Journal of Universal Computer Science, vol. 29, no. 9 (2023), 1069-1089
submitted: 10/9/2022, accepted: 2/5/2023, appeared: 28/9/2023

PlantKViT: A Combination Model of Vision Transformer and KNN for Forest Plants Classification

Nguyen Van Hieu
(The University of Danang - University of Science and Technology, Danang, Vietnam
https://orcid.org/0000-0001-5311-806, nvhieuqt@dut.udn.vn)

Ngo Le Huy Hien
(Leeds Beckett University, Leeds, United Kingdom
https://orcid.org/0000-0002-1439-8289, N.Hien2994@student.leedsbeckett.ac.uk)

Luu Van Huy
(The University of Danang - University of Science and Technology, Danang, Vietnam
https://orcid.org/0000-0002-0590-8582, lvhuy@dut.udn.vn)

Nguyen Huy Tuong
(The University of Danang - University of Science and Technology, Danang, Vietnam
https://orcid.org/0000-0003-1217-2195, huytuong010101@gmail.com)

Pham Thi Kim Thoa
(The University of Danang - University of Science and Technology, Danang, Vietnam
https://orcid.org/0000-0003-4988-7864, ptkthoa@dut.udn.vn)

Abstract: The natural ecosystem incorporates thousands of plant species and distinguishing them is normally manual, complicated, and time-consuming. Since the task requires a large amount of expertise, identifying forest plant species relies on the work of a team of botanical experts. The emergence of Machine Learning, especially Deep Learning, has opened up a new approach to plant classification. However, the application of plant classification based on deep learning models remains limited. This paper proposed a model, named PlantKViT, combining Vision Transformer architecture and the KNN algorithm to identify forest plants. The proposed model provides high efficiency and convenience for adding new plant species. The study was experimented with using Resnet-152, ConvNeXt networks, and the PlantKViT model to classify forest plants. The training and evaluation were implemented on the dataset of DanangForestPlant, containing 10,527 images and 489 species of forest plants. The accuracy of the proposed PlantKViT model reached 93%, significantly improved compared to the ConvNeXt model at 89% and the Resnet-152 model at only 76%. The authors also successfully developed a website and 2 applications called 'plant id' and 'Danangplant' on the iOS and Android platforms respectively. The PlantKViT model shows the potential in forest plant identification not only in the conducted dataset but also worldwide. Future work should gear toward extending the dataset and enhance the accuracy and performance of forest plant identification.

Keywords: Forest plants, Plant identification, Resnet-152, ConvNeXt, Transformer-Learning, Deep learning models, K- nearest-neighbor
Categories: I.0, I.2, I.3, I.4, I.5, I.6, J.3
DOI: 10.3897/jucs.94657

1 Introduction

Plants are an essential source of oxygen, raw materials, and nourishment for human life. The diversity of plant species plays a significant role in various fields, including food, industrial growth, medical science, and environmental protection [Chen et al., 2021]. Expertise in plant species is needed to exhaustively identify new, rare, or economically valuable plant species to support ecosystems and promote industries, sustainability, and energy productivity in agriculture. In addition, the increase in human productive activities, excessive logging, rapid urban development, global warming, and awareness of plant species have disrupted the ecological habitat of living organisms, leading to the extinction of hundreds of plant species every year. The extinction of a large number of plant species will cause severe consequences for humans and ecosystems, such as floods, flash floods, desertification, and continuous global climate change [Rawat et al., 2015]. Therefore, understanding and classifying plant species also helps people to protect them, which is also protecting human lives. Vietnam is a country with a forest system covering nearly half of the area, with biodiversity covering the length of the country. Excluding the plant species that have been found and recorded, many rare and valuable forest plant species have yet to be found [Minh-Hoang et al., 2015]. Consequently, the classification of plant species is a matter of current concern for economic development, which prompted the interest of this research.

The task of identifying plant species in Vietnam in particular and globally, in general, is manual, complicated, and time-consuming work, requiring specialized experts in plant management [Hieu et al., 2020b]. Therefore, it is critical to developing systems and models that can conduct the work automatically, saving time and other resources. At present, there are many methods of plant classification, such as plant genetics, plant chemical analysis, and plant cell analysis. However, only botanist researchers can apply these methods of classification or identification, which is unlikely to satisfy the need that everyone desires to quickly identify the type of plant or its origin [Selvaraju et al., 2017]. The advancement of pattern recognition and digital image processing has led to the emergence of object recognition technologies based on image processing in human life, including face and fingerprint recognition [Dougherty et al., 2020]. This technique provides a sufficient theoretical foundation and technological framework for image-based plant identification.

Many studies have been conducted in the past decade to build plant classification systems with positive findings [Selvaraju et al., 2017, Hieu et al., 2020b, Angelova et al., 2013]. Traditionally, worldwide researchers used leaves as a common physical characteristic to identify between distinct species, utilizing texture, color, and shape. Aakif et al. proposed a plant classification approach that involves three phases, starting with preprocessing to extraction and sorting. They classified the plant morphological features, shape, and Fourier descriptions, using an Artificial Neural Network (ANN). On 817 distinct leaf samples from 14 fruit trees, the system achieved an accuracy of more than 96% [Aakif et al., 2015]. Jeon and colleagues presented a new approach for classifying leaves by employing a Convolutional Neural Network (CNN), alongside the other 2 models utilizing GoogleNet to change network depth. Although the tested plant was 30% damaged, the models performed with 94% accuracy [Jeon et al., 2017]. These studies, however, contain limitations in terms of noise and background factors that influence low-level picture representation. This is because photos manually processed in a lab and subsequently classified yield a greater degree of accuracy than images shot on smartphones [Carranza-Rojas et al., 2016].

Therefore, it is difficult to use well-clean input photos with no background that the above studies used in actual applications. As a result, various scholars have both theoretically and empirically worked on developing a high-level representation of photos with minimal environmental impact. Sun et al. built the BJFU100 dataset by taking 10,000 photographs of 100 plant species around the Beijing Forestry University campus using mobile phones in an effort to construct a plant image dataset in the natural environment. A 26-layer deep learning model with 8 residual building blocks was used to produce an unsupervised plant classification system. In the end, they achieved a classification accuracy of 91.78% [Sun et al., 2017]. Following the trend, Al-Qurran et al. combined transfer learning and data augmentation on a dataset of plant species in wildlife to get a relatively good accuracy of 78.76% [Al-Qurran et al., 2018]. It can be observed that utilizing photos taken in nature makes it more difficult for plant classification algorithms to perform accurately than using datasets generated in labs.

Contributing to the research of improving the classification issues of plant images taken in the real environment, this study assembled a dataset named DanangForestPlant, which contains 10,527 images of 489 different plant species in 4 research areas in Danang, Vietnam. The dataset represents the botanical diversity in the city and is supplemented by other image sources to ensure training efficiency. Moreover, this research proposed a model called PlantKViT by using Resnet-152 and ConvNeXt networks to classify forest plants. Results showed that the PlantKViT model reached 93% in accuracy, significantly improved compared to the ConvNeXt model at 89% and the Resnet-152 model at only 76%. The outcomes are promising to apply to other conducted datasets worldwide.

The paper is organized as firstly delivering related studies in Section 2, before presenting recommended resources and methods in Section 3. Experiment with the DanangForestPlant dataset, the accuracy of the proposed model and comparison with other models is detailed in Section 4, before providing Conclusions in Section 5

2 Related Works

2.1 ResNet-152 Model - Deep Residual Learning for Image Recognition

In the past, automatic identification of plant species was often solved by mandating a photograph of specific plant organs, such as leaves [Kumar et al., 2018, Fiel et al., 2011, Sulc et al., 2014], flowers [Mattos et al., 2014, Li et al., 2020, Angelova et al., 2013], or bark [Fiel et al., 2011, Sulc et al., 2013, Boudra et al., 2015]. Furthermore, some of these datasets and systems have set additional constraints to the input image, like a white background behind the leaf image. In recent years, CNN has been successful in several computer vision tasks, particularly those involving the identification and detection of complex objects. The CNN models tested on Plant CLEF 2015 [Goeau et al. 2015] were significantly superior to the combination of older models. Inspired by those successes, this study built a model on top of a Deep Convolutional Neural Networks architecture called Resnet.

A very deep network has the benefit of being able to represent extremely complex functions since it can learn the distinctive features of images at various levels of abstraction. However, performing with deep networks is complex since it leads to the vanishing gradient issue [Alzubaidi, 2021]. Convolutional neural networks (CNN) have been applied successfully in image classification and pattern recognition techniques [Cui, 2018, Zhou et al., 2015, Hien et al., 2021, Hieu et al., 2020a]. Deep neural networks, on the other hand, require a large amount of data to avoid overfitting. Therefore, alternative approaches are desperately needed when the training data is limited and insufficient

[Zhao, 2017, Cogswell, 2016, Hieu et al., 2020b, Hien et al., 2020]. And one of them is the ResNet - Deep Residual Learning model.

ResNet (Residual Network) was introduced to the public in 2015 by [He et al., 2016] and even won first place in the 2015 ILSVRC competition [Russakovsky et al., 2015] with the top 5 error rate of only 3.57%. On the ImageNet dataset, the residual block of the ResNet model has a depth of up to 152 layers, 8 times deeper than the VGG grid but still holds lower complexity. This promising result has prompted this study to construct a model based on ResNet-152 architecture (Figure 1).

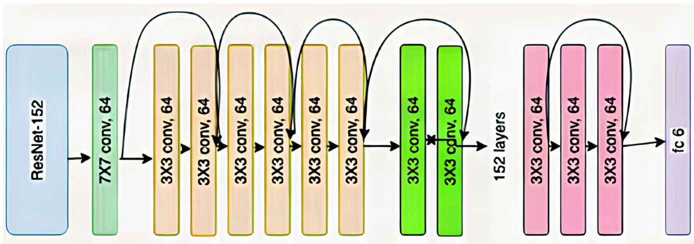


Figure 1: Resnet-152 Architecture

2.2 ConvNet Model

Neural networks, especially convolutional neural networks (ConvNets or CNNs) are a huge advancement and impact on the domain of Deep Learning. Many variants of ConvNets have been rapidly developed such as VGGNet [Szegedy et al., 2015], Inceptions [Szegedy et al., 2016], ResNet [He et al., 2016], DenseNet [Huang et al., 2017], MobileNet [Howard et al., 2017], EfficientNet [Tan et al., 2020] and RegNet [Xuet et al., 2021] focuses on various aspects of accuracy, efficiency, and scalability. ConvNets models almost dominate the field of computer vision.

Meanwhile, designing neural networks for natural language processing (NLP) is in a very different position, because Transformers replace ConvNets as the dominant architecture. Despite the almost complete difference between the two fields of computer vision and natural language processing, when the introduction of Vision Transformers (ViT) was announced in [Dosovitskiy et al., 2021] of Google Research, Brain Team created the convergence point of the two fields. The scientists were inspired by Transformers' scaling successes in NLP, and they applied them directly to the image with as little modification as possible. The best Vision Transformers model achieved 88.55% accuracy on ImageNet.

However, this does not necessarily mean that Vision Transformers have completely overwhelmed the ConvNets models. A sliding window self-attention mechanism implemented improperly could be very expensive [Ramachandran et al., 2019]. Speed could be optimized using sophisticated techniques like cyclic shifting [Liu et al., 2021a], but the system becomes more complex in design. Therefore, in some cases, ConvNet has met the requirements of the issue, although using a simpler and no-frills model.

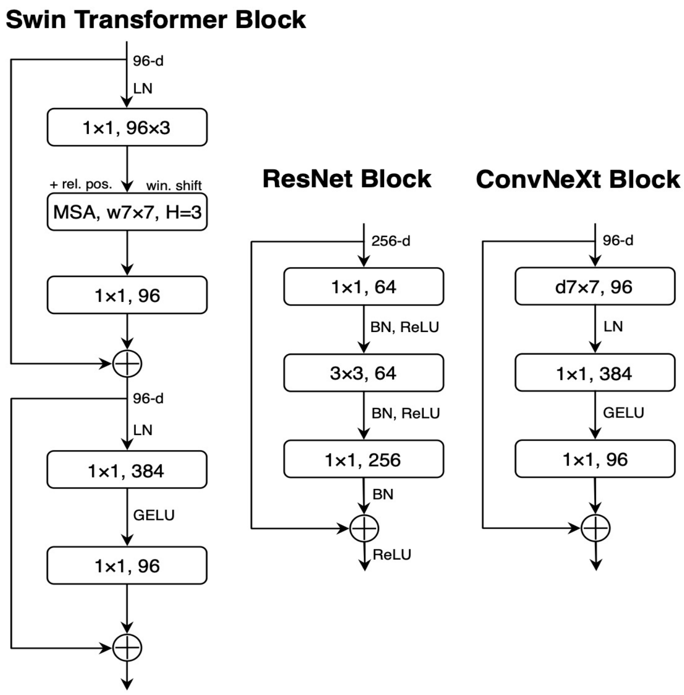


Figure 2: Block architecture of ResNet, Swin Transformer, and ConvNeXt [Liu et al., 2022]

2.3 ConvNeXt Model

Because of the neglect to fully exploit the performance of both ConvNet and Transformers in today's increasingly complex classification problems, research [Liu et al., 2022] by the Facebook AI Research team investigated the architectural differences between ConvNets and Transformers and attempted to identify confounding factors when comparing network performance. Scientists gradually 'modernize' the architecture to build decentralized visionary Transformers (e.g. Swin Transformers), defining how decisions in Transformers affect ConvNets performance. That was the foundation for the genesis of ConvNeXt. ConvNeXt is improved from the standard ResNet model, inspired by Swin Transformer [Liu et al., 2021b] (Figure 2) with an accuracy of 87.7%@1 on the ImageNet dataset, which is considered the strongest development today in the field of data classification (Figure 3)

2.4 Neural Network Embeddings

In recent years, neural network applications have increased dramatically, from natural language processing to image segmentation. Embeddings, a method operated to express discrete variables as continuous vectors, is one notable successful usage of deep learning models. This approach has found practical uses such as Embeddings for machine translation[Jansen, 2017] and Entity Embeddings of Categorical Variables [Guo et al., 2016].

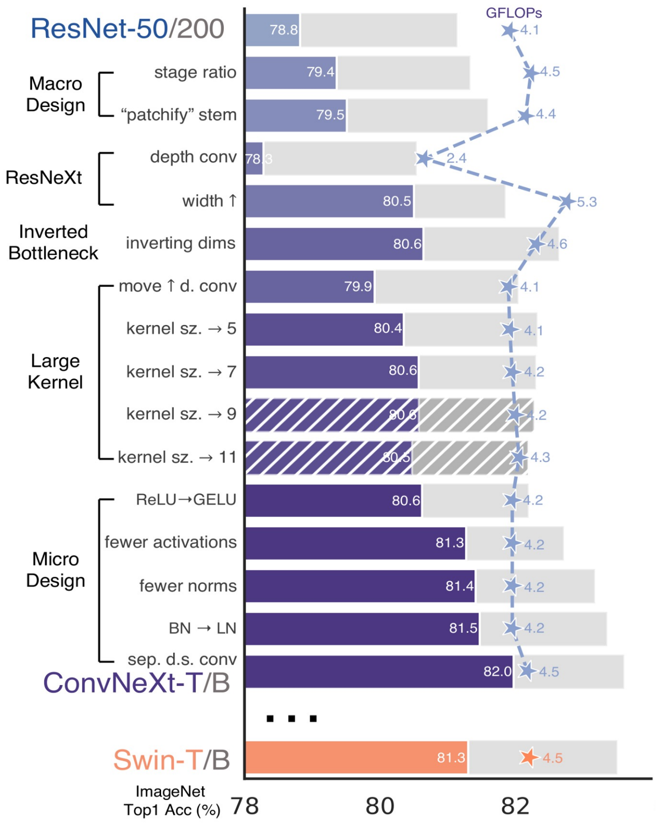


Figure 3: A block in ConvNeXt is improved from Resnet and Swin Transformer [Liu et al., 2022]

The transfer of a discrete variable to a vector of continuous numbers is known as embedding. Embeddings are low-dimensional continuous vector representations of discrete variables that are learned in the context of neural networks. Neural network embeddings are beneficial as they could diminish the size of category variables while still representing them meaningfully in the transformed space.

Neural network embeddings enclose 3 main purposes:

- Find the nearest neighbors (k-nearest neighbors) in the embedded space;
- Serve as input to the Machine Learning model for the supervised task;
- Visualize themes and relationships between categories.

Noticing the advantages of Neural network embeddings, this study arranged to embed the entire dataset collected through pre-trained neural networks. The benefit of this is that the embedded plant images become vectors of the same dimension and are easily classified by basic Machine Learning algorithms. Thanks to that, when new plants are added to the dataset, the proposed model does not require retraining from the beginning, but only embeds that plant type and assigns a label. Thus, the proposed approach has added the ability to identify a new plant species to the model while still saving resources.

3 Materials and Methods

3.1 Plant Classification Model using Resnet-152 Architecture

In the first model, this research built a model based on the Resnet-152 architecture that has been trained on the ImageNet dataset. In this test, the most basic classification model is built based on the Resnet-152 model, removing the 1,000 class classification layer in the Fully Connected section and replacing it with the 489 species classifier layer based on the dataset. The output of the model will be a vector of size 489. After going through the softmax function, it will reveal the probability that the image belongs to the corresponding 489 species. The loss function used is the Cross-Entropy function of the model when tested by the authors on the ImageNet set.

The training process was performed with the optimization algorithm SGD to optimize each data sample. Several results of plant identification of the model that was built based on ResNet architecture are indicated in Table 1.

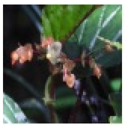
					
Ground Truth	Begonia eberhardtii Gagnep	Dicra-nopteris linearis	Heritiera littoralis Dryand	Neocin-namomum lecomtei H. Liu	Chro-molaena odor-ata (L.) R.M.King & H.Rob
Our ResNet model	Begonia eberhardtii Gagnep	Dicra-nopteris linearis	Heritiera littoralis Dryand	**Croton cascaril-loides Raeusch**	**Memecy-lon aff. ambrense Jacq.-Fél**

Table 1: Several results of plant identification of the model that was built based on ResNet architecture

The results demonstrate that the top 1 accuracy is about 59% and the top 5 accuracy is about 76%.

Despite the Resnet model having achieved impressive results on the Plant CLef dataset [ImageCLEF, 2022], the use of ResNet has not yielded satisfactory results on the DanangForestPlant dataset. One possible reason is that the better quality and quantity of

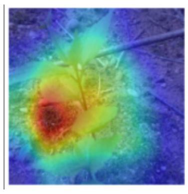
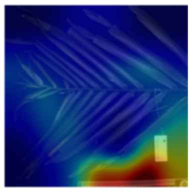
		
Chromolaena odorata (L.) R.M.King & H.Rob	Pothos repens (Lour.) Druce	Rhododendron simsii Planch

Table 2: Gradcam shows that the model remains noisy due to the background

data of the PlantClef set should lead to different model performance. An analysis based on the GradCAM technique [Selvaraju et al., 2017]shows that in some cases, the model has not focused on important details of the image. Specifically, in Table 2, the model also focuses on the ground area, leading to false predictions. This error can be remedied by manual data processing to remove the effect of unnecessary detail or by improving the focusability of the model. In the next section, the study details how to improve this model.

3.2 ConvNeXt Model

Recently, research [Liu et al., 2022] by Facebook AI Research about the development of ConvNeXt - an improved model from the standard ResNet model inspired by Swin Transformer has drawn concentration. With an accuracy of 87.7%@1 on the ImageNet dataset, it is considered to be the model with the best accuracy in the existing image classification models. Inspired by that research, this study utilized ConvNeXt as an enhancement to the ResNet network in our original model.

ConvNeXt model is built based on the Transfer Learning technique of the ConvNeXt Large model and the weights that have been trained on the ImageNet dataset. This model also changes the architecture at the Fully Connected layer to classify 489 forest plant species. Several results of plant identification of the model that was built based on the ConvNeXt architecture are indicated in Table 3.

Gradcam analysis of the images predicted by the new model shows that the model has overcome the disadvantages of the first model, it focuses on the details in the plant image to give a more accurate prediction (Table 4). The top 1 accuracy is around 77% and the top 5 accuracy is around 89%.

In the world in general or in Vietnam in particular, countless new plant species are discovered every month and every year. An issue arises when a plant that has just been discovered and put into the management system, has to be included in the plant classification system. Whether it ought to be added to the dataset or start the training from scratch, this leads to a costly, laborious, and time-consuming process since training complex deep learning networks takes a great deal of time. To solve that problem, this study proposed a new model PlantKViT to classify forest plant species even when they have just been added to the dataset.

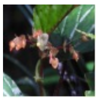
					
Ground Truth	Begonia eberhardtii Gagnep	Dicranopteris linearis	Heritiera littoralis Dryand	Neocinnamomum lecomtei H. Liu	Chromolaena odorata (L.) R.M.King & H.Rob
Our ConvNeXt model	Begonia eberhardtii Gagnep	Dicranopteris linearis	Heritiera littoralis Dryand	Neocinnamomum lecomtei H. Liu	Chromolaena odorata (L.) R.M.King & H.Rob

Table 3: Several results of plant identification of the model that was built based on the ConvNeXt architecture

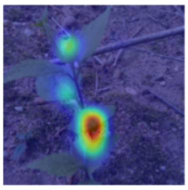
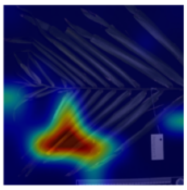
		
Chromolaena odorata (L.) R.M.King & H.Rob	Pothos repens (Lour.) Druce	Rhododendron simsii Planch

Table 4: Gradcam shows that the model focuses on important details

3.3 The Proposed PlantKViT Model

As mentioned in Section 2, the successful use of Transformer architecture in NLP has inspired scientific studies to apply Transformer architecture to image processing. Recently, a new Transformer architecture - Vision Transformer [Dosovitskiy et al., 2021] introduced by Dubovitskiy et al. has attracted the attention of this research. The proposed model uses Vision Transformer as an embedding network to embed the plant image and capture its feature vector. Feature vectors will be labeled and stored in the database. With a database of feature vectors representing plant species, it is straightforward to identify the learned plant species by comparing the feature vector of the image to be recognized with other vectors in the dataset using the KNN algorithm.

It is acknowledged that KNN may face a limitation when handling a large dataset. This is because the classification problem normally uses a dataset that changes over time. And after the feature extraction is completed, it will proceed to call a classification algorithm, such as multi-SVM. However, due to the nature of the context of plant image classification in the study area, new data over time does not increase considerably. Therefore, to avoid the problem of retraining on the entire dataset, the research team proposes KNN instead of the classification algorithm. In addition, using KNN also helps to add new species simply and quickly.

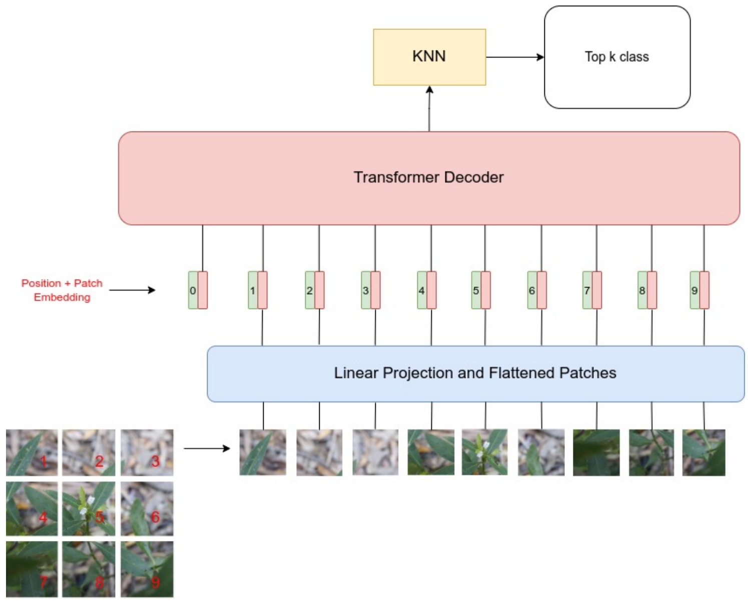


Figure 4: PlantKViT model architecture

3.3.1 Vision Transformer

This study uses the Vision Transformer model as an Embedding Network by removing the Head part, then the output of this process is a 1024 dimension vector, which is an attribute vector for the image.

With CNNs for image classification, the input is the entire image with a fixed size, however, Vision Transformer (ViT) has different processing. ViT processes the image by dividing it into equal-sized parts (patches) like each Token in NLP. For example, with the image displayed in Figure 4, the original image size is 480x480 pixels, and ViT transmits it into 9 patches of 160x160 pixels size. Thereafter the vector embeddings of 489 forest plant species were collected and labeled as corresponding species, then, stored in the database.

3.3.2 Proposed Model for Learning New Plant Species

With normal classification problems, after extracting image features, algorithms like Multi-SVM are applied to classify. However, for forest plant identification, there is a limitation that new plant species are discovered every month. This creates a barrier when training multiple times on a dataset with a small change, which is laborious and time-consuming. To solve this concern, the KNN algorithm is applied, as detailed in Figure 5. If there is a new plant species, its image is embedded through Vision Transformer to obtain the feature vector, label the vector and save it in the database. Therefore, the model has learned a new plant species. When there is another image of that new species, the model only needs to use KNN to match the embedding vectors available in the database to produce predictions. This makes the process of adding new plant species easier and minimizes the number of training times of the model.

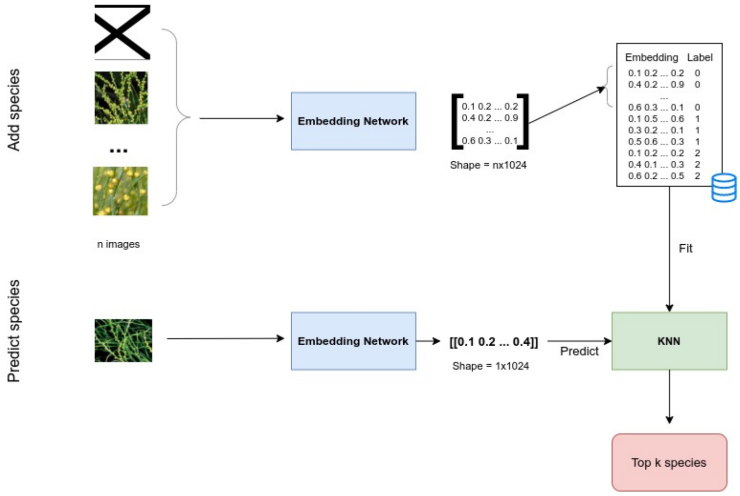


Figure 5: Proposed model for learning new plant species

4 Experiments and Results

4.1 Image Acquisition

The collected dataset contains 489 different plant species of project 36/HDKHCN/2020, which were manually collected from 4 typical research areas: Ba Na Nui Chua, Son Tra, Ngu Hanh Son, and Nam Hai Van in Da Nang, Vietnam. The manual collection process has occurred over a year by botanist experts from the Vietnam Academy of Forest Science, Vietnam National Forestry University, and Thu Dau Mot University under the project 'Research on building an intelligent management system of flora in Da Nang'. The dataset also represents the botanical diversity in the city, and several samples are presented in Table 5.

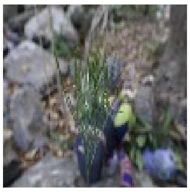
Psilotum nudum (L.) P. Beauv.			
Blechnum orientale L			
Pyrrosia lingua (Thunb.) Farw			
Dacrydium elatum (Roxb.) Wall. ex Hook			

Table 5: Samples of plant images in the collected dataset

Since the collected dataset contains many noisy and poor-quality images, they are manually filtered out. Duplicated images for the same plant species are also removed.

After cleaning, the dataset included several layers with insufficient data, mostly focused on one layer with roughly 8 images. Such a limited number of images cannot guarantee the training efficiency of Machine Learning models. Therefore, other data sources were used to supplement the number of images, including Google Images and the PlantNet website. The dataset after being extended has an improvement in image quantity and quality. The average number of images in a class is approximately 20 images and

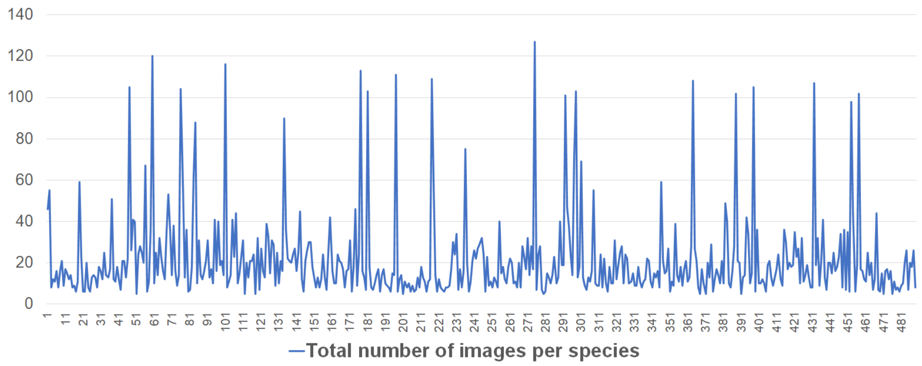


Figure 6: Distribution of the total number of images per species in the final dataset

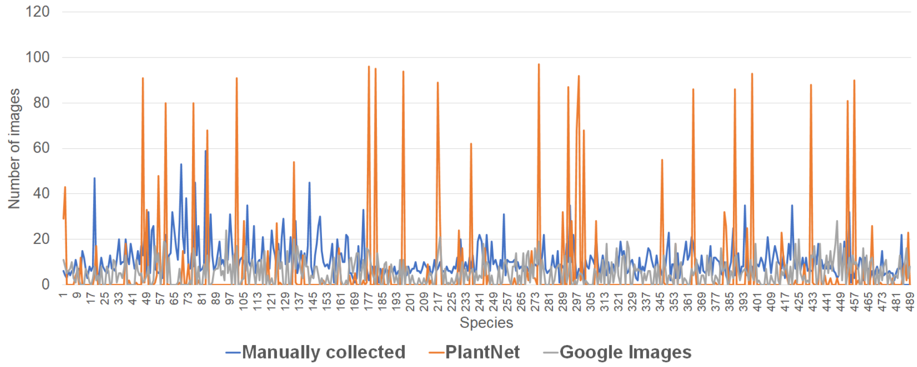


Figure 7: Distribution of the number of images per species from 3 different images sources in the final dataset

there are 489 species with 5 or more images (Table 6). Therefore, during the training and evaluation process, only 489 species with 5 images or more will be performed.

Figure 6 illustrates the distribution of the total number of images per species in the final dataset, and their distribution in 3 different data sources is also depicted in Figure 7. It can be seen that there are around 21 images on average in each species. And each of the 489 species has 5 to 127 images. From Table 6 and Figure 8, it is observed that nearly half of the images are manually collected from the actual location, and images from PlantNet account for more than one-fourth of the dataset. The proportion of images taken from Google Images is the least, with 24.54%. The final dataset is downloadable on request at https://bit.ly/PlantKViT.

4.2 Image Pre-processing

Before the image is fed into the model, image preprocessing is performed to optimize the predictive model, with the following steps.

- Crop to the center of images for better feature extraction;

Image Source	**Manually collected**	**PlantNet**	**Google Images**	**Total**
Number of Images	5136	2808	2583	10527
Percentage	48.79%	26.67%	24.54%	100.00%
Mean	10.50	5.74	5.28	21.53
Median	8	0	4	16
Standard Deviation	7.65	18.28	5.43	20.66
Min	0	0	0	5
Max	59	97	28	127

Table 6: Descriptive statistics of the training dataset

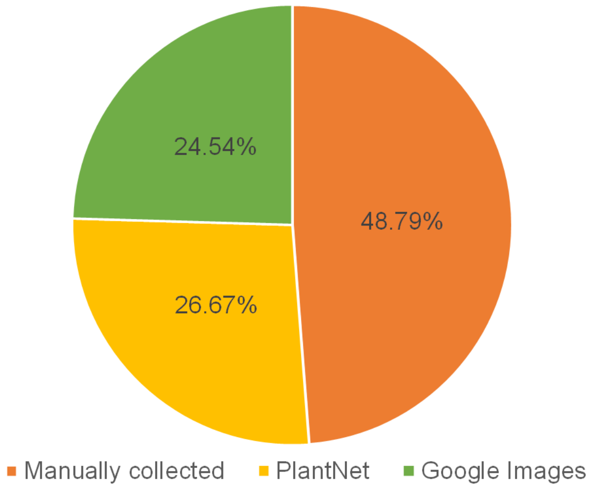


Figure 8: Percentage of images collected from 3 different sources

- Resize images to a fixed size (224, 224);
- Normalize by parameters: mean = [0.485, 0.456, 0.406], and std =[0.229, 0.224, 0.225] for 3 color channels;
- Data augmentation: Create more new plant images by randomly rotating the image from 30 to -30 degrees, and flipping the image horizontally.

4.3 Splitting the training and testing dataset

The data is split with 80% for the training set and 20% for the testing set. The testing and training sets have the same distribution of the number of species. Image data is stored in folders named by the name of the species. The splitting instruction for the training and testing sets is stored in a CSV file (Figure 9). Species names are encoded as integers for straightforward training and will be decoded to species' names after prediction.

label	image
92	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/186 _ 71465 _ Tam thụ hùng gaudichaud _ Trigonostemon gaudichaudii (Baill.) Müll.Arg./DSCN9323.JPG
307	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/390 _ 72322 _ Quắn hoa to _ Helicia robusta (Roxb.) R.Br. ex Blume/_DSC8778.JPG
45	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/140 _ 71268 _ Cỏ đầu rìu hoa chùy _ Floscopa scandens Lour./google_0009.jpg
129	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/22 _ 70764 _ Ráng yểm dực thay đổi _ Tectaria variabilis Tardieu & Ching/DSCN7272.JPG
368	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/46 _ 70867 _ Nóng roxburghi _ Saurauia roxburghii Wall./_DSC9464.JPG
86	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/180 _ 71441 _ Ba soi _ Mallotus paniculatus (Lamk.) Muell. - Arg./12.JPG
263	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/346 _ 72137 _ Cách hoa sp _ Cleistanthus sp1./_DSC2288.JPG
406	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/507 _ 72840 _ Máu chó trái dày _ Knema pachycarpa Wilde/IMG_3843.JPG
300	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/384 _ 72296 _ Thiên lý hương _ Embelia parviflora Wall. ex A. DC./DSCN1901.JPG
153	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/242 _ 71700 _ Dạ hợp nhỏ _ Magnolia coco (Lour.) DC./google_0019.jpg
107	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/20 _ 70756 _ Ráng chân xỉ lược _ Pteris semipinnata L./6 Ráng seo gà nửa lông chim (1).JPG
425	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/529 _ 72936 _ Thông tre _ Podocarpus neriifolius D.Don)/4a903f9dc3d4f3333cc02c81a38cf5465c3c8b19.jpg
469	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/81 _ 71013 _ Rau kem _ Ceropegia sp. nov./_DSC9763.JPG
148	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/238 _ 71682 _ Sắn dây _ Pueraria montana (Lour.) Merr./aad24783b40cd8978d4c1428c5fc1cd8dd8a0a98.jpg
80	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/175 _ 71421 _ Mã rạng ba thùy nhỏ _ Macaranga trichocarpa (Reichb. f. & Zoll.) Muell.-Arg. var. trilobulata (Muell.-Arg.) Gagnep./google_0006.jpg
15	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/112 _ 71144 _ Cứt lợn _ Ageratum conyzoides (L.) L./google_0017.jpg
292	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/377 _ 72268 _ Trọng đũa tuyến _ Ardisia crenata Sims./_DSC3931.JPG
368	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/46 _ 70867 _ Nóng roxburghi _ Saurauia roxburghii Wall./_DSC9468.JPG
465	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/78 _ 71000 _ Rau má _ Centella asiatica (L.) Urb./b10055a7dc78882b85129b16bdeb221bdc282531.jpg
204	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/291 _ 71908 _ Nắp ấm hoa dồi _ Nepenthes mirabilis (Lour.) Druce/04177454ecdf32aa88ac3325c7838852a0c3452f.jpg
477	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/89 _ 71046 _ Giom nam bộ _ Melodinus cochinchinensis (Lour.) Merr./_DSC1485.JPG
344	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/428 _ 72491 _ Đuôi chuột _ Stachytarpheta jamaicensis (L.) Vahl/bc7cdefa16fb5cfcdf5ee887597f03409939fbce.jpg
437	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/543 _ 72993 _ Găng cao _ Rothmannia eucodon (K.Schum.) Bremek./IMG_6573.JPG
64	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/16 _ 70739 _ Hỏa mạc thon _ Pyrrosia lanceolata (L.) Farw./86f129416d4bd0001ba4ac298d0c3efc4be7954b.jpg
227	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/312 _ 71997 _ Hài đài cuốn _ Paphiopedilum appletonianum (Gower) Rolfe/DSCN8953.JPG
368	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/46 _ 70867 _ Nóng roxburghi _ Saurauia roxburghii Wall./Nóng số 4.png
65	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/160 _ 71360 _ Đỗ quyên hoa đỏ _ Rhododendron simsii Planch./_DSC8839.JPG
204	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/291 _ 71908 _ Nắp ấm hoa dồi _ Nepenthes mirabilis (Lour.) Druce/45eaadcf742bb9fac12155efff9b5ed2f59e46ae.jpg
232	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/317 _ 72017 _ Bàn long sâm _ Spiranthes sinensis (Pers.) Ames./_DSC1531.JPG
116	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/208 _ 71561 _ Quế rừng _ Cinnamomum aff. iners Reinw. ex Blume/_DSC8739.JPG
285	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/370 _ 72238 _ Thồm lồm _ Polygonum chinense L./google_0005.jpg
271	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/354 _ 72169 _ Diệp hạ châu đỏ _ Phyllanthus ruber (Lour.) Spreng./_DSC7746.JPG
54	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/150 _ 71314 _ Dây pọp nhỏ _ Zehneria maysorensis (Wight & Arn.) Arn./_DSC9743.JPG
349	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/432 _ 72510 _ Hương bài _ Dianella ensifolia (L.) DC./9863a8c5f40d2b735f172b82fc6eb127fdc5cd9b.jpg
214	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/300 _ 71949 _ Lan tổ yến ấn _ Acripsis indica Wight/DSCN6765.JPG
80	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/175 _ 71421 _ Mã rạng ba thùy nhỏ _ Macaranga trichocarpa (Reichb. f. & Zoll.) Muell.-Arg. var. trilobulata (Muell.-Arg.) Gagnep./_DSC8955.JPG
383	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/482 _ 72731 _ Bão táp _ Scaevola taccada (Gaertn.) Roxb/136bdc9070a6228eea514918af59ff03d5d1682c.jpg
462	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/74 _ 70983 _ Nhọc lá nhẵn _ Polyalthia sp2./_DSC9075.JPG
291	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/376 _ 72264 _ Cơm nguội mộc _ Ardisia attenuata Wall. ex A.DC./_DSC5396.JPG
379	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/475 _ 72701 _ Giẻ trung quốc _ Desmos chinensis Lour./google_0009.jpg
226	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/311 _ 71993 _ Lan móng rùa _ Oberonia sp./google_0023.jpg
398	/content/drive/MyDrive/plantid/datasets/RawDaNangV3-HandCraftClean/498 _ 72791 _ Dẻ gai lá nhọn _ Castanopsis acuminatissima (Blume) A.DC/IMG_4470.JPG

Figure 9: Inside the CSV file of the testing set

4.4 Configuration

In this study, the dataset used for implementation contains 489 plant species collected in Da Nang (10,527 images, 50Gb). All 3 of our models are trained on CPU configuration - Intel Xeon Processor and GPU - Tesla K80. The parameters when training the 3 models are also similar.

- Size of input and output images: (224, 224);
- Initial learning rate: 0.001;
- Momentum: 0.9.

In the experiment, the learning rate is varied and monitored during training to obtain the optimum value. When performed with 100 loops, an early stopping mechanism is also implemented to stop the training when the validation accuracy is not improving within 3 consecutive epochs. By executing this, the model can obtain the optimal number of epochs while saving time during the training and validation process.

4.5 Results and Discussion

After training, several plant identification results of our proposed model are indicated in Table 7. It can be seen from Table 9 that our proposed model architecture classifies plants significantly better, compared to Resnet and ConvneXt models while using the same dataset. In particular, the F1 score, which measures the test's accuracy, of the proposed model is significantly higher at 0.83, compared to 0.77 and 0.56 of ConvneXt and Resnet models respectively. Similar results were also attained from the proposed model with precision and recall (Top 1 and Top 5 Accuracy). Notably, the recall (Top 5 Accuracy) of the proposed PlantKViT model reached 0.93 which is 122.4% when testing on the Resnet model. Moreover, the proposed model can learn new plant data, as a prominent advantage compared to Resnet and ConvneXt models, which is the key to applying the proposed model to practice.

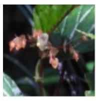
					
Ground Truth	Begonia eberhardtii Gagnep	Dicra-nopteris linearis	Heritiera littoralis Dryand	Neocin-namomum lecomtei H. Liu	Chro-molaena odor-ata (L.) R.M.King & H.Rob
Our proposed model	Begonia eberhardtii Gagnep	Dicra-nopteris linearis	Heritiera littoralis Dryand	Neocin-namomum lecomtei H. Liu	Chro-molaena odor-ata (L.) R.M.King & H.Rob

Table 7: Several plant identification results of the proposed model

To make the results more comparable, this research also applied the K-fold cross-validation technique to our proposed model. To implement this, the data is split into 5 partitions of equal size (k=5). The next step involves fitting the model into four subsets and evaluating the test error by using the fitted model on the fifth subset. This process is repeated 5 times, with each subset utilized as the test set once, as indicated in Figure 10.

As can be seen in Table 8, the F1 score, precision, and recall when using the K-fold cross-validation technique are marginally comparable to those attained without cross-validation (seen in Table 9), which is 0.816, 0.838, and 0.833 respectively. This substantiates the assertion that the dataset possesses adequate magnitude to effectively train and assess outcomes.

4.6 Plant Image Identification Application

To deliver a convenient lookup for plant management supervisors and other users, the study has successfully developed the website http://danang.plantid.com.vn. Thanks to this, users can easily upload images and perform the plant classification by accessing http://danang.plantid.com.vn/predict. Moreover, the team has launched an Andriod application on CH Play called 'Danangplant' (https://play.google.com/store/apps/details?id=com.plan_app), as well as an iOS application on App Store named 'Plant Id' (https://apps.apple.com/vn/app/plant-id/id1628335447). Since the team targets Vietnamese users as the main customer segmentation, the applications currently remain in a Vietnamese version, and the authors are putting effort to upgrade the English version in the next release.

The systems have the following primary functions.

- Plant species information lookup: species name, family, phylum, and detailed description;

Figure 10: K-fold cross-validation technique with k = 5

Fold	**F1 (weighted average)**	**Precision (weighted average)**	**Recall (weighted average) Top 1 accuracy**	**Recall (weighted average) Top 5 accuracy**	**Execu-tion time (s)**
Fold 1	0.821	0.839	0.844	0.949	0.9682
Fold 2	0.814	0.840	0.830	0.940	0.3188
Fold 3	0.801	0.821	0.819	0.945	0.3192
Fold 4	0.823	0.846	0.840	0.956	0.5779
Fold 5	0.820	0.844	0.831	0.940	0.3173
Average	0.816	0.838	0.833	0.946	0.5003

Table 8: Comparison of recall and F1 score using K-fold cross-validation technique with k = 5

- Search by distribution area (Son Tra, Ngu Hanh, Son, Ba Na, Nam Hai Van);
- Create a project of vegetation area: allows admin users to perform professional functions in the process of investigation, storing images, and coordinates of the study area;
- Plant identification: the PlantKViT model has been assessed and put into operation.

5 Conclusions and Future Work

In this study, a new dataset on Vietnamese plants has been collected and tested on newly built models inspired by several models that achieved outstanding identification results on the ImageNet dataset. Unlike common plant species collected in the Plant Clef dataset [ImageCLEF, 2022], the collected dataset contains rare forest plants with high economic value but in uncommon locations, representing unique biodiversity.

Despite substantial efforts in the past [Zhou et al., 2015, He et al., 2016, Szegedy et al., 2016], a proposed model was built on the Vision Transformer architecture combined

Model	**F1 (weighted average)**	**Precision (weighted average)**	**Recall (weighted average) Top 1 accuracy**	**Recall (weighted average) Top 5 accuracy**	**Execu-tion time (s)**
Our Resnet model	0.56	0.56	0.59	0.76	0.5670
Our ConvNeXt model	0.77	0.79	0.77	0.89	0.0209
Our proposed model (using K-fold)	0.82	0.84	0.83	0.95	0.5003
Our proposed model	0.83	0.86	0.84	0.93	0.0881

Table 9: Comparison of accuracy and execution time of the 3 models on plant image classification

with the K-nearest neighbors algorithm, which is more efficient than the conventional ResNet model commonly used in today's recognition problems. With impressive results with top@5 accuracy up to 93% and execution time of only 0.08s, the proposed model significantly outperformed Resnet and ConvneXt models. This is a promising result with further applications in practice. Therefore, the study also built a website and 2 mobile application systems so that people can look up plants conveniently on their gadgets, encouraging local people to improve their knowledge of plant taxonomy and interested parties to conduct plant species identification research. The future scope focuses on improving the performance of the plant image recognition system in the natural environment, by continuing to enrich the dataset and improve the recognition model. While the scope of this study can not only be applied to the diversity of flora in a city in Vietnam, further research can be steered to any location worldwide.

Acknowledgements

This study is funded and implemented for the project 'Research on building an intelligent management system of flora in Da Nang with number 36/HDKHCN/2020'. This work is supported by the People's Committee, Da Nang, and the University of Science and Technology, University of Danang.

References

[Aakif et al., 2015] A. Aakif and M. F. Khan (2015). Automatic classification of plants based on their leaves, Biosystems Engineering, vol. 139, pp. 66–75.

[Al-Qurran et al., 2018] Al-Qurran, R., Al-Ayyoub, M. and Shatnawi, A. (2018). Plant classification in the wild: a transfer learning approach. In 2018 International Arab Conference on Information Technology (ACIT), pp. 1-5.

[Alzubaidi, 2021] L. Alzubaidi et al. (2021). "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions". Journal of Big Data, 8(1), 1–74.

[Angelova et al., 2013] Angelova, A., Zhu, S., Lin, Y.: "Image segmentation for large-scale subcategory flower recognition". In: Applications of Computer Vision (WACV), 2013 IEEE Workshop on. pp. 39–45. IEEE (2013).

[Boudra et al., 2015] Boudra, S., Yahiaoui, I., Behloul, A.: "A comparison of multi-scale local binary pattern variants for bark image retrieval". In: Advanced Concepts for Intelligent Vision Systems. pp. 764–775. Springer (2015).

[Carranza-Rojas et al., 2016] Carranza-Rojas, J. and Mata-Montero, E. (2016). Combining leaf shape and texture for Costa Rican plant species identification. CLEI Electronic Journal, 19(1), pp.7-7.

[Chen et al., 2021] Chen, Y.G., Huang, J.H., Luo, R., Ge, H.Z., Wołowicz, A., Wawrzkiewicz, M., Gładysz-Płaska, A., Li, B., Yu, Q.X., Kołodyńska, D. and Lv, G.Y. (2021). Impacts of heavy metals and medicinal crops on ecological systems, environmental pollution, cultivation, and production processes in China. Ecotoxicology and Environmental Safety. No: 219. pp: 112336.

[Cogswell, 2016] Cogswell, M., Faruk Ahmed, Ross Girshick, (2016). "Reducing Overfitting in Deep Networks by Decorrelating Representations". International Conference on Learning Representations (ICLR).

[Cui, 2018] Cui, N. (2018). "Applying Gradient Descent in Convolutional Neural Networks". Journal of Physics: Conference Series, Vol. 1004(1).

[Dosovitskiy et al., 2021] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby (2021). An image is worth 16x16 words: transformers for image recognition at scale. International Conference on Learning Representations 2021.

[Dougherty et al., 2020] Dougherty, E.R. (2020). Digital image processing methods. CRC Press.

[Fiel et al., 2011] Fiel, S., Sablatnig, R.: "Automated identification of tree species from images of the bark, leaves and needles". In: Proc. of 16th Computer Vision Winter Workshop. pp. 1–6. Mitterberg, Austria (2011).

[Goeau et al. 2015] Goeau, H., Bonnet, P., Joly, A.: "Lifeclef plant identification task 2015". In: Working Notes of CLEF 2015 - Conference and Labs of the Evaluation forum, Toulouse, France, September 8-11, 2015. CEUR-WS (2015).

[Guo et al., 2016] Cheng Guo and Felix Berkhahn.2016. Entity Embeddings of Categorical Variables. arXiv preprint arXiv:1604.06737.

[He et al., 2016] He, K., Zhang, X., Ren, S., Sun, J. (2016). "Deep Residual Learning for Image Recognition". 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[Hien et al., 2020] Ngo Le Huy Hien, Thai Quang Tien, Nguyen Van Hieu (2020). Web Crawler: Design And Implementation For Extracting Article-Like Contents. Cybernetics and Physics. No: 9.3. pp: 144-151.

[Hien et al., 2021] Ngo Le Huy Hien, Luu Van Huy, Nguyen Van Hieu (2021). Artwork style transfer model using deep learning approach. Cybernetics And Physic. No: Vol. 10, Is.3.pp: 127-137.

[Hieu et al., 2020a] Nguyen Van Hieu, Ngo Le Huy Hien (2020). Recognition of Plant Species using Deep Convolutional Feature Extraction. International Journal on Emerging Technologies. No: 11(3).pp: 904-910.

[Hieu et al., 2020b] Nguyen Van Hieu, Ngo Le Huy Hien (2020). Automatic Plant Image Identification of Vietnamese species using Deep Learning Models. SSRG International Journal of Engineering Trends and Technology. No: 68.4. pp: 25-31.

[Howard et al., 2017] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv: 1704.04861.

[Huang et al., 2017] Huang, G., Liu, Z., Maaten, L. van der, Weinberger, K. Q. (2017). "Densely Connected Convolutional Networks". 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2017.243.

[ImageCLEF, 2022] ImageCLEF / LifeCLEF - Multimedia Retrieval in CLEF ImageCLEF (2022). PlantCLEF2022.

[Jansen, 2017] S. Jansen (2017). Word and phrase translation with word2vec.

[Jeon et al., 2017] Jeon, W.S. and Rhee, S.Y. (2017). Plant leaf recognition using a convolution neural network. International Journal of Fuzzy Logic and Intelligent Systems, 17(1), pp.26-34.

[Kumar et al., 2018] Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C., Soares, J.V.: "Leafsnap: A computer vision system for automatic plant species identification". In: Computer Vision–ECCV 2012, pp. 502–516. Springer (2012).

[Li et al., 2020] Z. Li et al. (2020). "A review of computer vision technologies for plant phenotyping". Computers and Electronics in Agriculture, 176, p.105672.

[Liu et al., 2021a] Liu, Z., Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo (2021). Swin transformer: Hierarchical vision transformer using shifted windows. International Conference on Computer Vision 2021.

[Liu et al., 2021b] Z. Liu et al., 'Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,' 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 9992-10002.

[Liu et al., 2022] Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T. and Xie, S., 2022. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11976-11986).

[Mattos et al., 2014] Mattos, A.B., Herrmann, R.G., Shigeno, K.K., Feris, R.S.: "Flower classification for a citizen science mobile app". In: Proceedings of International Conference on Multimedia Retrieval. p. 532. ACM (2014).

[Minh-Hoang et al., 2015] Minh-Hoang, N. (2021). Multifaceted Interactions between Urban Humans and Biodiversity-related Concepts: A Developing-country Data Set. Data Intelligence, 3(4), pp.578-605.

[Ramachandran et al., 2019] Prajit Ramachandran, Niki Parmar, Ashish Vaswani, Irwan Bello, Anselm Levskaya, and Jonathon Shlens. Stand-alone self-attention in vision models. NeurIPS, 2019.

[Rawat et al., 2015] Rawat, U.S. and Agarwal, N.K. (2015). Biodiversity: Concept, threats and conservation. Environment Conservation Journal, 16(3), pp.19-28.

[Russakovsky et al., 2015] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z. (2015). "ImageNet Large Scale Visual Recognition Challenge". International Journal of Computer Vision, 115(3), 211–252.

[Selvaraju et al., 2017] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. International Conference on Computer Vision 2017.

[Sulc et al., 2013] Sulc, M., Matas, J.: "Kernel-mapped histograms of multi-scale lbps for tree bark recognition". In: Image and Vision Computing New Zealand (IVCNZ), 2013 28th International Conference of. pp. 82–87. IEEE (2013).

[Sulc et al., 2014] Sulc, M., Matas, J.: "Texture-based leaf identification". In: Computer Vision-ECCV 2014 Workshops. pp. 185–200. Springer (2014).

[Sun et al., 2017] Sun, Y., Liu, Y., Wang, G. and Zhang, H. (2017). Deep learning for plant identification in natural environment. Computational intelligence and neuroscience, 2017.

[Szegedy et al., 2015] Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. (2015). "Going deeper with convolutions". 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[Szegedy et al., 2016] Szegedy, Christian; Vanhoucke, Vincent; Ioffe, Sergey; Shlens, Jon; Wojna, Zbigniew (2016). "Rethinking the Inception Architecture for Computer Vision".2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2818–2826.

[Tan et al., 2020] Mingxing Tan, Quoc V. Le (2020) . EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 36th International Conference on Machine Learning, PMLR 97:6105-6114.

[Xuet et al., 2021] Jing Xu, Yu Pan, Xinglin Pan, Steven Hoi, Zhang Yi, Zenglin Xu (2021). RegNet: Self-Regulated Network for Image Classification. IEEE Transactions on Neural Networks and Learning Systems.

[Zhao, 2017] Zhao, W. (2017). "Research on the Deep Learning of the Small Sample Data based on Transfer Learning". AIP Conference Proceedings, 1864.

[Zhou et al., 2015] Zhou, L., Qingwu Li, GuanyingHuo, Yan Zhou, (2017). "Image Classification Using Biomimetic Pattern Recognition with Convolutional Neural Networks Features". Computational Intelligence and Neuroscience.