

Article

GRASP-125: A Dataset for Greek Vascular Plant Recognition in Natural Environment

Kosmas Kritsis ^{1,*}, Chairi Kiourt ^{1,*}, Spyridoula Stamouli ¹, Vasileios Sevetlidis ¹, Alexandra Solomou ², George Karetsos ², Vassilis Katsouros ¹ and George Pavlidis ¹

- ¹ Athena Research Center, Institute for Language and Speech Processing, 15125 Athens, Greece; pstam@athenarc.gr (S.S.); vasiseve@athenarc.gr (V.S.); vsk@athenarc.gr (V.K.); gpavlid@athenarc.gr (G.P.)
² Institute of Mediterranean Forest Ecosystems, Hellenic Agricultural Organization DEMETER, 11528 Athens, Greece; solomou@fria.gr (A.S.); karetsos@fria.gr (G.K.)
* Correspondence: kosmas.kritsis@athenarc.gr (K.K.); chairiq@athenarc.gr (C.K.)

Abstract: Plant identification from images has become a rapidly developing research field in computer vision and is particularly challenging due to the morphological complexity of plants. The availability of large databases of plant images, and the research advancements in image processing, pattern recognition and machine learning, have resulted in a number of remarkably accurate and reliable image-based plant identification techniques, overcoming the time and expertise required for conventional plant identification, which is feasible only for expert botanists. In this paper, we introduce the GRASP dataset, a set of images composed of 125 classes of different species, for the automatic identification of vascular plants of Greece. In this context, we describe the methodology of data acquisition and dataset organization, along with the statistical features of the dataset. Furthermore, we present results of the application of popular deep learning architectures to the classification of the images in the dataset. Using transfer learning, we report 91% top-1 and 98% top-5 accuracy.

Keywords: deep learning; image classification; plant identification; transfer learning



Citation: Kritsis, K.; Kiourt, C.; Stamouli, S.; Sevetlidis, V.; Solomou, A.; Karetsos, G.; Katsouros, V.; Pavlidis, G. GRASP-125: A Dataset for Greek Vascular Plant Recognition in Natural Environment. *Sustainability* **2021**, *13*, 11865. <https://doi.org/10.3390/su132111865>

Academic Editor: Giulia Caneva

Received: 22 September 2021
Accepted: 25 October 2021
Published: 27 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Image-based plant identification is a relevantly recent interdisciplinary research field, combining computer science and botany. On the one hand, it addresses the current demands for biodiversity monitoring and conservation, and the growing interest in ecology and plant phenotyping. On the other hand, it covers the needs of the general public for plant species knowledge, in cases such as ecotourism.

The field of automated plant species identification presents some significant challenges to researchers, which need to be considered in each methodological step of the identification process. The first one is the large number of taxa, which represent thousands of classes. The vascular flora of Greece includes 5758 species and 1970 subspecies (native and naturalized), representing 6620 taxa, belonging to 1073 genera and 185 families [1,2]. This research focused on a particularly important region located in the mountainous complex of central Greece, which includes the National Parks of mountains Oiti and Parnassus. In Oiti National Park, around 1150 species and subspecies have been recorded, and in Parnassus National Park, around 854 species and subspecies. Within species, there are *significant morphological variations*, depending on developmental stage, the time of day, the season and even the location where each individual is growing.

These factors can affect the morphological characteristics of the whole plant (e.g., height) or the appearance of its individual organs, such as flowers, leaves and fruits at various stages in its life cycle. Quantitative descriptions of the growth of these complex and deforming plant organs are vital, requiring robust optical segmentation and identification methods [3–5] and tracking and scene flow estimation systems [6]. In some cases, images

may be captured under controlled conditions in the laboratory or greenhouses, but they are more likely to be taken in more challenging natural environments. Thus, automated image acquisition protocols are highly desirable, generating large numbers of images [7]. Moreover, visual differences of individual plant images may appear as results of different capture devices or of variations in focus, focal length, shooting angle and natural light. Inherently, differences are found among plants of the same species. There also exist *similarities among different species*, making the visual identification process difficult, even for experienced botanists.

Different image-based plant identification methods have been studied and applied for about two decades, analyzing images of different plant organs or organ features, which present various degrees of complexity. The most studied plant organ, especially in earlier works, has been the leaf, mostly those of trees, and flowers have also often been investigated. In the most recent years, methods shifted from restricted single-organ and single-view to more realistic multi-organ and multi-view [5,6,8]. More specifically, earlier studies required images of individual plant organs, such as leaves [9–11] or flowers [12,13] (see [14] for an extensive literature review), to identify plant specimens, and specific constraints were imposed for image acquisition, such as a controlled background behind the plant organ. More recent proposals, stimulated by the advancements in the areas of machine learning and computer vision, have been utilizing deep architectures, such as convolutional neural networks (CNNs), in order to analyze more complex images of plants, pushing research on plant identification towards more accurate, applicable and effective applications in natural environments. In such cases, the images may contain multiple organs captured from multiple viewpoints in a variety of complex and natural environments. Different approaches used for image-based plant identification, and different methods for image acquisition, are reflected on a number of benchmark datasets of plant images.

This paper introduces a *new dataset of Greek vascular plant images*, the GRASP-125, which will be publicly available for real-world plant identification applications. To the best of our knowledge, this is the first public dataset of its kind. The dataset consists of 125 species with an average of 130 images per class, which amounts to about 16,000 images. To date, research on plant identification has been either performed on proprietary datasets created in a controlled laboratory environment, or on “noisy” crowd sourced image sets. The open character of GRASP-125 will facilitate further developments in the domain. Furthermore, this paper presents a comparative analysis of the application of popular CNN architectures on the GRASP-125 dataset, with results that show top-1 accuracy of 91% and top-5 accuracy of 98% with pre-trained models (transfer learning).

The rest of the paper is organized as follows: Section 2 provides a quick overview of related studies regarding publicly available datasets of images for plant identification. Section 3 describes the GRASP-125 dataset and provides the methodology and specifications for data acquisition and organization. Next, in Section 4 we describe the employed transfer learning approach for training various deep learning models along with the experimental setup. In Section 5, we present results on the performance of the trained models. Finally, Section 6 summarizes the contributions of our dataset to the plant identification task, especially in natural environments, and concludes by listing some future challenges.

2. Related Work

The most important and widely used image datasets of plants consist either of (a) scans or pseudo-scans of leaves, as a result of scanning leaves in the laboratory or flattening and photographing them on a flat surface, such as laboratory tables with homogeneous background, or (b) images of different plant organs, such as leaves, flowers, stems and fruits, and more complex photos, for instance, branches of trees or the entire plant, captured in their natural environment.

Plant leaf datasets are very popular for the recognition of the plant species [13] and for identification of plant diseases [15]. Some popular datasets of *leaf images* are the following:

- The *Swedish leaf* (<http://www.cvl.isy.liu.se/en/research/datasets/swedish-leaf/>, accessed on 20 September 2021) dataset is the result of a leaf classification project of Linköping University and the Swedish Museum of Natural History [16]. The dataset contains leaf scans on a plain background of 15 Swedish tree species, with 75 leaves per species (total 1125 images).
- The *Flavia* (<https://sourceforge.net/projects/flavia/>, accessed on 20 September 2021) dataset consists of 1907 leaf images, either scans or pseudo-scans, of 32 species with 50–77 images per species. The images were captured in the campus of the Nanjing University and the Sun Yat-Sen arboretum, Nanking, China [17].
- The *Leafsnap* (<http://leafsnap.com/dataset/>, accessed on 20 September 2021) dataset contains 30,866 leaf images of 185 tree species of the Northeastern United States, along with automatically-generated segmented data [10]. The images fall into two categories: (a) “clean” laboratory images of pressed leaves captured with a high-quality digital camera (around 80% of the dataset), and (b) field images captured with mobile devices (around 20% of the dataset), of varying degrees of blur and illumination.
- The *ICL* (<http://english.iim.cas.cn>, accessed on 20 September 2021) dataset contains leaf images, scans and pseudo-scans, of 220 plant species, mainly trees and herbs, with 26 to 1078 images per species (total 17,032 images) [18]. Images were acquired at the Hefei Botanical Garden in Anhui, China by researchers from the Intelligent Computing Laboratory (ICL) at the Institute of Intelligent Machines.

Some of the most popular datasets containing *flower images* captured in natural environments are the following:

- The *Oxford Flower 17 and 102* (<https://www.robots.ox.ac.uk/~vgg/data/flowers/>, accessed on 20 September 2021) datasets contain images of common flowers of the UK, which belong to 17 and 102 herb species, respectively. Oxford Flower 17 contains 80 images per species with a total number of 1360 images, while Oxford 102 has a total of 8189 images, with 40 to 258 images per species [12,13]. Images were acquired by internet search and were complemented by field photos. All images depict flowers in their natural environment with significant variations in viewpoint, focus and illumination conditions. Moreover, they were chosen to represent morphologically unique species, and species very similar to each other.
- The *Jena Flower 30* (<https://doi.org/10.7910/DVN/QDHYST>, accessed on 20 September 2021) dataset contains images of 30 herb species during their flowering season, acquired in the area around Jena, Germany, with mobile devices [19]. Species are represented by 11 to 70 images with a total of 1479 images.

Apart from the above presented datasets that focus on individual organs, i.e., leaves and flowers, the richest source of training data for plant identification is provided by the Image Cross Language Evaluation Forum (ImageCLEF), which has been organizing a plant identification challenge since 2011. Each year, a dataset of plant images is provided, including a variety of species, organs and views, accompanied by a set of rich metadata, which include date, content type, GPS coordinates, etc. The initial challenge of 2011 provided a dataset of 5436 images of leaves (scans, pseudo-scans and free natural photos) from 71 tree species of the French Mediterranean area. In the following years, the PlantCLEF dataset grew to 1,132,015 images of trees, herbs and ferns, representing 1000 different taxa growing in France and in neighbouring countries (PlantCLEF2015/2016). These images were acquired through a crowd sourcing initiative by members of a social network of amateur and expert botanists who used the mobile application *Pl@ntNet* (<https://identify.plantnet.org>, accessed on 20 September 2021) to submit their observations. The images contain leaf scans, pseudo-scans and photographs of plant organs, namely, leaves, flowers, fruits, stems, branches and the entire plant. In LifeCLEF 2017 and 2018 challenges, impressive identification performance was achieved for over 10,000 species, mostly living in Europe and North America, whereas PlantCLEF2019 focused on the flora

of data deficient countries, providing a dataset of 10,000 species of the Guiana Shield and the Amazon rainforest (<https://www.imageclef.org>, accessed on 20 September 2021).

More recently, a study on plant identification of German flowering plants [20] introduced a new, partly crowd-sourced, image dataset acquired through a systematic collection protocol. The dataset comprises 50,500 images of 101 plant species of a balanced distribution of observations (100 images per class), since each individual was photographed from five predefined perspectives (entire plant, flower frontal, flower lateral, leaf top and leaf back). All images were collected using the Flora Capture smartphone application (<https://floraincognita.com>, accessed on 20 September 2021). The researchers suggested that the fusion of multi-organ images and a combination of different perspectives are beneficial for plant identification, especially in cases of inconspicuous species. This dataset is available upon request.

All the aforementioned datasets are summarized in Table 1. From this short overview of available image datasets for training plant identification algorithms, it is concluded that most of them, especially the earlier ones, contain images of individual plant organs, mainly leaves, acquired over a limited period of time, in a specific geographical area with very strict specifications. The latest datasets, and more specifically those that were released in the last 5 years of the PlantCLEF challenge, and the German flowering plants dataset, include more diverse and realistic images of multiple plant organs as they appear in their natural environments. This makes them more suitable for the development of reliable real-life plant identification applications (see [14,21] for a more detailed evaluation of the available datasets of plant images).

Table 1. Summary of the most popular image datasets for plant recognition.

Dataset	Num. of Taxa	Total Num. of Images	Organ	Life Form
Swedish leaf	15	1125	leaf	tree
Flavia	32	1907	leaf	tree
Leafsnap	185	30,866	leaf	tree
ICL	220	17,032	leaf	tree, herb
Oxford Flower 17	17	1360	flower	herb
Oxford Flower102	102	8189	flower	herb
Jena Flower 30	30	1479	flower	herb
German flowering plants	101	50,500	flower, leaf	herb
PlantCLEF2015/2016	1000	113,205	leaf, flower, fruit, stem	tree, herb, fern
GRASP-125	125	16,367	leaf, flower, fruit, stem	tree, herb, fern

The GRASP-125 dataset introduced with this paper is in line with state-of-the-art approaches to plant identification, following a multi-organ and multi-view approach to data acquisition. Compared to the majority of the available datasets, it contains a variety of life forms of the Greek flora, which include some important, rare and endemic species. Additionally, each species is represented by different organs and organ combinations, captured from multiple perspectives in a variety of shooting conditions. These characteristics make the GRASP-125 suitable to train algorithms for plant identification, which may serve real-life applications for biodiversity observation in the natural environment.

3. The GRASP-125 Dataset

The GRASP-125 (<http://advent.athenarc.gr/grasp/>, accessed on 20 September 2021) dataset has been created for the purpose of developing a plant identification mobile application for the general public, and more specifically, for hikers and nature enthusiasts who visit the National Parks of Oiti and Parnassus in central Greece. Both parks belong to the same mountainous complex and host a great wealth of flora. According to the AdVenT flora database [22], 819 plant taxa (species and subspecies) have been recorded in Oiti and Parnassus mountains at an altitude of over 1000 m. Apart from the quantitative aspect,

the flora of both Oiti and Parnassus include some rare, endemic and impressive herbaceous plant species of unique beauty, mostly during their blossom in early spring.

3.1. Methodology of Data Acquisition

The methodology of creating the GRASP-125 dataset was based on acquisition specifications, aiming to ensure realistic and complex scenarios of capturing and identifying plants as they appear in their natural environments. The goal was to establish a reliable set of plant images with a representative distribution of the species, in order to train an accurate and generalizable computational model suitable to serve the needs of a useful mobile application for the identification of plants in Oiti and Parnassus National Parks. The data collection specifications were the result of joint work between forest ecologists, botanists and computer scientists.

In terms of plant distribution, the areas chosen to perform data collection cover the mountainous region of the National Parks at an altitude of over 1000 m, where the most popular and interesting hiking trails, in terms of natural beauty and biodiversity, are found. Photos of plants were taken exclusively by the research team of the Institute of Mediterranean and Forest Ecosystems (IMFE), including researchers in the areas of ecology, systematic botany and forestry, experienced in field research in the region under study. The data collection process covered a period approximately between May and September.

In order to acquire images that represent a large variety of real-life scenarios, an elaborated manual containing specific guidelines for image capturing, with relevant illustrations, was provided to the team of data collectors. The main guidelines specify that photos should cover:

- Numerous individuals of the same plant taxa, found in different locations, at different times of the day and in different seasons.
- Different developmental stages of the same plant taxa.
- Different lighting conditions, excluding overlit and dark photos.
- Different organs of the same individuals, such as leaves, flowers, fruits and stem.
- Different parts of the plant, e.g., images of branches, and the whole plant.
- Simple photos, including a single organ, and complex photos, combining multiple organs (e.g., foliage) or different organs (e.g., foliage together with fruit and/or flowers), in the same frame as a natural background. It was noted that in case of complex photos there should be a main “theme,” focusing on a particular organ. In both cases, i.e., simple and complex photos, the “theme” should be centralized and should cover 70–80% of the frame.
- Multiple viewpoints and shooting angles (top, frontal and side views) of the same “theme”.
- Multiple zoom scales so to achieve, together with multiple viewpoints, a variety of visually distinguishable photos at least 30–40% different from one another.

The photos were captured by utilizing different devices, ranging from professional and high quality digital cameras, to mobile phone cameras. After the shooting session, the ground-truth, i.e., the correct taxon (species or subspecies) for all images, was assigned and validated by the team of botanists.

Further quality assurance of the collected field photos was performed by the team of computer scientists, who removed images unsuitable for inclusion in the dataset; these included too blurry or overlit photos or multiple, nearly identical shots of the same individual. As in many datasets of plant images [8], the distribution of the number of images per taxon was unequal. A few taxa were well represented, and many taxa had very few images. Given this limitation, an algorithm was developed for retrieving additional images using Web-crawling techniques by exploiting the YANDEX (<https://yandex.com/>, accessed on 20 September 2021) search engine and using the scientific names of taxa as search terms. The corpus collected with this method was screened by the botanists, who removed images corresponding to wrong taxa or disputed ones.

3.2. Dataset Description and Statistics

The GRASP-125 dataset consists of two complementary sets of plant images that follow the acquisition approaches presented in the previous section: (a) field photos from the region of interest and (b) images automatically collected through the Web and verified by expert botanists.

The life forms represented in the GRASP-125 dataset consist of trees, shrubs and herbs growing at an altitude of over 1000 m in Oiti and Parnassus in central Greece. The dataset includes images of plant taxa which are very unique morphologically, and of plant taxa with high degrees of visual similarity. The classes corresponding to the plant taxa, which are included in the dataset, represent plant species and subspecies. Depending on the visual similarity, images of the subspecies were incorporated into the corresponding species. More specifically, in cases in which two or more plant subspecies of the same species presented extreme visual similarity (are virtually indistinguishable even for the experts), their images were merged and they were considered of the same class, corresponding to the relevant plant species. In rare cases, where two or more plant subspecies presented extremely low similarity (again, according to the experts), they were considered as separate classes that correspond to the relevant plant subspecies. Some sample images from the GRASP-125 dataset are presented in Figure 1.



Figure 1. Samples from the GRASP-125 dataset.

The initial number of field photos collected by the botanists was 8082. Following a manual inspection procedure and after deleting inappropriate images, 5455 samples were kept, which were unequally distributed in classes (species), with an average value of 43.6 images and a standard deviation of 30.7. The set was largely unbalanced, as depicted in Figure 2a.

The Web-crawling technique resulted in 500 images per species, resulting in a total of 62,500 images. As expected, several of these images were not related to the species in query, requiring extensive verification by the experts. After screening the results, a total of 10,872 images were kept, corresponding to 113 of the total 125 classes with an average number of 96.2 images and a standard deviation of 93.1. This is still an unbalanced dataset as shown in Figure 2b.

At the final stage, the GRASP-125 dataset was formed by the union of the two aforementioned sets, including 16,327 images, which are still unbalanced but with an adequate representation of all included species. The least represented class comprises 40 images, whereas the most represented class contains 474 images. A sorted representation of the dataset is shown in Figure 2c, where only every other class is shown for better visualization. It is obvious that the relevant samples collected by the Web-crawling technique were not enough to balance the data, resulting an average of 130.6 images per class with a standard deviation of 96.3. At this point it should be highlighted that the entire dataset was re-scaled, while keeping the images' original aspect ratio and fixing the largest dimension to span 500 pixels.

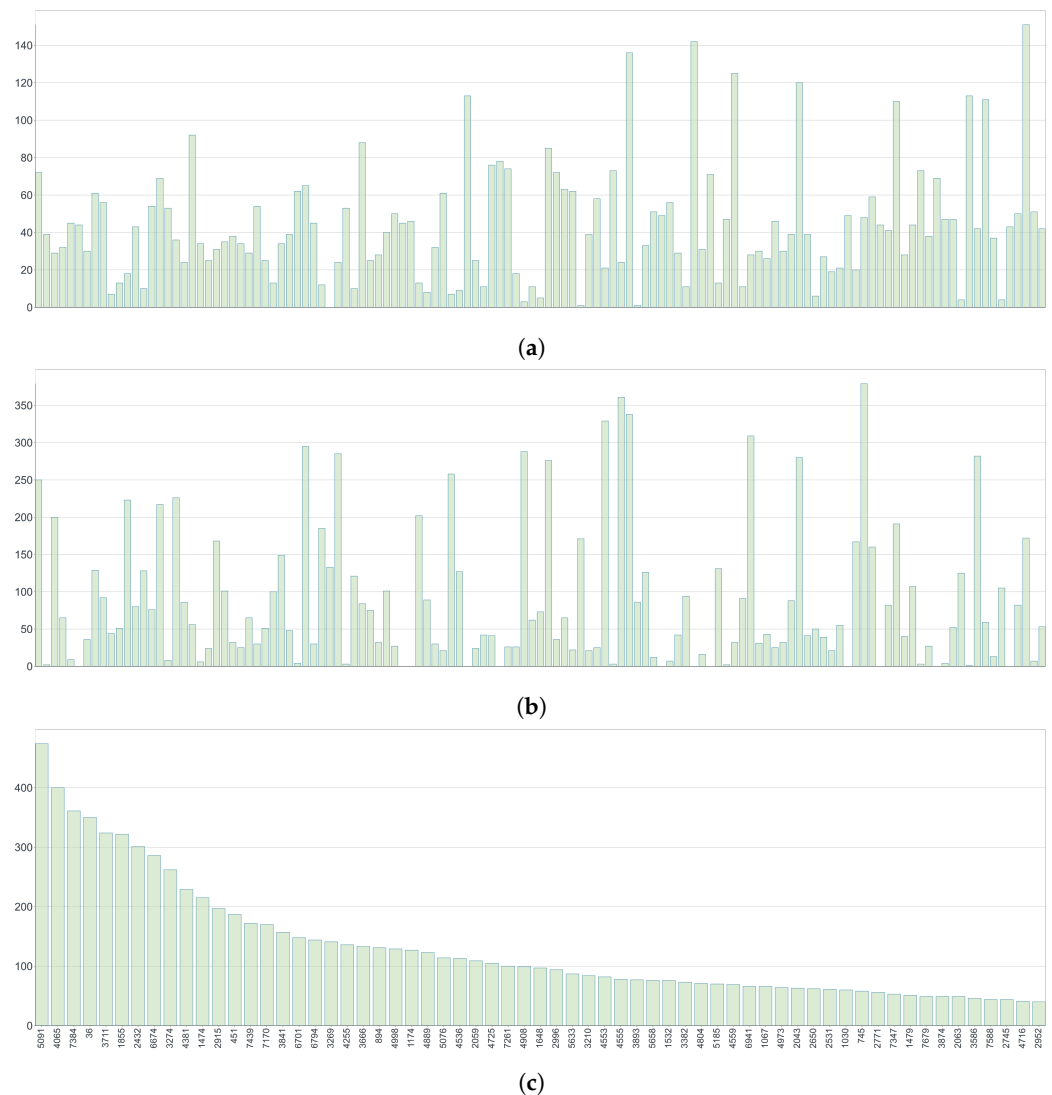


Figure 2. Distributions of the images that constitute the GRASP-125 dataset. (a) Distribution in classes of the collected field images. (b) Distribution in classes of the images retrieved with Web-crawling. (c) The profile of GRASP-125 sorted by class representation.

4. Experiments

Transfer learning [23] is a methodology of overcoming the restricted learning procedure and utilizing knowledge acquired for one task to solve related ones. This approach is often applied to solutions based on deep neural networks, since large amounts of data are usually required for faster convergence and better generalization of a model. Transfer learning involves the concepts of a domain and a task [24]. A domain D consists of a feature space \mathcal{X} and a marginal probability distribution $P(X)$ over the feature space, where $X = \{x_1, \dots, x_n\} \in \mathcal{X}$. For instance, in an image classification task, \mathcal{X} is the space of all image representations, x_i is the i -th image sample and X represents all the sample images used for training.

Given a domain, $D = \{\mathcal{X}, P(X), X \in \mathcal{X}\}$, a task T consists of a label space Y and a conditional probability distribution $P(Y|X)$ that is typically learned from the training data, consisting of pairs $x_i \in \mathcal{X}$ and $y_i \in Y$. In image classification tasks, Y is the set of all labels, i.e., the classes of the image dataset, and y_i is an image label. Given a source domain D_S , a corresponding source task T_S , a target domain D_T and a target task T_T , the objective of transfer learning is to enable the algorithm to learn the target conditional probability distribution $P(Y_T|X_T)$ in D_T with the information gained from D_S and T_S , where $D_S \neq D_T$ or $T_S \neq T_T$. In most cases, a limited number of labeled target examples,

which is exponentially smaller than the number of labeled source examples, is assumed to be available.

In this regard, given the heterogeneity, the unbalanced image data distribution per class and relatively medium number of images in the GRASP-125 dataset, transfer learning can help achieve higher classification rates in the plant identification task, with less technical skills and theoretical knowledge, and less time in contrast to the development of a new model. Given that transfer learning was chosen for the experiments conducted in this paper, the setup, basically, relied on fine tuning selected widely-used deep learning architectures, pre-trained and validated on the ImageNet dataset [25] VGG16 [26], InceptionV3 [27], ResNet50 [28], InceptionResNetV2 [29], MobileNetV2 [30], DenseNet121 [31], NASNetLarge and NASNetMobile [32] and EfficientNetB0 and EfficientNetB4 [33].

Fine tuning is the process in which a pre-trained model is stripped of its top layers, which are replaced by new and more appropriate layers to the new task. Then the model is re-trained with the dataset at hand. Figure 3 shows the considered architecture, which has already provided very good results in another application domain, tasks with similarities with plant identification. The red-colored block represents the input to the model along with the basic parameters (*batch size, image width, image height and image channels*). The blue-colored block depicts the pre-trained stack of layers of the considered popular architectures (pre-trained with ImageNet). The green-colored block highlights the new top layers, which are designed according to the new task.

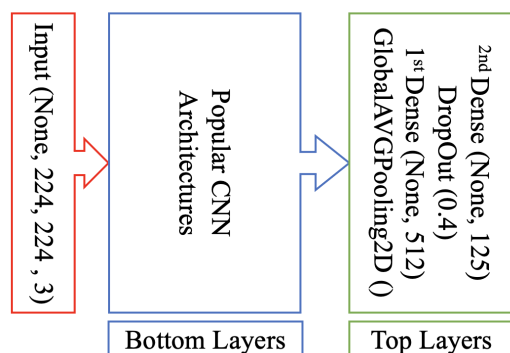


Figure 3. Transfer learning architecture.

The experiments were conducted with the machine learning framework TensorFlow (<https://www.tensorflow.org/>, accessed on 20 September 2021) and the Keras (<https://keras.io/>, accessed on 20 September 2021) high-level API. The bottom layers of the models were initialized with the weights resulted by the training with ImageNet. Then, the architectures were re-trained. Typically, in transfer learning only a small number of epochs is expected to be enough to attain convergence on the best possible result by any model. In our case we found that almost all models could not provide considerable accuracy improvements after around 25 epochs. We selected a batch size of 16 images, and the Adam optimizer [34] algorithm with learning rate of 0.0001. A 90–10% training-validation approach was employed, leading to using about 14,400 images for training and about 1600 images for evaluation.

5. Results and Discussion

Since the main goal of the presented dataset is to train a reliable image classifier with real-world images of plants which would be easily deployable as part of a mobile application, our results focus on answering two questions:

- Which architecture achieves the highest classification rate?
- How computationally expensive is this architecture for deployment on mobile devices?

By analyzing the accuracy rates depicted in Figure 4, it is evident that both EfficientNet architectures were quite successful, reporting the best performances with 92% and 91% top-

1 recognition accuracy, followed by NASNetMobile and InceptionResNetV2 architectures, with accuracy rates of 90% and 89%. MobileNetV2 and DenseNet121 reached 88% and 87%, respectively. The VGG16 architecture failed almost completely to fit its parameters to the objective, resulting the lowest accuracy rates.

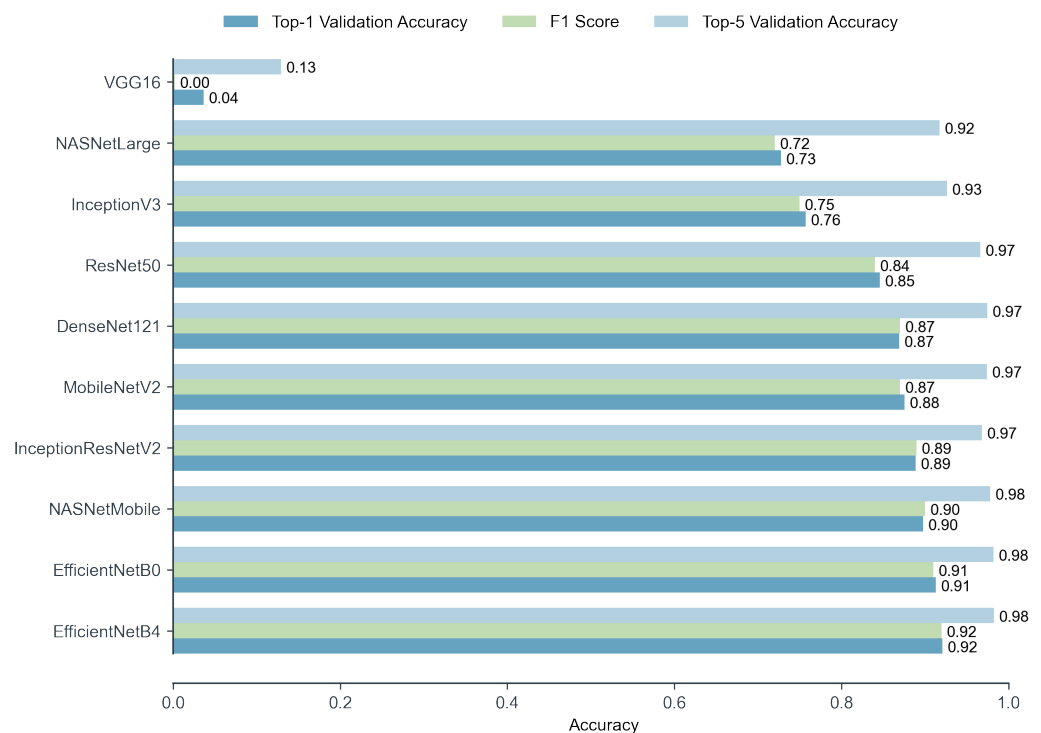


Figure 4. Validation accuracy (top-1 and top-5) and F1 score of the different models trained on the GRASP-125 dataset.

The training progress was monitored and validated by the categorical cross-entropy loss function and the classification accuracy of each model, as it is depicted in Figure 5a. Figure 5b shows the corresponding validation accuracy attained by each model. It should be highlighted that there is no reason to illustrate the corresponding training curves (training loss and accuracy), since their evolution was quite smooth and all models converged before reaching the 25th epoch.

Given the technical specifications and requirements of a mobile application, the ideal model should be relatively fast and able to allocate computational resources efficiently, while providing high accuracy rates for the task at hand. Taking into consideration the total number of required parameters in the different statistical objective functions presented in Table 2 and the accuracy of each model presented in Figure 4, we can remark that:

- MobileNetV2 (38 MB), EfficientNetB0 (60 MB), NasNetMobile (74 MB) and DenseNet121 (95 MB) can be considered as the models with fewer trainable parameters but with relatively high classification accuracy performances.
- Larger models, such as NASNetLarge (1020 MB) and InceptionResNetV2 (645 MB), are more suitable for server-side deployments, due to their size and overall high computational resource requirements.
- Models with more complex architectures, such as NasNetLarge and EfficientNetB4, tend to require more training epochs to converge (see Figure 5a).
- EfficientNetB0 and EfficientNetB4 are very accurate and are medium-sized models, about 60 MB and 214 MB, respectively.

Consequently, we can conclude that the most suitable architecture that balanced the trade-off between the resource consumption efficacy and recognition accuracy rates was EfficientNetB0.

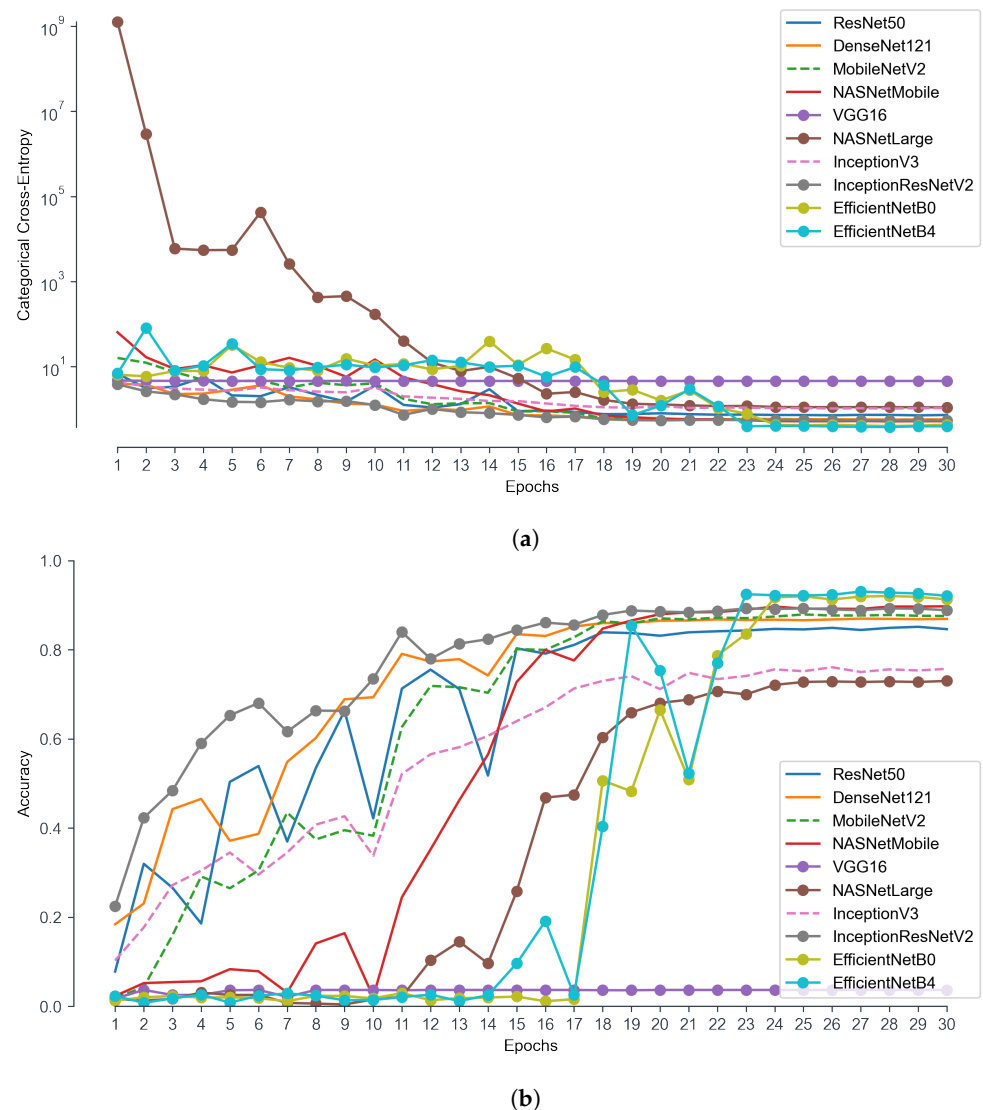


Figure 5. Validation loss and accuracy curves for the considered architectures. (a) Validation loss; (b) validation accuracy.

Table 2. Number of parameters for each model.

Models	Parameters (Million)	Trainable Parameters (%)
ResNet50	24.7	99.8
DenseNet121	7.6	98.9
MobileNetV2	3.0	98.8
NASNetMobile	4.9	99.2
VGG16	15.0	100.0
NASNetLarge	87.0	99.8
InceptionV3	22.9	99.8
InceptionResNetV2	55.2	99.9
EfficientNetB0	4.8	99.1
EfficientNetB4	18.7	99.3

An additional approach to evaluating the performance of trained models on unbalanced datasets is to analyze their ability in recognizing the different target classes correctly. To this end, by selecting the model with the highest reported validation accuracy, we further

evaluated its performance over each individual class. According to the EfficientNetB4 architecture, the 10 classes of the GRASP-125 that resulted the lowest recognition rates are shown in Figure 6. A detailed confusion matrix of all classes can be found online (<http://advent.athenarc.gr/grasp/>, accessed on 20 September 2021). We expected that most of these classes would be represented by fewer image examples. However, further investigation indicated that this is not entirely true. For instance, class “3269” (*Calystegia silvatica* (Kit.) Griseb.) included 141 image examples and class “114” (*Abies cephalonica* Loudon) included 132 images.

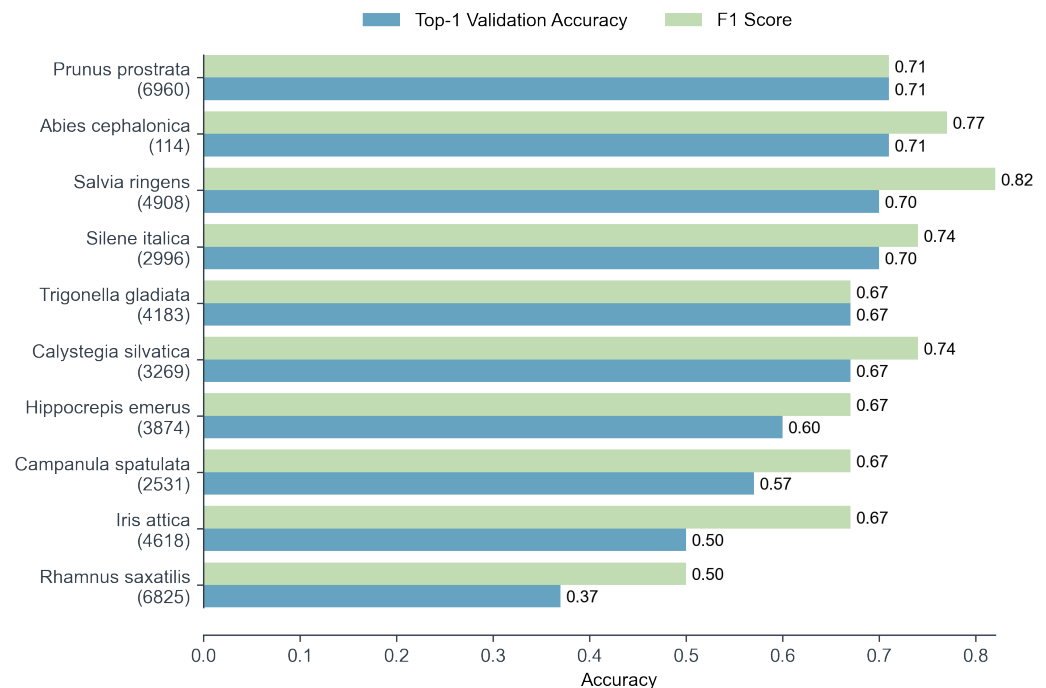


Figure 6. The 10 classes in which EfficientNetB4 performed poorly.

Furthermore, by inspecting the misclassified images that belong to *Calystegia silvatica*, we discovered that the majority was identified as *Convolvulus arvensis* L., both belonging to the same family of *Convolvulaceae*. In order to understand deeper and explain the decisions of the trained model on the provided classification predictions, we have adopted the SmoothGrad [35] and GradCAM++ [36] (an improved version of GradCAM [37]) algorithms, which provide visualizations of the learned features by exploiting class activation maps.

In this regard, for each of the two aforementioned classes we chose to visualize the activation maps of two images, one that was identified correctly to its actual class and one that was confused as belonging to the other related *Convolvulaceae* species, as it is illustrated in Figure 7. The dashed lines over the presented heat-maps mark the areas of attention for the model, in order to determine the class of the different input images. The color gradients of the heat-maps, provide additional information regarding the semantic importance of the image features, ranging from deep blue for low importance, to deep red that corresponds to the most important pixel features.

In the GRASP-125 dataset, the majority of images that represent the *Convolvulus arvensis* species contain more than a single flower within the image frame, and *Calystegia silvatica* includes images that mostly depict a single flower. By inspecting the GradCAM++ image of the correctly predicted *Convolvulus arvensis*, we observe that the model identified multiple areas of importance. Similarly, the *Calystegia silvatica* that was misclassified as *Convolvulus arvensis*, activated various areas, thereby affecting the model towards the wrong decision.

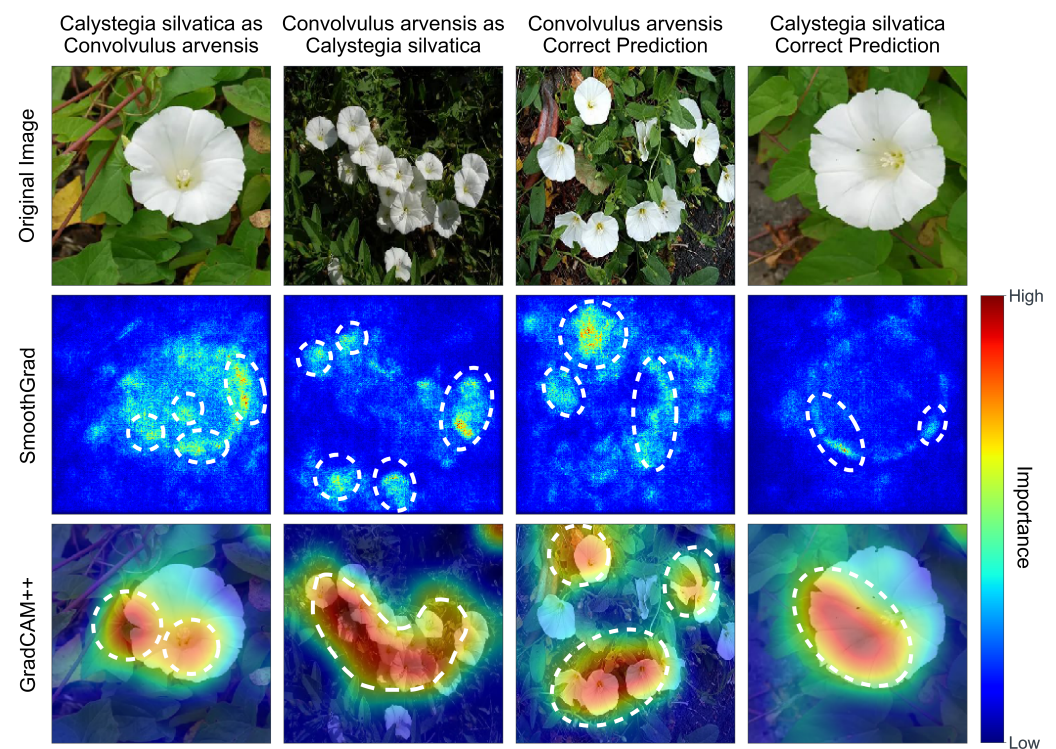


Figure 7. SmoothGrad and GradCAM++ visualisations of the learned features for the *Calystegia silvatica* and *Convolvulus arvensis* species.

From a taxonomic perspective, *Convolvulus arvensis* has twining or prostrate stems, leaves ovate-oblong to lanceolate, funnel-shaped flowers, usually solitary or in cymes of 2–3 flowers, with small bracts and corolla 1.9–2.5 cm long in white. The stigma is filiform and pollen is tricolpate (has three ridge-like apertures). *Calystegia sylvatica* has twining stems, leaves ovate to broad-ovate, shortly acuminate with apex acute to narrow-obtuse, flowers solitary, calyx hidden by bracts, corolla 4–7 cm long, white, or rarely pinkish. Stigma is clavate (swollen at the top) and pollen is pantoporate [38]. Within the *Calystegia* group, floral morphology is relatively invariant, and the lineage is composed of self-incompatible, white-flowered perennials that are prone to introgression. *Calystegia* is well separated from the genus *Convolvulus* and forms a monophyletic group. However, a worldwide contemporary monographic evaluation for this genus is not yet available, while there is still a conflict on whether it should be merged with the genus *Convolvulus* or remain a separate genus [39]. Further analysis of molecular data has indicated that it is in fact nested in the *Convolvulus* [40], although a recently published monograph of the *Convolvulus* genus did not include *Calystegia*, for pragmatic reasons [41]. Overall we can say that it is a taxonomically difficult species complex, in which hybridization and introgression frequently occur, thereby entangling the process of visually identifying the two species based solely on morphological characters, which can be further considered as a quite challenging task for the trained computational model.

6. Conclusions

Plant identification is a very challenging task in the fields of machine learning and computer vision, due to the morphological complexity of the plants and the similarities that several plant species present. In this paper, we introduced the *GRASP-125* dataset that is in line with the recent advancements in the field of image-based plant identification, which follow a more realistic, accurate and applicable multi-organ and multi-view approach to data acquisition. The proposed dataset contains images of various taxa of plants of the mountain flora of Greece, some of which are not so well-known to the public, and include special characteristics. The dataset contains various life forms, namely, trees, shrubs and

herbs. Each taxon is represented by images of different individuals, which depict different plant organs, namely, leaves, flowers, fruits and stems/branches, and combinations of organs, in a variety of structured views (top, front, side, entire plant views), backgrounds and shooting conditions, which represent realistic, natural and complex real-life scenarios of plants capturing.

By using popular deep learning models and transfer learning for the classification of the GRASP-125 dataset we achieved 92% top-1 and 98% top-5 accuracy. The brief comparative analysis of the popular pre-trained architectures showed that almost all of them can perform quite well in GRASP-125 dataset, except the VGG16 model. When the objective is the application in mobile devices, some models presented several advantages, such as the MobileNetV2, EfficientNetB0, NasNetMobile and DenseNet121. These models can be considered favorable for plant identification mobile applications, due to their size and low number of parameters (faster processing). On the other hand, more complex architectures, though more accurate, are computationally expensive; thus, they are favorable for server-side applications and online services.

As stated in this paper, the development of the dataset is still ongoing in two directions: (a) to increase its size by introducing new classes and (b) to balance more appropriately the distribution of the images per class, by enriching existing classes with new images. Moreover, based on the experimental results we will focus to increase the top-1 accuracy of the presented model with further investigation of the considered pre-trained models and by developing new architectures.

Author Contributions: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, and funding acquisition, K.K., C.K., S.S., V.S., A.S., G.K., V.K. and G.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been co-financed by the European Union and Greek national funds through the Operational Programme “Competitiveness, Entrepreneurship and Innovation”, under the call “RESEARCH—CREATE—INNOVATE” (project code: T1EDK-03844, project title: AdVENT—Augmented Visitor Experience in National Parks). The authors would like also to thank their colleagues at the Institute of Mediterranean Forest Ecosystem for their valuable support in collecting, organizing and verifying field images from the considered plant species.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <http://advent.athenarc.gr/grasp/> (accessed on 20 September 2021).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Dimopoulos, P.; Raus, T.; Bergmeier, E.; Constantinidis, T.; Iatrou, G.; Kokkini, S.; Strid, A.; Tzanoudakis, D. *Vascular Plants of Greece: An Annotated Checklist*; Botanic Garden and Botanical Museum Berlin-Dahlem Berlin: Berlin, Germany, 2013; Volume 31.
2. Dimopoulos, P.; Raus, T.; Bergmeier, E.; Constantinidis, T.; Iatrou, G.; Kokkini, S.; Strid, A.; Tzanoudakis, D. Vascular plants of Greece: An annotated checklist. Supplement. *Willdenowia* **2016**, *46*, 301–347. [[CrossRef](#)]
3. Dobrescu, A.; Valerio Giuffrida, M.; Tsiftaris, S.A. Leveraging Multiple Datasets for Deep Leaf Counting. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops, Venice, Italy, 22–29 October 2017.
4. Giuffrida, M.V.; Minervini, M.; Tsiftaris, S. Learning to Count Leaves in Rosette Plants. In Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP), Swansea, UK, 10 September 2015; pp. 1.1–1.13. [[CrossRef](#)]
5. Dobrescu, A.; Valerio Giuffrida, M.; Tsiftaris, S.A. Understanding Deep Neural Networks for Regression in Leaf Counting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 16–20 June 2019.

6. Schunck, D.; Magistri, F.; Rosu, R.A.; Cornelißen, A.; Chebroly, N.; Paulus, S.; Léon, J.; Behnke, S.; Stachniss, C.; Kuhlmann, H.; et al. Pheno4D: A spatio-temporal dataset of maize and tomato plant point clouds for phenotyping and advanced plant analysis. *PLoS ONE* **2021**, *16*, e0256340. [[CrossRef](#)] [[PubMed](#)]
7. Kolhar, S.; Jagtap, J. Plant trait estimation and classification studies in plant phenotyping using machine vision—A review. *Inf. Process. Agric.* **2021**, in press. [[CrossRef](#)]
8. Joly, A.; Goëau, H.; Bonnet, P.; Bakić, V.; Barbe, J.; Selmi, S.; Yahiaoui, I.; Carré, J.; Mouysset, E.; Molino, J.F.; et al. Interactive plant identification based on social image data. *Ecol. Inform.* **2014**, *23*, 22–34. [[CrossRef](#)]
9. Fiel, S.; Sablatnig, R. Automated identification of tree species from images of the bark, leaves or needles. In Proceedings of the Computer Vision Winter Workshop, Mitterberg, Austria, 2–4 February 2011.
10. Kumar, N.; Belhumeur, P.N.; Biswas, A.; Jacobs, D.W.; Kress, W.J.; Lopez, I.C.; Soares, J.V.B. Leafsnap: A Computer Vision System for Automatic Plant Species Identification. In *Computer Vision—ECCV 2012*; Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 502–516.
11. Sulc, M.; Matas, J. Texture-Based Leaf Identification. In *Computer Vision—ECCV 2014 Workshops*; Agapito, L., Bronstein, M.M., Rother, C., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 185–200.
12. Nilsback, M.; Zisserman, A. A Visual Vocabulary for Flower Classification. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; Volume 2, pp. 1447–1454.
13. Nilsback, M.; Zisserman, A. Automated Flower Classification over a Large Number of Classes. In Proceedings of the 6th Indian Conference on Computer Vision, Graphics Image Processing, Bhubaneswar, India, 16–19 December 2008; pp. 722–729.
14. Wäldchen, J.; Mäder, P. Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review. *Arch. Comput. Methods Eng.* **2017**, *25*, 507–543. [[CrossRef](#)] [[PubMed](#)]
15. Jadhav, S.; Udipi, V.; Patil, S. Identification of plant diseases using convolutional neural networks. *Int. J. Inf. Technol.* **2020**. [[CrossRef](#)]
16. Söderkvist, O. Computer Vision Classification of Leaves from Swedish Trees. 2001. Available online: <http://www.diva-portal.org/smash/get/diva2:303038/FULLTEXT01.pdf> (accessed on 20 September 2021).
17. Wu, S.G.; Bao, F.S.; Xu, E.Y.; Wang, Y.; Chang, Y.; Xiang, Q. A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network. In Proceedings of the 2007 IEEE International Symposium on Signal Processing and Information Technology, Giza, Egypt, 15–18 December 2007; pp. 11–16.
18. Hu, R.; Jia, W.; Ling, H.; Huang, D. Multiscale Distance Matrix for Fast Plant Leaf Recognition. *IEEE Trans. Image Process.* **2012**, *21*, 4667–4672.
19. Seeland, M.; Rzanny, M.; Alaqrara, N.; Wäldchen, J.; Mäder, P. Plant species classification using flower images—A comparative study of local feature representations. *PLoS ONE* **2017**, *12*, e0170629. [[CrossRef](#)]
20. Rzanny, M.; Mäder, P.; Deggelmann, A.; Chen, M.; Wäldchen, J. Flowers, leaves or both? How to obtain suitable images for automated plant identification. *Plant Methods* **2019**, *15*, 1–11. [[CrossRef](#)] [[PubMed](#)]
21. Wäldchen, J.; Rzanny, M.; Seeland, M.; Mäder, P. Automated plant species identification—Trends and future directions. *PLOS Comput. Biol.* **2018**, *14*, e1005993. [[CrossRef](#)] [[PubMed](#)]
22. Solomou, A.; Karetos, G.; Trigas, P.; Proutsos, N.; Avramidou, E.; Korakaki, E.; Kougioumtzis, K.; Goula, A.; Pavlidis, G.; Stamouli, S.; et al. Vascular Plants of Oiti and Parnassos National Parks of Greece, as Important Components of Biodiversity and Touring Experiences. In Proceedings of the 9th International Conference on Information and Communication Technologies in Agriculture, Food & Environment, Thessaloniki, Greece, 24–27 September 2020.
23. Torrey, L.; Shavlik, J. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques*; IGI Global: Hershey, PA, USA, 2009; pp. 242–264.
24. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
25. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS'12), Lake Tahoe, NV, USA, 3–6 December 2012; Curran Associates Inc.: Red Hook, NY, USA, 2012; Volume 1, pp. 1097–1105.
26. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 7–9 May 2015.
27. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [[CrossRef](#)]
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [[CrossRef](#)]
29. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
30. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. [[CrossRef](#)]

31. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
32. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning Transferable Architectures for Scalable Image Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8697–8710.
33. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning (ICML 2019), Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
34. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
35. Smilkov, D.; Thorat, N.; Kim, B.; Viégas, F.B.; Wattenberg, M. SmoothGrad: Removing noise by adding noise. *arXiv* **2017**, arXiv:1706.03825.
36. Chattopadhyay, A.; Sarkar, A.; Howlader, P.; Balasubramanian, V.N. Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 839–847. [[CrossRef](#)]
37. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–27 October 2017; pp. 618–626. [[CrossRef](#)]
38. Lewis, W.H.; Oliver, R.L. Realignment of Calystegia and Convolvulus (Convolvulaceae). *Ann. Mo. Bot. Gard.* **1965**, *52*, 217–222. [[CrossRef](#)]
39. Spaulding, D.D. Key to the bindweeds (Calystegia and Convolvulus, Convolvulaceae) of Alabama and adjacent States. *Phytoneuron* **2013**, *83*, 1–12.
40. Stefanović, S.; Austin, D.F.; Olmstead, R.G. Classification of Convolvulaceae: A phylogenetic approach. *Syst. Bot.* **2003**, *28*, 791–806.
41. Wood, J.R.; Williams, B.R.; Mitchell, T.C.; Carine, M.A.; Harris, D.J.; Scotland, R.W. A foundation monograph of Convolvulus L.(Convolvulaceae). *PhytoKeys* **2015**, *51*, 1–282. [[CrossRef](#)]