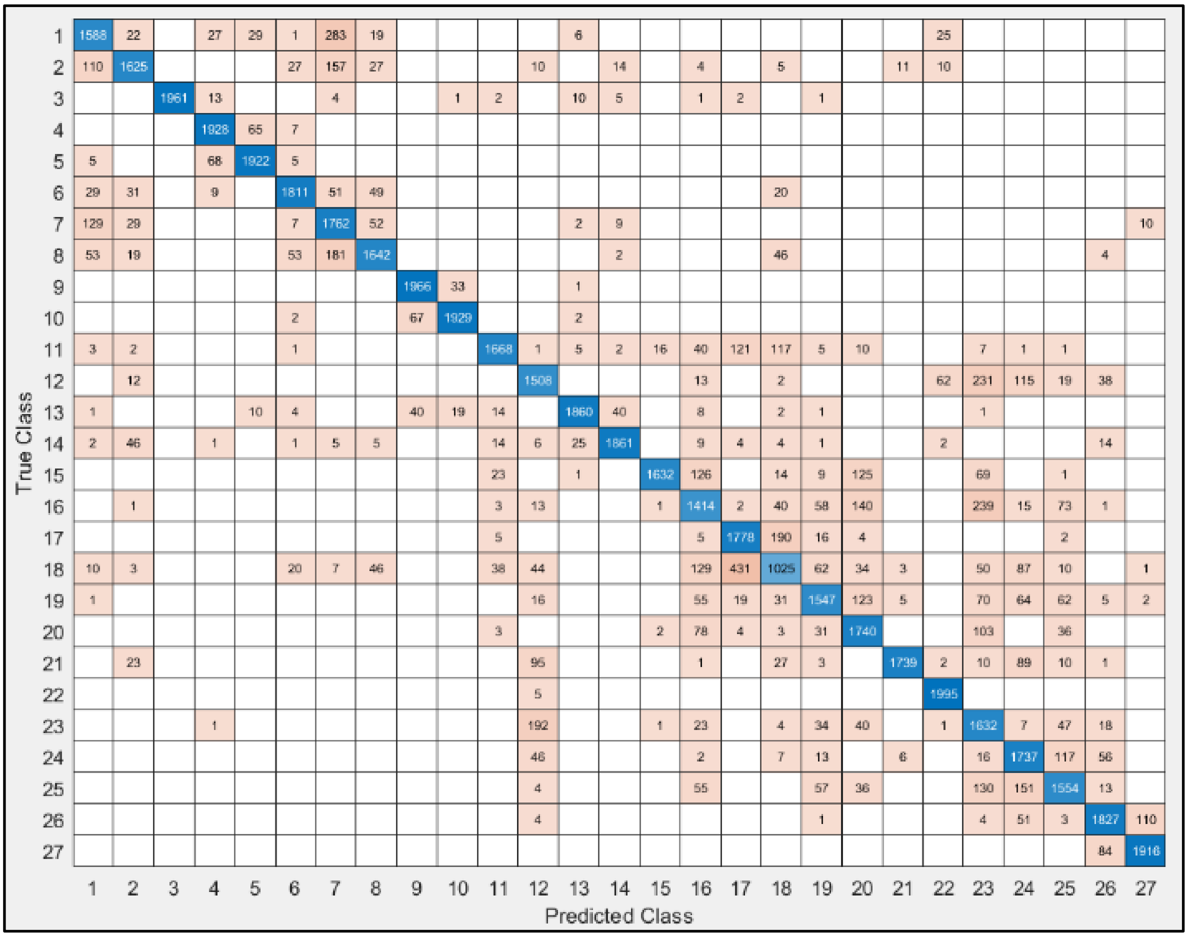

Figure 6- Overall Performance Confusion Matrix

Figure 6 shows the complexity matrix of the total 54,000 image comparisons obtained by accessing 20 similar images of each of the proposed model's 2,700 images.

When Figure 5 and Figure 6 are examined, it is evaluated that the proposed model with the Cosine similarity measurement method is more successful than other architectures in the general average.

The architectures examine the average of the 27 classes in the dataset, and Table 2 gives the AP values of the 11-point sequential access P-R Graph.

Table 2- Comparative AP values of the models for all classes

CLASS	*Average Precession (AP)*						
	Xception	*Densenet201*	*Darknet53*	*SqueezeNet*	*EfficientNetb0*	*MobileNetV2*	*Proposed Model*
1	0.88518	0.91546	0.92923	0.75318	0.91784	0.89946	0.93041
2	0.89554	0.86244	0.92558	0.73379	0.88665	0.89007	0.95714
3	0.99136	0.99441	0.99532	0.98557	0.99699	0.9852	0.99631
4	0.98666	0.95700	0.98089	0.90870	0.94061	0.97036	0.99422
5	0.98763	0.97010	0.98286	0.90278	0.95553	0.97031	0.98852
6	0.95511	0.93166	0.96591	0.76824	0.94237	0.94445	0.97954
7	0.90623	0.89201	0.91057	0.74454	0.91811	0.91626	0.97259
8	0.91148	0.91655	0.93032	0.77554	0.94867	0.91222	0.96104
9	0.99039	0.98244	0.98938	0.97844	0.98751	0.98807	0.99574
10	0.97194	0.97034	0.96989	0.93614	0.98228	0.97132	0.99365
11	0.89535	0.93354	0.93863	0.81214	0.86427	0.8825	0.95696
12	0.93852	0.93596	0.93908	0.82205	0.93753	0.90841	0.95569
13	0.95291	0.96704	0.94727	0.88453	0.94822	0.96167	0.98339
14	0.96565	0.97910	0.97012	0.85038	0.95747	0.95318	0.98486
15	0.93268	0.94413	0.93011	0.93167	0.9654	0.95339	0.94963
16	0.89949	0.88436	0.89229	0.76802	0.86022	0.84844	0.95158
17	0.90826	0.89132	0.93270	0.77774	0.87092	0.87359	0.95852
18	0.81765	0.82061	0.83702	0.69827	0.75402	0.79471	0.88051
19	0.91885	0.94192	0.92277	0.80066	0.90000	0.90453	0.97067
20	0.92050	0.93173	0.93877	0.79536	0.88077	0.87155	0.96991
21	0.98052	0.98589	0.97353	0.92893	0.97640	0.97454	0.99134
22	0.99662	0.99959	1.00000	0.99749	1.00000	0.99765	1.00000
23	0.90036	0.89745	0.91260	0.78537	0.86101	0.86774	0.95144
24	0.92044	0.93033	0.90579	0.81819	0.94711	0.9153	0.97069
25	0.89755	0.90662	0.87519	0.79116	0.89663	0.87772	0.94795
26	0.94750	0.95488	0.95508	0.90093	0.9523	0.93679	0.96654
27	0.96404	0.96996	0.98265	0.96434	0.96647	0.96585	0.97948

Table 2 shows the AP values of the P-R Graphs of 27 classes in the dataset. When the 27 classes in the dataset were compared according to CNN and Machine Learning architectures, the proposed model showed the highest success in all 27 classes. Thus, when we consider all classes separately, it was proven that the proposed model showed higher success. Among the classes, other architectures performed less than the proposed model. Our proposed model showed the highest success among the classes in the 22 class with AP=1 and the lowest success in the 18 class with AP=0.88051.

When the general performance of the proposed CBIR system is examined in Figure 5, the highest performance is shown by the proposed model method. When the performance between classes is examined in Table 2, the proposed model again showed the highest performance in all classes. Squeezenet architecture generally show the lowest performance between classes. In this study, when the general performance in classes and the performance in 27 classes are examined, it is evaluated that the proposed method is the method that achieves the highest performance in increasing the number of images compared to other methods by using the Cosine similarity measurement method.

4. Conclusions

In this study, the results obtained using CNN architectures was compared with the proposed model results. Using artificial intelligence techniques, it has been demonstrated that seed morphological characteristics are crucial for taxonomic differentiation in the genus *Euphorbia*. This gives both professionals and non-experts access to a valuable data source. One limitation of the study is the small number of images in the dataset. Improving the study's suggested model's performance requires adding more images to the dataset. The similarity measurement methods, along with the feature extraction method used in the CBIR system, were evaluated, and the method that showed the highest performance was selected. First, the performance metrics of Densenet201, Darknet53, Xception, Squuzenet, Efficientnetb0, MobilenetV2, and the proposed model were compared to develop the most successful model in feature map extraction for use in the CBIR system. The Xception, Darknet53, and Resnet201 architectures that achieved the highest performance in these architectures were taken. Darknet53, Xception, and Resnet201 architectures were combined with the feature map merging method to create a feature map. A feature map of 2700x3000 size was obtained for 2700 images in the dataset. The feature vector size was reduced with the NCA dimension reduction method to increase the performance of working on the feature vector with the combination of these three architectures. The feature vector dimensions were reduced from 2700x3000 to 2700x900 with NCA. The proposed model was evaluated with other architectures using the proposed feature map extraction in the CBIR system. In evaluating the similarity measurement methods, 11-point sequential access P-R Curves were used for each similarity method. The general performance in the dataset and the last 27 classes were evaluated separately. When our proposed model and 27 classes were evaluated separately in the general performance in accessing 20 images in the dataset, it showed the highest performance in all classes. When the general performance evaluation was made, the highest performance was obtained in the proposed model in 20 images. The Squeezenet showed the lowest performance. When the 27 classes in the dataset were evaluated separately, the proposed model showed the highest performance in all classes. It has been observed that 27 different *Euphorbia* seed types belonging to the same species were successfully detected thanks to the developed CBIR-based system.

Declaration of Competing Interest

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

Consent statement/Ethical approval: Not required.

Funding Statement: Received no funding to support our paper.

Data availability: The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Authors' contributions: The authors have an equal amount of contribution to the paper.

References

Apriyanti D H, Arymurthy A M & Handoko L T (2013, November). Identification of orchid species using content-based flower image retrieval. In 2013 International conference on computer, control, informatics and its applications (IC3INA) (pp. 53-57). IEEE.

Auni A R & Sugiharti E (2024). Optimization of Mango Plant Leaf Disease Classification Using Concatenation Method of MobileNetV2 and DenseNet201 CNN Architectures. *Scientific Journal of Informatics* 11(4): 1023-1034

da Silva O L M, Cordeiro I & Caruzo M B R (2016). Seed morphology in Euphorbia and its taxonomic applications: a case study in São Paulo, Brazil. Brazilian Journal of Botany, 39, 349-358.

Davis P H (1965). 1985. Flora of Turkey and the Aegean Islands, Vol.: I-IX.

Dayrell L, Ott T, HorrocksT & Poschlod P (2023). Automated extraction of seed morphological traits from images. Methods in Ecology and Evolution, 14(7), 1708-1718.

Ding Z, Liu H, Demartino C, Feng M & Sun Z (2024). Neighborhood component analysis-based feature selection in machine learning to predict tendon ultimate stress of unbonded prestressed concrete beams. Case Studies in Construction Materials, 21, e03428.

Fayed A A & Hassan N M (2007). Systematic significance of the seed morphology and seed coat sculpture of the genus Euphorbia L.(Euphorbiaceae) in Egypt. Flora Mediterrinea 17: 47-64

Frodin D G (2004). History and concepts of big plant genera. Taxon, 53(3), 753-776.

Gawli K S & Gaikwad A S (2020). Deep learning for plant species classification. JETIR, 7(11), 99-105.

Gulzar Y, Hamid,Y, Soomro A.B, Alwan A A, & Journaux, L (2020). A convolution neural network-based seed classification system. Symmetry, 12(12), 2018.

Gyires-Tóth B P, Osváth M, Papp D & Szucs G (2019). Deep learning for plant classification and content-based image retrieval. Cybern. Inf. Technol, 19(1), 88-100.

Haupt J, Kahl S, Kowerko D & Eibl M (2018). Large-Scale Plant Classification using Deep Convolutional Neural Networks. CLEF (Working Notes), 3.

Huang Z, Wang R, Cao Y, Zheng S, Teng Y, Wang F & Du J (2022). Deep learning based soybean seed classification. Computers and Electronics in Agriculture, 202, 107393.

Ibrahim N M, Gabr D G I, Rahman A U, Dash S & Nayyar A (2022). A deep learning approach to intelligent fruit identification and family classification. Multimedia Tools and Applications 81(19): 27783-27798

Ibrahim N M, Gabr D G, Rahman A, Musleh D, AlKhulaifi D & AlKharraa M (2023). Transfer learning approach to seed taxonomy: a wild plant case study. Big Data and Cognitive Computing, 7(3), 128.

Karthikeyan M & Raja D (2023). Deep transfer learning enabled DenseNet model for content based image retrieval in agricultural plant disease images. Multimedia Tools and Applications, 82(23), 36067-36090.

Kumar M S, Neeraja P, Gunamaneni A, Nirmala A V & Nikhil G (2023, September). Deep Learning for Plant Species Classification. In 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET) (pp. 559-564). IEEE.

Kursat M, Başer B, Özbey F & Emre İ (2023). Seed morphology of some taxa of the genus Euphorbia L. (Euphorbiaceae) in Turkey and its taxonomic significance. *Turkish Journal of Botany* 47(4): 307-322

Kursat M, Yildirim M, & Emre I (2023). Classification of the images (Plant-21) in the dataset created with 21 different Euphorbia Taxons with the developed AI-based hybrid model. Signal, Image and Video Processing 17(8): 4153-4161

Lahitani A R, Permanasari A E & Setiawan N A (2016, April). Cosine similarity to determine similarity measure: Study case in online essay assessment. In 2016 4th International conference on cyber and IT service management (pp. 1-6). IEEE.

Loddo A, Loddo M & Di Ruberto C (2021). A novel deep learning based approach for seed image classification and retrieval. Computers and Electronics in Agriculture, 187, 106269

Lorenzo P, Luca P & Giorgio G. (2020). Convolutional neural networks for relevance feedback in content based image retrieval. Multimedia Tools and Applications, 79(37-38), 26995-27021

Muneesawang P & Guan L (2004). An interactive approach for CBIR using a network of radial basis functions. IEEE Transactions on multimedia, 6(5): 703-716

Pahlevani A H & Akhani H (2011). Seed morphology of Iranian annual species of Euphorbia (Euphorbiaceae). Botanical journal of the Linnean Society 167(2): 212-234

Pandey N, Krishna S, & Sharma S (2013). Automatic seed classification by shape and color features using machine vision technology. International Journal of Computer Applications Technology and Research 2(2): 208-213

Patil J K & Kumar R (2017). Analysis of content based image retrieval for plant leaf diseases using color, shape and texture features. Engineering in agriculture, environment and food 10(2): 69-78

Picek L, Šulc M, Patel Y & Matas J (2022). Plant recognition by AI: Deep neural nets, transformers, and kNN in deep embeddings. Frontiers in plant science, 13, 787527.

Sandler M, Howard A, Zhu M, Zhmoginov A & Chen L C (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

Saritha R R, Paul V & Kumar P G (2019). Content based image retrieval using deep learning process. Cluster Computing, 22(Suppl 2), 4187-4200.

Sunnetci K M, Ulukaya S & Alkan A (2022). Periodontal bone loss detection based on hybrid deep learning and machine learning models with a user-friendly application. Biomedical Signal Processing and Control, 77, 103844.

Sutaji D & Yıldız O (2022). LEMOXINET: Lite ensemble MobileNetV2 and Xception models to predict plant disease. Ecological Informatics, 70, 101698.

Tan M & Le Q (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

Theerthagiri P, Ruby A U, Chandran J G C, Sardar T H & Shafeeq BM A (2024). Deep SqueezeNet learning model for diagnosis and prediction of maize leaf diseases. *Journal of Big Data* 11(1), 112.

Toğaçar M (2022). Using DarkNet models and metaheuristic optimization methods together to detect weeds growing along with seedlings. Ecological Informatics, 68, 101519.

Yildirim M (2024). Content-Based Image Retrieval and Image Classification System for Early Prediction of Bladder Cancer. Diagnostics, 14(23): 2637

Yucel N & Yildirim M (2024). Automated deep feature fusion based approach for the classification of multiclass rice diseases. *Iran Journal of Computer Science* 7(1): 131-138

Yue J, Li Z, Liu L & Fu Z (2011). Content-based image retrieval using color and texture fused features. Mathematical and Computer Modelling 54(3-4): 1121-1127