CONTENTS

I Introduction 2
- I-A **The plant classification** 2
- I-B **Species vs Cultivar** 2
- I-C **Variety (cultivar) registers** 2
- I-D **The main problem of Plant Classification** 2
- I-E **Artificial Intelligence vs human** . . . 3
- I-F **Proposal** 3

II Related Works. 3
- II-A **Machine Learning Algorithms for Plant Classification.** 4
- II-B **Applying The CNN algorithms for Image Classification.** 4
- II-C **Deep Transfer Learning technique and Stack Method.** 4
- II-D **Image preprocessing.** 5

III Data. 5
- III-A **Importance of the Data.** 5
- III-B **Data collection sources.** 5
- III-C **Enrich DataSet with Data Augmentation.** 5

IV Algorithms. 5
- IV-A **Reasons for choosing CNN.** 6
- IV-B **What is CNN.** 6
  - IV-B1 Convolutional Layer 6
  - IV-B2 Pooling Layer 6
  - IV-B3 Activation Layer 6
  - IV-B4 Fully Connected Layer . . . 7
- IV-C **Proposal algorithms.** 7
  - IV-C1 Inception Network 7
  - IV-C2 Xception Network 7
  - IV-C3 Residual Network (ResNet) . 7
  - IV-C4 Inception Combine ResNet . 8
  - IV-C5 DenseNet: Densely Connected Convolutional Networks 8

V Method. 8
- V-A **Transfer Learning and Stacking Method.** 8
- V-B **Overfitting.** 8
- V-C **The Dropout method.** 9
- V-D **Final Batch Normalization layer.** . . . 9

VI Experiment - Result. 9
- VI-A **Benchmarking.** 9
- VI-B **Architectures** 9
- VI-C **Train/Validation/Test split.** 9
- VI-D **Image pre-processing.** 9
- VI-E **Algorithm training.** 9
- VI-F **Verification process.** 10
- VI-G **Quantitative analyses.** 10
- VI-H **Testing process.** 11
  - VI-H1 Experiment 1 11
  - VI-H2 Experiment 2 11
  - VI-H3 The majority voting-based classifier ensemble. 11
- VI-I **Result.** 11

VII Discussion - ANALYSIS OF EXPERIMENTAL RESULTS 13
- VII-A **Limitation.** 13
  - VII-A1 Limitation of value of dataset and quality of images 13
  - VII-A2 Multiple comparisons without corrections 13
  - VII-A3 A limitation of the voting ensemble 13
  - VII-A4 Power of machine 13
- VII-B **Future work.** 13
  - VII-B1 Tune the model 13
  - VII-B2 Image segmentation and image preprocessing 13

VIII Conclusion. 14

References 14

Appendix 16

***Abstract*—The current research project intends to study an innovative approach to plant classification by comparing four CNN models. The multiple CNN architecture models such as DenseNet-201, Xception, Inception-ResNet-V2, and EfficientNet-B7 were taught in two ways: beginning only by the name of the cultivar, then by the name of the cultivar after a preliminary determination of the genus. Later, it was assessed using images of flowers of two genus and 66 cultivars. All four CNN models were found to be capable to generate well an automated plant classification system for Genus from 89,1% for EfficientNet-B7 to 98,9% for Xception. Not all CNN models such as EfficientNet-B7 use their parameters with the same level of efficiency, which achieves 1.86% for Cultivar and 2.56% for Dahlia - two largest groups of classes and 20,16% for Helleborus classes with 89.1% for Genus class (7 and 2 classes in each group respectively). This result was unpredictably quite unclear.**

***Index Terms*—Cultivar CNN DenseNet-202 Xception Inception-ResNet-V2 EfficientNet Image Classification Plant Classification Flower Classification Majority voting-based method**

I. INTRODUCTION

A. *The plant classification*

Classification is a general scientific method used to methodize and systematise knowledge, which aims at organizing a set or collection of objects under study.

Classification of plants intends to rationalize knowledge about plants to scientifically understand their properties, kinship, characteristics and other features of a particular plant to make knowledge about them more attainable and observable. It allows to correctly identify the plant and its place in the classification, which in turn permit to make scientific research and exploit it for practical and functional purposes.This classification is based on the establishment of an evolutionary tree diagram in order to depict relationships of plants in evolution. Depending on the main properties, appearance, characteristic and features of a plant, it is feasible to determine which group (taxon) it belongs to.

Starting from the highest category, plants have generally been classified as follows: - Class - Subclass - Superorder - Order - Family - Subfamily - Tribe - Subtribe - Genus - Species - Variety - Form - Cultivar. Each group comprise the characteristics of a level above it but has some particularly unique elements and features. The further you go down the scale, the more inconsequential the differences become, at the end of the scale, the classification is left to only one plant.

The characteristic of newly discovered plants is compared to the characteristics to other already collected groups accordingly. Thus, classification is crucially important for the scientific study of plants.

B. *Species vs Cultivar*

The current research scrutinizes the classification of plants according to cultivar name. Species and variety (cultivar) are different concepts. Species - biological, natural formation, a result of natural variability. Variety or cultivar is a selection concept, formed as a result of artificial selection of particulars with the traits necessary for a researcher, which are fixed within the species by artificial means.

C. *Variety (cultivar) registers*

Every unique cultivar has a unique name within its denomination class (which is almost always the genus). Names of cultivars are registered with an International Cultivar Registration Authority (ICRA). The main purpose is to avoid duplication of variety names within a class (usually a genus) and to make sure that the nomenclature is congruent and consistent with the latest editions of the International Code of Nomenclature for Cultivated Plants (ICNCP). The ICRA system is voluntary, non-state but provides no legal protection, generally overall formed by institutions and societies specializing in particular plant genera such as Dahlia or Rhododendron and is currently located in Europe, America, and other countries (Spencer, 2007). New names and other relevant data are gathered and submitted to the ICRA. They note down details about the plant, such as parentage, highlight a basic description of its special, distinctive features and also the names of those who were involved with its development. Most ICRAs can be contacted electronically and with many of them maintaining web sites for an up-to-date listing (information) (Brickell, 2009). There are independent registration authorities for different plant types such as roses and camellias. In addition, cultivars may be associated with commercial marketing names referred to in the Cultivated Plant Code as” trade designations”.

D. *The main problem of Plant Classification*

Identification of plant species is usually done by Botanists and Plant researchers and scientists. This data is also needed by farmers, foresters, nature lovers, biologists, etc. There are countless plant species in the world. For example, there are now more than 57,000 registered cultivars of Dahlia and over three hundred species and tens of thousands of cultivars of Rosa (Bates, 2015), which are officially registered through the Royal Horticultural Society (RHS) (RHS, 2015). Furthermore, a cultigen - a plant whose origin or selection is primarily due to intentional human activity – are being developed and patented every day. The official register control lists are updated annually.

As the field of cultivar is quite large and has a large number of plants in it, hence it becomes very difficult and laborious to properly identify a particular plant cultivar just by looking at it. The plant identification by conventional methods is difficult, time-consuming and frustrating for a non-professional. Besides, due to great resemblance between some plant species, it may be difficult to differentiate them easily even for a specialist. Thus, the theory-based method is found incomparable in its efficiency and accuracy with a well-trained machine. The conventional ways of classifying plants have been based on the visible physical characteristics of the plant. However, since the discovery of DNA, biologists have attempted to classify plants more accurately, and to group them according to the similarities of their DNA. The DNA fingerprinting works by juxtaposing the DNA profile of a sample to the database of known cultivar profiles. It is quite costly for a single sample. Hence, there is a need to create a computerized or automated system to recognize and classify the plants.

Fruits, flowers, leaves, stem, habit are characteristics which are available to a majority of plants. Its shape, colours, veins, and textures are the key features used to compose such an automated plant classification system. Since the scientific study of plant classifications quite often involves tasks that require storing large amounts of information in memory, which may not be feasible for humans, therefore computers are used for this task.

*E. **Artificial Intelligence vs human***

Plant recognition is a task that classifies plant species using photographed images. It is exceedingly difficult tasks rather than conventional image recognition because some species look closely similar to each other . Photographed plant images are strongly influenced by the angle at which the photo was taken and the direction in which the plant is grown differing even for plants of the same species. The color of the plant and flower also can be differed along with the seasonal changes.

Recently in computer science research, lots of new techniques like image processing, pattern recognition, and machine learning are widely used to increase human's ability to identify plant species. Image-processing Machine Learning (ML) techniques such as Artificial Neural network (ANN), Convolution Neural Network (CNN), Support Vector Machines (SVM), k-Nearest Neighbor (KNN), and others are used in order to perform pattern recognition and classification.

Deep learning, a subfield of Artificial Intelligence (AI), is a well-known and widely employed technique that has been used in miscellaneous fields including medical science, biology, speech recognition and computer vision. The deep learning methods have risen as a promising and auspicious alternative in plant recognition and it outperforms other handmade approaches in feature extraction. A usual deep learning system comprise of multiple linear and non-linear layers. The word "deep" comes from deeply stacking the layers. These methods are able to train the weights of networks on large datasets as well as to calibrate the weights of pre-trained networks on small datasets (Kumar Shukla et al., 2020).

Principally, the convolutional neural networks (CNN) obtained highest success lately in the field of image recognition and is widely implemented to the most of the modern deep learning techniques. CNN could impart vastly in managing agricultural and forestry production, gardening, and nurseries, due to its vast competency to accurately identify plant species in a much smaller time frame. Hence in order to solve the problem of Plant Classification, it was necessary to develop a solution based on Artificial Intelligence. Thus, a great demand for AI tools to identify a plant species exists in the present day.

*F. **Proposal***

The current research project intends to study an innovative approach to plant classification by comparing four CNN models such as DenseNet-201, Xception, Inception-ResNet-V2, and EfficientNet-B7. The multiple CNN architecture models were taught in two ways: beginning only by the name of the cultivar, then by the name of the cultivar after a preliminary determination of the genus. Later, it was assessed using images of flowers of two genus and 66 cultivars. All four CNN models were found to be capable to generate well an automated plant classification system for Genus from 89,1% for EfficientNet-B7 to 98,9% for Xception. The suggested model will help the professional and non-professional to correctly classify the plant by photographing the flower.

The main stages of this research are:

1) First, the dataset of different plant flower was collected from various sources: private photo, internet open sources images and extracted private videos. All these images were color images.
2) Second, images by data augmentation were added to reduce the imbalance of classes which gave outcomes result of 1183 images with 25 images in each class on average. After that the dataset was split on train/validation/test images by 80/15/5%.
3) Third, four CNN algorithms are trained and compared on the basis of accuracy and loss, using the transfer learning-based approaches to advance the performance of the network on a small dataset such as in this research.

The most important findings are:

1) the recognition accuracy does not increase if a number of classes slightly decreases;
2) there is a linear relationship between class complexity and accuracy;
3) not all CNN models use their parameters with the same level of efficiency.

Forth, relationships between these performance indices were analysed providing insights for:

1) understanding what solutions have been explored so far and what direction would it be appropriate to go in the future;
2) selecting the CNN architecture that better fits the resource constraints of practical deployments and applications.

Experimental results show that the flower features are well learned and can be used for classifying flower images effectively with further improvement.

II. Related Works.

A large number of research has been done to classify the Plant using the Artificial Intelligence (AI) techniques, from handcrafted Iris flower data to resourceful hi-tech mobile phones. Some of the research is described below. Some algorithms and methods were transferred and applied to the current research to make a comparison and to make an attempt at achieving the same results or higher; some other algorithms and methods are stored and considered for future work due to time and computational limitations.

A. *Machine Learning Algorithms for Plant Classification.*

The Iris flower data set or Fisher's Iris data set is a multivariate data set introduced by the British statistician and biologist Ronald Fisher in his 1936 paper 'The use of multiple measurements in taxonomic problems as an example of linear discriminant analysis'(Dua and Graff, 2017). This data set consists of 3 different types of irises' (Setosa, Versicolour, and Virginica) petal and sepal length, stored in a 150x4 numpy.ndarray. The rows being the samples and the columns being: Sepal Length, Sepal Width, Petal Length and Petal Width. Different Machine Learning Algorithms were training, testing and improving on that dataset.

B. *Applying The CNN algorithms for Image Classification.*

Great success came to visual imagery analysis with the development of Deep Learning and, particularly, Convolutional Neural Network.

Tremendous work (Bianco et al., 2018) was accomplished in-depth analysis of the majority of the deep neural networks (DNNs). The diversity and complexity of existing algorithms are shown in a Fig.* Recognition accuracy, model complexity, computational complexity, memory usage, and inference time was observed for each DNN. This study is useful as for researchers as for practitioners, it helps to select the DNN architecture(s) and guides in research directions.

One of the popular new algorithms from Deep Convolutional Neural Network (CNN) family is DenseNet. In (Alipour et al., 2021), the transfer learning approach was used employing DenseNet121 architecture to categorize various species of oxford-102 flowers dataset. Classifying 102 flower species 98.6% accuracy was achieved the highest amidst the other employed methods.

A new image classification algorithm based on the saliency map was proposed in (Lv et al., 2021), which integrates the saliency map into the image feature extraction process, thus avoiding image segmentation and enhancing the adaptability and reliability of the algorithm. The original image is processed using the visual saliency map model based on the spectral residual method. In the experiment, it was observed that this algorithm was more accurate in the extraction of the salient regions of flowers for different types of flowers in a complex background, which provides a basis for the classification of migration learning later. But this method was unable to classify the cultivar name of a particular plant, when the shape of the flower was identical. Experiments on the Oxford flower-102 data set showed that the model had better classification results than traditional methods and other deep neural network architectures with a classification accuracy of 91.9%, which verifies the accuracy of this method for flower image classification tasks.

In (Abai and Rajmalwar, 2019) paper, an approach for classification models was presented, which was used for a subset of ImageNet dataset with 200 different classes called the Tiny ImageNet. No pre-trained network has been used; neither were any dense or fully connected layers, or dropout layers, while more than 500 training epochs were used. New architecture and some new techniques were applied to improve accuracy. Two custom DenseNet models were designed that were well suited for this challenge, which helped to achieve results with Top-1 accuracy of 59.5% and 62.7% for the networks.

Scaling up ConvNets by using more layers is widely used to achieve better accuracy. However, the process of scaling up ConvNets has never been well understood and there are currently many ways to do it. The most common way is to scale up ConvNets by their depth(Tan and Le, 2019) or width. Another less common, but increasingly popular, method is to scale up models by image resolution. This empirical study shows that it is critical to balance all dimensions of network width/depth/resolution, and such balance can be achieved by scaling each of them with a constant ratio. Based on this observation, a simple yet effective compound scaling method was proposed. In particular, the proposed EfficientNet-B7 achieves state-of-the-art 84.3% top-1 accuracy on ImageNet, while being 8.4x smaller and 6.1x faster on inference than the best existing ConvNet. The EfficientNets also transfer well and achieve state-of-the-art accuracy on CIFAR-100 (91.7%), Flowers (98.8%), and 3 other transfer learning datasets, with an order of magnitude fewer parameters.

C. *Deep Transfer Learning technique and Stack Method.*

In (Guan, 2021) study, a new plant disease detection approach by combining four CNN models has been developed. The experiment used an open-source database of 36258 images classified into 10 plant species and 61 classes of healthy and disease plant leaves. Four CNN models including Inception, Resnet, Inception Resnet, and Densenet were deployed and the results of the CNN models were processed by a stacking method. The use of the stacking method achieved an 87 accuracy rate, which is a significant improvement compared to the result of using a single CNN model. The relatively high accuracy rate indicates that using a combination of CNN models by a stacking method could be an appropriate approach that can be extended to practical cultivation conditions as an advanced plant disease warning tool.

The DenseNet algorithm is employed for different image classification tasks, for example, medical. Transfer learning can train the weights of networks on large datasets as well as fine-tuning the weights of pretrained networks on small datasets. A DenseNet201 based deep transfer learning (DTL) was proposed (Jaiswal et al., 2021) to classify the patients as COVID infected or not i.e. COVID-19 () or COVID (). The proposed model is utilized to extract features by using its own learned weights on the ImageNet dataset along with a convolutional neural structure. Extensive experiments are performed to evaluate the performance of the proposed DTL model on COVID-19 chest CT scan images. Comparative analyses reveal that the proposed DTL based COVID-19 classification model outperforms the competitive approaches.

In (Lee and Chan Yoon, 2019) paper, a very inspiring model for plant image classification task was proposed by using a deep learning framework that combines a Machine Leaning model with Deep Learning model. A digital photo also has a lot of useful metadata which can be acquired through the electronic sensors, a smartphone is already equipped with. Exchangeable Image File format (EXIF) is the metadata of a digital photo. It contains useful information such as date and time, lens exposure, image resolution, and geographical position if possible. This metadata can help classify, not only plant species, but for other image recognition tasks. Shooting time, geographical location, and camera condition information is embedded within the image file as metadata. Such information may improve the performance of the image classification. The linear discriminant analysis (LDA) algorithm was incorporated to implement linear combinations between selected features and EXIF metadata. EXIF metadata is used as the input value of the wide model whereas image content is used for the deep convolutional neural net component. Within the framework, were combined each output linearly to predict the species of the input image. The accuracy was achieved the 78% maximum.

D. ***Image preprocessing.***

In (Sharma et al., 2020) approach, firstly, images were resized to smaller pixel size in order to speed up the computations. The acquired images contain some noise. This noise was removed using some filtering techniques like Gaussian Blur. After that images were present in RGB format which is not appropriate for further work as RGB format is unable to separate image intensity. Hence it was converted to another color space that is HSV which separates color from the intensity. Also, RGB color space is noisier than HSV. Secondly, the segmentation of images was done in order to separate the leaves from the background. The accuracy was achieved in range of 53.4% - 66.4% for Machine Learning algorithms and 98% for CNN that can prepossess the algorithm. According to Fei Hu[] as far as the flower images, the accuracy of image segmentation is strongly influenced by the background soil, weeds, and fallen leaves. Hence, the classification effect is overwhelmingly dependent on the accuracy of image segmentation. Thus, the current research was refused to use of image segmentation but this method inspiring to further investigation.

III. Data.

A. ***Importance of the Data.***

For the AI system to learn and generate a desired result; it must receive, process, and learn from a large quantity of data. Obviously, by feeding an AI with images of cats, you cannot gain as an output images of dogs. That is why it is crucial to select the quality dataset for a system, otherwise the output could yield a total, undesirably different result than expected, even if data is numerous. This requires targeted testing and sampling to ensure that the training exercises are being perfectly optimized. This type of work is time-consuming but essential. Therefore, proper Data is key for helping AI systems to learn effectively.

B. ***Data collection sources.***

There are no known open databases that would regulate not only the genus but also the plant variety. So, for this project, it was required to collect the data set. The copyright law ought to be taken into consideration, so the images were collected from open internet sources only, such as Wikipedia, and from small private collections of photographs. Additionally, the video of plants was taken from plant nurseries and converted into. jpg files with the Python program. The combined dataset was thoroughly revised and cleared from purport less images, non-contrasting or blurred.

C. ***Enrich DataSet with Data Augmentation.***

In view of the fact, that assembling a dataset is a very time-consuming procedure and with timing of the project being quite limited, the dataset only includes 2 genera Dahlia and Helleborus which includes 59 and 7 cultivars, respectively. The cultivar that consists of just 1-3 images was deleted as it could not be correctly split between training and validation. Fig. 1 shows a Classes counting.

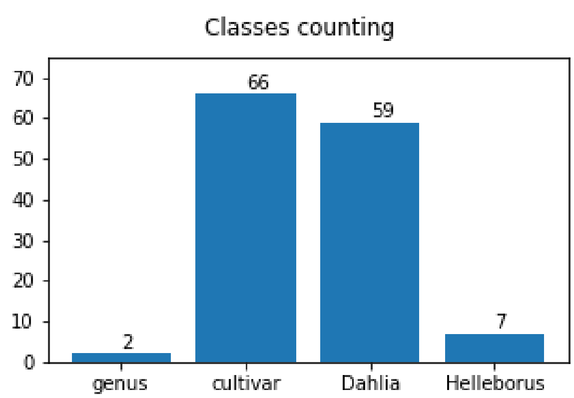


Fig. 1. Classes counting.

In order to increase the quantity of data, the Data Augmentation technique was implemented with the up-sampling method to avoid a severe issue with class imbalance. To fix the quantity of images in the classes that appear the most, images were copied from the remaining classes and processed with augmentation; later, they were added to the dataset. The process was repeated several times until the number of images in each class was equal. The images were randomly processed with a horizontal flip, up to 10% zoom and up to 15-degree rotation with black filling the edge pixels. Cultivar images have been added up to 25 on average for each class. Still for genus classification the imbalance remains as in Helleborus include only 7 classes versus 59 in Dahlia. As a result, the dataset contains a total of 1501 cultivar images. The sample of training images is shown in Appendix Fig.20. The Data Structure is shown in a Table I and in a Fig,2.

IV. Algorithms.

The selection of correct Data is as effective in achieving significant results in problem classification as the selection of

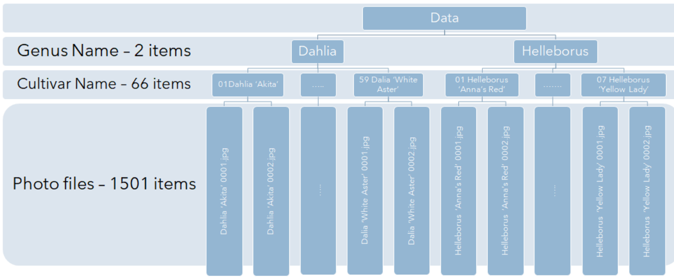


Fig. 2. Data Structure

TABLE I
DATA STRUCTURE

Genus	N	Cultivar Name
Dahlia	1	Dahlia 'Arabian Night'
	2	Dahlia 'Alauna Clair-obscur'
	3	Dahlia 'Akita'
		....
	57	Dahlia 'Totally Tangerine'
	58	Dahlia 'Teesbrooke Audrey'
	59	Dahlia 'White Aster'
Helleborus	1	Helleborus 'Anna_s Red'
	2	Helleborus niger (Christmas Rose)
	3	Helleborus 'Frostkiss Glenda_s Gloss'
	4	Helleborus 'Penny's Pink'
	5	Helleborus 'Molly's White'
	6	Helleborus 'Yellow Lady'
	7	Helleborus 'Walhelivor (Ivory Prince)'

a right algorithm both of which either improve or impair the results further. In this section, some background information about the CNN algorithms was given in order to better understand the contributions of this research.

A. *Reasons for choosing CNN.*

For plant classification, the features of plant such as leaves, flowers, or bark, their' textures, colour, or shape were mainly used. This information was thoroughly collected and loaded for image processing by Machine Learning (ML) techniques such as Support Vector Machines (SVM) or k-Nearest Neighbour (KNN). Later, for visual image analysis rapidly developing Deep learning methods were applied such as Artificial Neural network (ANN) and Convolution Neural Network (CNN), which outperform other hand-crafted methods in feature extraction. The latest (CNN) achieved the most success in the field of image recognition.

B. *What is CNN.*

CNN is a neural network which comprises of four layers namely Convolutional layer, Pooling layer, Activation function layer, and Fully connected layer as shown in Fig. 3.

A typical deep learning system consists of several linear and non-linear layers. The word "deep" comes from deeply stacking the layers. These techniques are able to train the weights of networks on large datasets as well as fine tuning the weights of pre-trained networks on small datasets (Kumar Shukla et al., 2020).

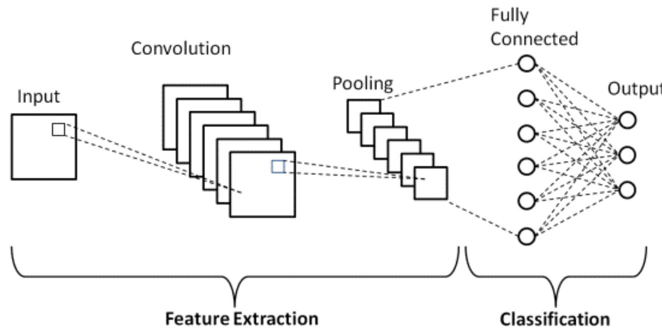


Fig. 3. General CNN architecture. Source: https://medium.com/techiepedia/binary-image-classifier-cnn-using-tensorflow-a3f5d6746697

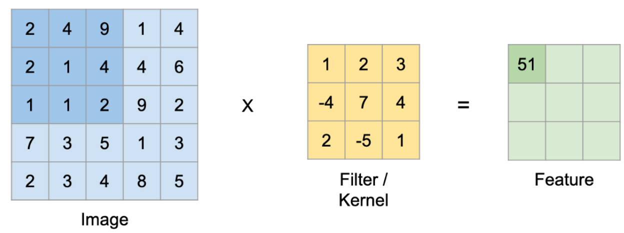


Fig. 4. Convolution operation step — 1. Source:https://towardsdatascience.com/convolution-neural-networks-a-beginners-guide-implementing-a-mnist-hand-written-digit-8aa60330d022.word

1) Convolutional Layer: It is the most important layer in the CNN model and is also responsible for the naming of this network. In this layer shown in Fig. 4 , some mathematical computations are performed to get the features of the image. It consists of filters that have a lower width and height than an input image and a depth is the same as an input image. If an image of size 64*64*3 is fed in the CNN model and we had a total of 10 filters then the output of this layer would have the dimension of 64*64*10. There are a total of 5 convolutional layers where the number of filters is 32,64,64,128,128 respectively. The kernel size is 3*3 for all layers.

2) Pooling Layer: This is the layer, shown in Fig. 5 that is mainly responsible for the output size reduction of the previous layer. Filters of different sizes can be used in this layer but 2*2 size is preferred in general. There are two major types of pooling layers that are used namely- max pooling and average pooling. As the name suggests max pooling takes the maximum value out of the filter and average pooling takes the average. In our model, we have used average pooling with a pool size of 3*3 and with stride being default which is equal to pool size.

3) Activation Layer: In any neural network, the activation layer plays an important role as it is responsible for the nonlinear learning of the network. There are different types of activation functions such as sigmoid, tanh, ReLU, and

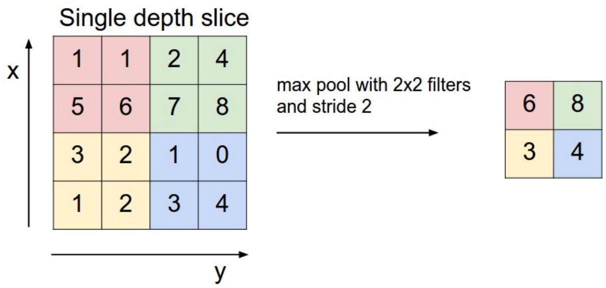


Fig. 5. Max Pooling with 2*2 filters and stride 2. Source: https://cs231n.github.io/convolutional-networks/.

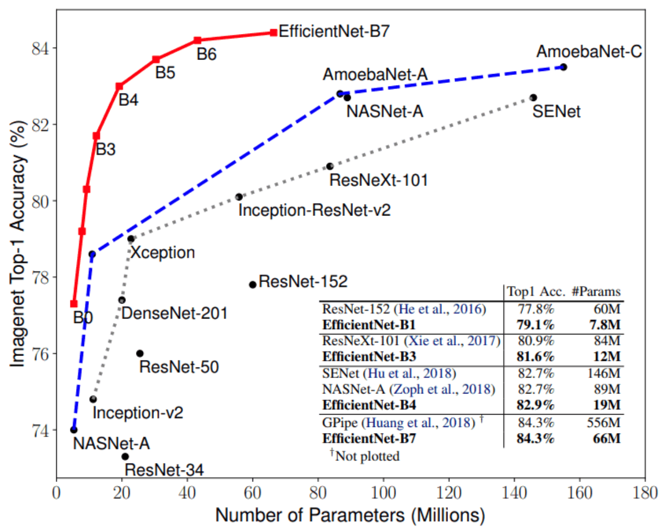


Fig. 6. Model Size vs. ImageNet Accuracy. Source: (Tan and Le, 2019).

LeakyReLU. In our model, we have used ReLU for all the layers except the output layer for which we have used softmax.

4) Fully Connected Layer: After performing all the computations in previous layers, the output is fed into a normal neural network for classification purposes. The model had 2 dense layers with 1024 and 19 units respectively.

C. Proposal algorithms.

The selection of algorithms was inspired by paper (Tan and Le, 2019). The figure 6 shows different models depending on the accuracy and number of parameters.

After the discussion DenseNet201, Xception, InceptionResNetV2, EfficientNetB7 algorithms were chosen.

1) Inception Network: An Inception module computes multiple different transformations over the same input map in parallel, concatenating their results into a single output. In other words, for each layer, Inception does a 5x5 convolutional transformation, and a 3x3, and a max-pool. And the next layer of the model gets to decide if (and how) to use each piece of information. The naive Inception module just tripled or quadrupled the number of filters. In order to solve the computational bottleneck, the authors of Inception used 1x1 convolutions to "filter" the depth of the outputs. The basic blocks of Inception Module shown in Fig, 7 and Fig. 8.

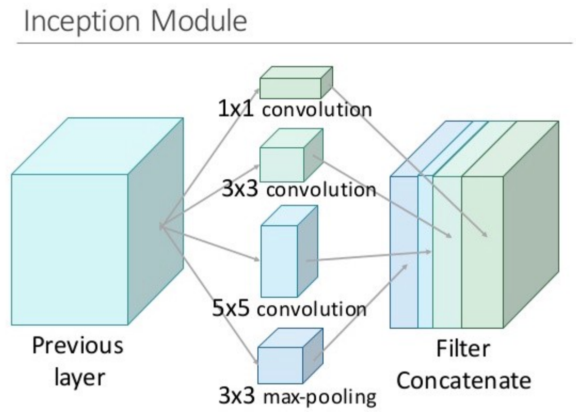
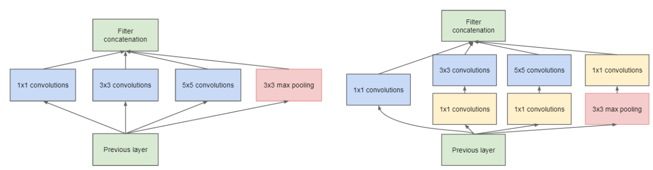


Fig. 7. Inception module. Source:https://towardsdatascience.com/an-intuitive-guide-to-deep-network-architectures-65fdc477db41#:~:text=Xception%20slightly%20outperforms%20Inception%20v3,implying%20a%20greater%20computational%20efficiency.



(a) Inception module, naïve version (b) Inception module with dimension reductions

Fig. 8. Inception module - naive and with dimension reductions. Source: https://towardsdatascience.com/an-intuitive-guide-to-deep-network-architectures-65fdc477db41#:~:text=Xception%20slightly%20outperforms%20Inception%20v3,implying%20a%20greater%20computational%20efficiency.

2) Xception Network: Xception stands for "extreme inception." Xception takes this one step further. Instead of partitioning input data into several compressed chunks, it maps the spatial correlations for each output channel separately, and then performs a 1x1 depthwise convolution to capture cross-channel correlation. Xception slightly outperforms Inception v3 on the ImageNet dataset, and vastly outperforms it on a larger image classification dataset with 17,000 classes. Most importantly, it has the same number of model parameters as Inception, implying a greater computational efficiency. The basic blocks of Xception Module shown in Fig, 9

3) Residual Network (ResNet): The authors of ResNet reduced these problems down to a single hypothesis: direct mappings are hard to learn. And they proposed a fix: instead of trying to learn an underlying mapping from x to H(x), learn the difference between the two, or the "residual." In general, ResNet gives layers a "reference" point — x — to start learning from. However, because the gradient signal in ResNets could travel back directly to early layers via shortcut connections, we could suddenly build 50-layer, 101-layer, 152-layer, and even (apparently) 1000+ layer nets that still performed well.

The ResNet structure is developed to solve the degradation problem. According to Kaiming He and his partners' paper, it

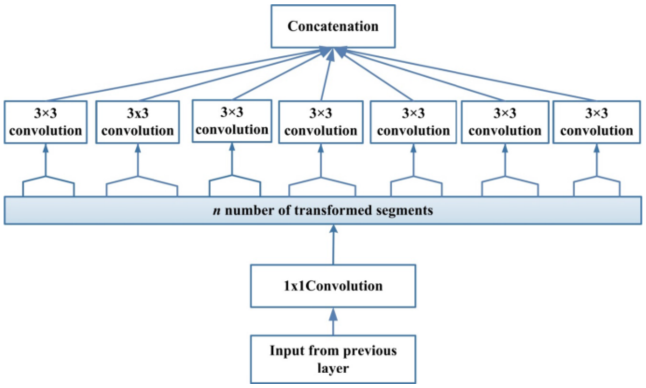


Fig. 9. Xception module. Source: https://link.springer.com/article/10.1007/s10462-020-09825-6

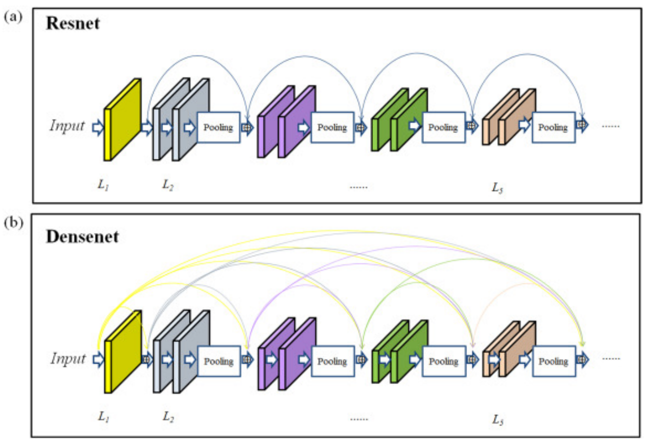


Fig. 10. DenseNet vs ResNet. Source: (Zhang et al., 2021)

points out that building a deeper neural network by stacking layers of identity mapping on the trained shallower network would produce a result which should be exactly the same as the trained shallower network since the layer stacked are identity map (He et al., 2016). In reality, because of the difficulty of training, an extremely deep network will cause degradation problem and the result is not as accurate as a relatively shallow network. However, the ResNet can solve this problem. The rationale of ResNet is to build multiple residual blocks and stack them together. Each of the residual blocks of ResNet consists of a series of layers and a shortcut connection that connects the input and output of the module. Then it performs an addition act between input and output. If the input and output sizes are different, then zero padding or projection (by 1×1 convolution) can be used to get the matching size (He et al., 2016).

4) Inception Combine ResNet: Inspired by the residual network, the idea of residual blocks is applied in the Inception Network in which each residual block comprises a series of layers of inception module and a shortcut connection between the input and output (Szegedy et al., 2016).

5) DenseNet: Densely Connected Convolutional Networks: the main difference with ResNets is that we will concatenate

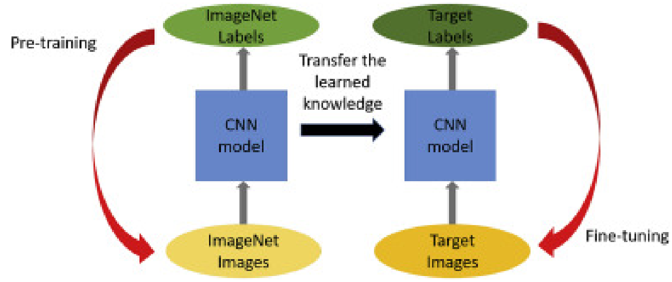


Fig. 11. The main steps of Transfer learning for classification. Source: (Zhu et al., 2021)

instead of adding the feature maps. The underlying logic of DenseNet is consistent with ResNet, but it establishes a solid connection between all the front and back layers. As a result, it requires fewer parameters than other CNNs, as there are no repeated feature maps. It makes it possible for feature reuse by different layers which increases variation in the subsequent layer input and improves the performance (Huang et al., 2017).

V. Method.

A. *Transfer Learning and Stacking Method.*

A CNN models, which are built on Deep Transfer Learning (DTL), were proposed to classify the plant's cultivar name. Those models were utilized to extract features by using its own learned weights on the ImageNet dataset along with a convolutional neural structure. Transfer learning (Pan and Yang, 2009) is about leveraging feature representations from a pre-trained model, so there would be no need to train a new model from scratch (see Fig. 11). The pre-trained models are usually trained on massive datasets such as the ImageNet dataset, which holds over 1 million images and over 1000 classes. These models can be applied directly in making predictions on new tasks or integrated into the process of training a new model, wherein, the weights obtained from the models can be reused. At the same time, due to the small number of images used in this research and the excessive cost of collection, the use of migration learning methods can solve the problem of insufficient image data.

B. *Overfitting.*

The over-fitting phenomenon in convolutional neural networks is an important problem. Overfitting is a concept in data science, which occurs when a statistical model fits perfectly against its training data. When this happens, the algorithm unfortunately cannot perform accurately against unseen data, defeating its purpose. Overfitting (Hawkins, 2004) happens when a model learns the detail and noise in the training data to such an extent that it negatively impacts the performance of the model on new data. This therefore means that the noise or random fluctuations in the training data is picked up and learned as concepts by the model. Overfitting is avoidable by early stopping, training with more data, data augmentation, feature selection, regularization, or ensemble methods. This paper uses the Dropout method, the transfer learning method, the Final Batch Normalization Layer and Data augmentation to reduce the overfitting of models. Just retrain the model or part of it using a low learning rate. This is important because it prevents significant updates to the gradient. These updates result in inferior performance.

C. *The Dropout method.*

The Dropout method (Srivastava et al., 2014)-(Qu et al., 2020) decreases the chance of the over-fitting phenomenon of the model from occurring by randomly discarding the training information. This technique randomly samples the hidden nodes of the activation layer in a certain proportion when the model is backpropagated, so the fully connected network had a certain sparseness, thereby reducing the synergistic effect of distinctive characteristics. Since the hidden nodes appear randomly with a specified probability, the two neurons will not appear repeatably in a similar time period, thus reducing the coadaptation relationship between neurons to improve the overall stability of the network by dropout method.

D. *Final Batch Normalization layer.*

In this work, Final Batch Normalization (BN) layer was attached before each Dense Layer for reducing overfitting problem. It was inspired by the study done by (Kocaman et al., 2021), it shows that the Final Batch Normalization (BN) layer, when placed before the softmax output layer, results in a considerable impact in highly imbalanced image classification problems, also it undermines the role of the softmax outputs as an uncertainty measure. It reduces the dependency on initialization and then decreases overfitting due to its minor regularization effect. Similarly, to dropout, it also adds some noise to each hidden layer's activation. Batch normalization, it is a process to make neural networks stable and faster by adding extra layers in a deep neural network. The new layer performs the standardizing and normalizing operations on the input of a layer coming from a previous layer.

VI. Experiment - Result.

A. *Benchmarking.*

In this experiment, the hardware of Nvidia K80 / T4 of Google Colab was employed for the study. The programming environment was Python 3.7.13, and the deep learning tool used was Tensorflow.

B. *Architectures*

The current research purpose was to study a plant classification approach by comparing four CNN models including DenseNet-201, Xception, Inception-ResNet-V2, and EfficientNet-B7. The multiple CNN architecture models were trained in two ways: first only by the name of the cultivar, and the second one by the name of the cultivar after a preliminary determination of the genus. Later those trained models were evaluated on images of flowers of two genera and different cultivars. The scheme of the proposed model is shown in Fig.12.

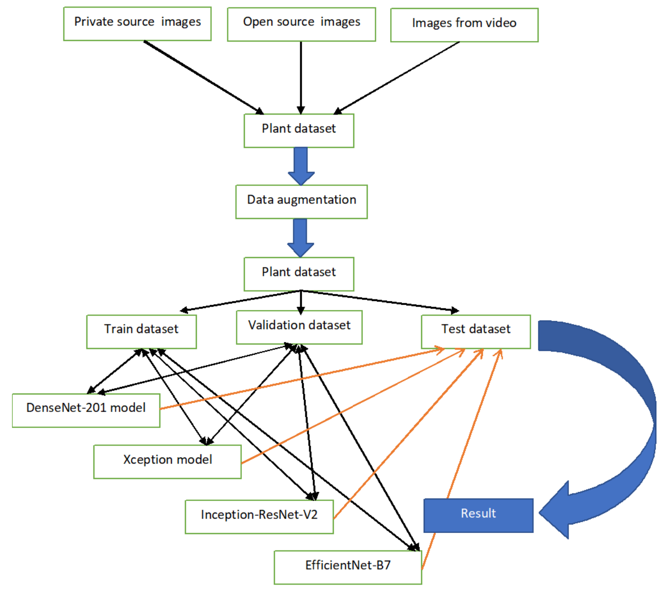


Fig. 12. Method - First step.

The construction process was divided into 5 fundamental components including:

1) Train/Validation/Test split;
2) Image pre-processing;
3) Algorithm training;
4) Verification process;
5) Testing process.

C. *Train/Validation/Test split.*

After collecting the data the dataset was split into Train/Validation/Test datasets. The 80% of the total dataset was used for training the model. 15% of the dataset was used for validation purposes. The remaining 5% of the dataset was used for testing purposes. Some samples of the dataset for the training of the algorithm are shown in Appendix Fig. 20.

D. *Image pre-processing.*

All of the images were rescaled into a constant size of 224×224 in order to be suitable for the training of the model and later they were normalized. Data Augmentation was also used in all classification models to slightly distort an image, and reduce the over-fitting in the training session. The images were randomly processed with horizontal flip, up to 10% zoom, and up to 15 degree rotation with black constantly filling the edge pixels. The images were labeled in the dataset.

E. *Algorithm training.*

This section focuses on a comparative analysis of the existing four plant classification models based on deep transfer learning. The transfer learning technique of a pre-trained model was used and fine-tuned for flower image classification.

It gave faster and more cost-effective training without requiring high computational power and a time-consuming process. The pre-trained models DenseNet201, Xception, InceptionResNetV2, and EfficientNetB7 were used to extract features by using learned weights from the ImageNet dataset. An average pooling was used in this model with a pool size of 3*3 and stride assigned as default which was equal to pool size. Then, the final last layer was retrained by feeding the dataset. The result of this process was the change of the output nodes from 1000 to the number of required classes. For classification, 2 dense layers with 512 and 256 neurons respectively were added with the use of 2 Dropuot layers 0.5 and 2 Batch Normalization (BN) layers. The feature extraction network i.e. DenseNet201 (top removed) followed by softmax activation function for categorical classification.

The objective of the current research is to compare the 4 CNN algorithms to select the one with better performance. Due to the possibility of the model to "remembers" the previous fitting, which was used in the stacking method, and to achieve a clear result from the unpredictable influence of previous models, the separation of coding for each model in a different program with saving the models in files for further evaluation on the test dataset was used. For resetting the model to the "initial" state by applying the K.clear_session() function the following steps should be taken:

1) set (reset) random seeds
2) reset TensorFlow default graph
3) delete previous model

The reset state should be sufficient and reassuring enough in order to prove its reproducibility.

*F. **Verification process.***

The training datasets were randomly split into multiple batches with the batch size set to 10. These batches were fed into the model with a stochastic order, once all batches were fed, one epoch was finished. Each model was trained for 20 epochs. In addition, Adaptive Moment Estimation (Adam) (Kingma and Ba, 2014) was used as the optimization algorithm for computing loss function. Adam is a replacement optimization algorithm for stochastic gradient descent for training deep learning models. Adam combines the best properties of the AdaGrad and RMSProp algorithms to provide an optimization algorithm that can handle sparse gradients on noisy problems. The empirically shown that Adam works well in practice and compares favorably to other adaptive learning-method algorithms (Ruder, 2016). The learning rate of 0.001 and the activation function in each layer was ReLU. Under these settings, the weights and parameters of previous layers were kept fixed, and only removed the last classifier layer and initialized as a new one. In total 16 models were saved: 4 CNN models for each of 4 classes - Genus, Cultivar, Dahlia, and Helleborus.

*G. **Quantitative analyses.***

The proposed deep transfer learning (DTL) based plant classification models were also compared by considering the various confusion matrix-based performance metrics such as precision, recall, F1- score, and accuracy.

There are four ways to check if the predictions are right or wrong:

1) TN / True Negative: when a case was negative and predicted negative
2) TP / True Positive: when a case was positive and predicted positive
3) FN / False Negative: when a case was positive but predicted negative
4) FP / False Positive: when a case was negative but predicted positive

Precision is the ability of a classifier not to label an instance positive that is actually negative. For each class, it is defined as the ratio of true positives to the sum of true and false positives, where TP – True Positives and FP – False Positives. Precision – Accuracy of positive predictions (Sokolova et al., 2006).

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$$

Recall is the ability of a classifier to find all positive instances. For each class, it is established as the ratio of true positives to the sum of true positives and false negatives, where FN – False Negatives. Recall: Fraction of positives that were correctly identified (Sokolova et al., 2006).

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN})$$

The F1 score is a weighted harmonic mean of precision and recall such that the best score is 1.0 and the worst is 0.0. Generally speaking, F1 scores are lower than accuracy measures as they embed precision and recall into their computation. As a rule of thumb, the weighted average of F1 should be used to compare classifier models, not global accuracy (Sokolova et al., 2006).

$$\text{F1 Score} = 2*(\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

Accuracy (Ac) is the measure and it is a ratio of correctly predicted observations to the total observations. Ac can be computed as (Sokolova et al., 2006):

$$\text{Accuracy} = (\text{TP}+\text{TN})/(\text{TP}+\text{FP}+\text{FN}+\text{TN})$$

An assessment was done after every epoch training and measured the performance of the model classification by the Cross-Entropy Loss function to adjust weights and calculate the difference between the output and the label of our image dataset by accuracy and top-k categorical accuracy. Top-1 accuracy is the conventional accuracy, which means that the model answer (the one with the highest probability) must be exactly the expected answer. Top-k accuracy means that any of your models that give k highest probability answers that

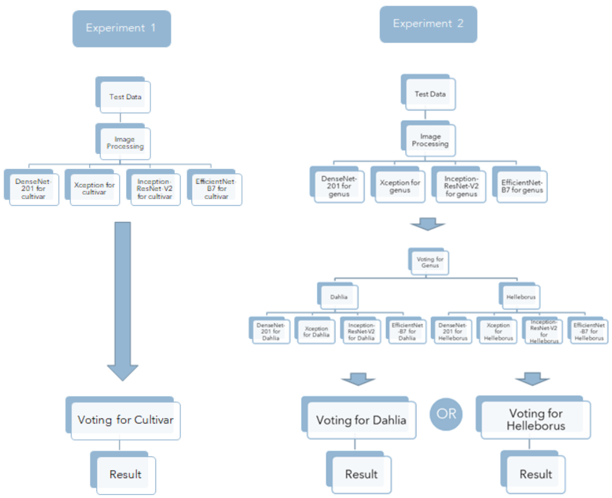


Fig. 13. Structure of experiments.

must match the expected answer. In this research was used the k equals 5.

The Table II shows the training and validation accuracy and lost analyses of the proposed DTL models with respect to models.

H. Testing process.

To provide an unfavoured estimate of the final models the performance of the proposed architecture was evaluated through two experiments and later provided comparisons with 4 CNN models on the test dataset. The schematic of the experiments is shown in Fig.13.

1) Experiment 1: In this experiment the trained models were evaluated on the remain test dataset by classifying the cultivar name of plants such as: Dahlia 'Fancy Pants'.

2) Experiment 2: In comparison, in the second experiment the trained models were evaluated by classifying the same cultivar name of plants but through previously classified genus name of the plant, such as, Dahlia or Helleborus.

3) The majority voting-based classifier ensemble.: The majority voting-based method was selected to measure the performance of stated experiments (Chandra et al., 2021) to combine the results due to the Accuracy for training models being exceptionally low for a multi-class task. The majority voting-based classifier ensemble could act as a multi-expert recommendation and decrease the probable chance of false conclusion. In this research the soft voting ensemble was selected, which involves summing the predicted probabilities for class labels and predicting the class label with the greatest sum probability. Otherwise, if a hard voting ensemble is chosen, it would include summing the votes for crisp class labels from multiple other models and predicting the class with the most votes.

Example of the majority voting-based method shown in Appendix Fig. 18 and 19. Evaluation of the majority voting-based method was done through two steps:

1) was found the probabilities from all 4 trained CNN models and collect in a table.
2) it was summed and sorted, only 5 of them with largest sum can be seen in the results.

The result of testing experiments is shown in Table III

I. Result.

It clearly shows that the proposed models receive significantly more accuracy and loss values if the number of classes is small. When determine the genus class with 2 options the Dahlia or Helleborus the training and validation accuracy convergence at significantly better speed and approaches 100%, which affirms that the model is overfitting extremely quickly after 10 epochs. For more complicated tasks as determining the Helleborus cultivar, name, class the result is slightly lower as it contains 7 options, but model overfitting later after the 15 epochs or more. For task involving class determination for 59 options of Dahlia and 66 options of Cultivar the models are more likely to underfit with the Top-1 accuracy fluctuating within 15.94% (Inception-ResNet-V2 model) - 28.75% (Xception model) accordantly. The Top-5 accuracy get increased to 50.09% - 66.41% for corresponding models.

All these results do not concern the EfficientNet-B7 model which achieves 1.86% for Cultivar and 2.56% for Dahlia - two largest groups of classes and 20,16% for Helleborus classes with 89.1% for Genus class (7 and 2 classes in each group respectively). This result was unpredictably quite unclear. The problem of EfficientNet-B7 model see Confusion Matrix Fig. (VI-I).

Therefore, according to graphs the further increase in accuracy is not expected. Furthermore, comparing performance of each model for given corresponding task indicates that Cultivar with 66 classes outperform better then Dahlia with 59 classes by 3% on average.

While observing metric graphs, it was evident that validation accuracy was much higher than training accuracy. Using the heavy Dropout of 50% with Butch normalisation prevents overfitting, but at same time, due to disabling neurons, some of the information about each sample is lost, and the subsequent layers try to construct predictions based on incomplete representations. Models' overfitting is widely recognized as a concern. It is less so recognized however that overfitting is not an absolute but involves a comparison. A model overfits if it is more complex than another model that fits equally well. This means that recognizing overfitting involves not just the comparison of the simpler and the more complex model but also the issue of how you measure the fit of a model (Hawkins, 2004). The confusion matrixes clearly show that the proposed DenseNet-201 and Xception models provide lesser false negative and false positive rate as it is shown in Appendix Fig. (21 - 36).

TABLE II
METRICS RESULT

Name of the Model	Training Accuracy, %	Top 5 Training accuracy, %	Validation accuracy, %	Top 5 Validation accuracy, %	Training loss	Validation loss
DenseNet201 model for Dahlia	21,44	56,93	51,18	84,71	2,84	1,93
Xception model for Dahlia	28,75	66,41	52,94	87,06	2,49	1,75
InceptionResNetV2 model for Dahlia	15,94	50,09	31,76	70,59	3,01	2,46
EfficientNetB7 model for Dahlia	2,56	11,01	2,94	10,59	4,16	4,07
DenseNet201 model for Helleborus	84,50	100	95,00	100	0,50	0,29
Xception model for Helleborus	84,50	100	100	100	0,55	0,19
InceptionResNetV2 model for Helleborus	65,12	100	85,00	100	0,90	0,44
EfficientNetB7 model for Helleborus	20,16	79,84	15,00	75,00	2,26	1,97
DenseNet201 model for cultivar	24,60	60,19	49,47	90,00	2,76	1,89
Xception model for cultivar	24,68	61,71	49,47	84,74	2,69	1,80
InceptionResNetV2 model for cultivar	18,68	52,49	37,37	68,95	2,99	2,40
EfficientNetB7 model for cultivar	1,86	8,88	2,63	10,00	4,26	4,18
DenseNet201 model for genus	98,82	100	100	100	0,03	0,00
Xception model for genus	98,90	100	100	100	0,04	0,00
InceptionResNetV2 model for genus	97,72	100	100	100	0,06	0,01
EfficientNetB7 model for genus	89,10	100	89,47	100	0,35	0,34

TABLE III
TEST DATA PREDICTION RESULT

	Genus - Cultivar			Cultivar	
	True Name	Predicted Name	Sum of %	Predicted Name	Sum of %
1	Dahlia _Melody Harmony_	Dahlia	389,28		
	Dahlia _Melody Harmony_	Dahlia _Maldini_	25,58	Dahlia _Creme de Cognac_	46,04
2	Dahlia _Cornel Brons_	Dahlia	389,31		
	Dahlia _Cornel Brons_	Dahlia _Creme de Cassis_	106,72	Dahlia _Creme de Cassis_	150,77
3	Dahlia _Bishop of Oxford_	Dahlia	389,20		
	Dahlia _Bishop of Oxford_	Dahlia _Bishop of Llandaff_	114,30	Dahlia _Bishop of Llandaff_	100,49
4	Helleborus _Anna_s Red_	Helleborus	310,42		
	Helleborus _Anna_s Red_	Helleborus _Anna_s Red_	224,83	Helleborus niger (Christmas Rose)	49,76
5	Dahlia _Bishop of York_	Dahlia	389,27		
	Dahlia _Bishop of York_	Dahlia _Black Jack_	120,27	Dahlia _Black Jack_	93,12
6	Dahlia _Fancy Pants_	Dahlia	388,69		
	Dahlia _Fancy Pants_	Dahlia _Fancy Pants_	69,67	Dahlia _Fancy Pants_	58,86
7	Dahlia _Akita_	Dahlia	389,31		
	Dahlia _Akita_	Dahlia _Arabian Night_	52,90	Dahlia _Arabian Night_	79,62
8	Dahlia _David Howard_	Dahlia	388,90		
	Dahlia _David Howard_	Dahlia _Karma Choc_	40,15	Dahlia _Karma Choc_	47,85
9	Dahlia _Kelsey Annie Joy_	Dahlia	389,29		
	Dahlia _Kelsey Annie Joy_	Dahlia _Black Jack_	66,10	Dahlia _Tam Tam_	65,11
10	Dahlia _Blue Boy_	Dahlia	388,98		
	Dahlia _Blue Boy_	Dahlia _Bohemian Spartacus_	23,86	Dahlia _Bohemian Spartacus_	34,05

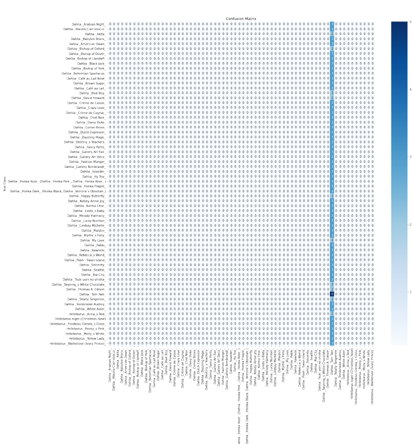


Fig. 14. Confusion Matrix for EfficientNetB7 model for cultivar.

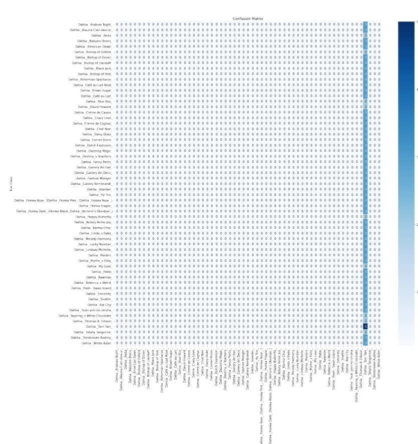


Fig. 15. Confusion Matrix for EfficientNetB7 model for Dahlia.

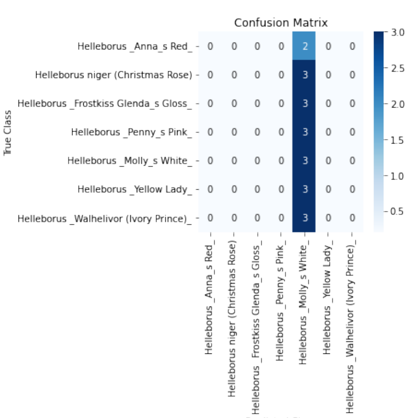


Fig. 16. Confusion Matrix for EfficientNetB7 model for Helleborus.

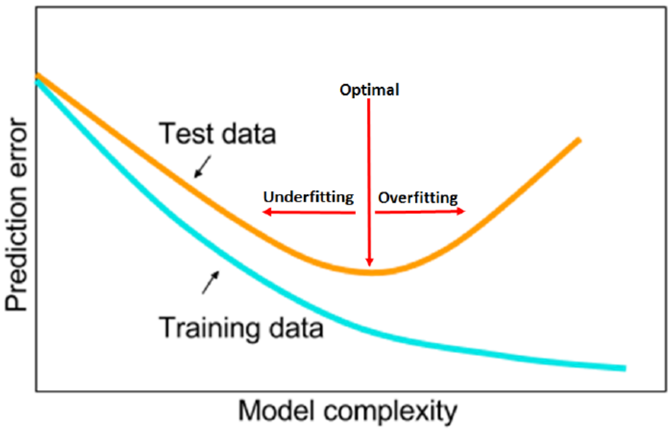


Fig. 17. Overfitting vs Underfitting. Source: (Smith, 2018).

VII. Discussion - ANALYSIS OF EXPERIMENTAL RESULTS

In curent research was analyzed and compared four CNN models. such as DenseNet-201, Xception, Inception-ResNet-V2.

Observing the images in a dataset was descried that large part of them have items possibly affecting the accuracy such as bee, hand, other flower or couple of flowers like bush. Every mentioned images could significantly decrease the performance of whole model not speaking about all of stack of them. Further work is needed to closely examine the impact of regularization strategies such as decreasing Dropout percent or presence of Batch Normalisation Layer. In addition, learning rate variation, pooling layer, size of batches and epochs, size of images, and others.

A. *Limitation.*

This study also has four main limitations:

1) Limitation of value of dataset and quality of images: To prevent an over-fitting the current research has required more data for training. It is exceedingly difficult to find a quality photo in an open internet source. It is preferred to collect the dataset by hand as each image should be inspected manually not just for blur but for background, lighting, and composition as well. The dataset should include images from a controlled environment as well as from a wild field. Some plant cultivars differ in size under corresponding characteristics, differing in terms of flowering season and geographical location, as well as different flowering phases, easy deformation, and withered petals. Additionally, each plant can also be grown differently, in a sunny spot or in a shadow. All these dependencies should be reflected to make a full correct dataset, but due to time limitations they were not. In addition, all images are made in the RGB format what, according to (Sharma et al., 2020) is noisier than HSV. The acquired images contain some noise. This noise should preferably be removed.

2) Multiple comparisons without corrections: In the current research, the main task was to compare some CNN algorithms, thus, all parameters which could change the output of results should be frizzed, even if they produce the overfitting or underfitting and low accuracy. For example, the large classes classification requires a lowered percent of dropping and, potentially, deletion or non-inclusion of the Batch Normalisation layer.

3) A limitation of the voting ensemble: The voting ensemble treats all models the same. This was a problem for the current research as EfficientNet-B7 model revealed the minimum performance in others. Due to the fact that the voting ensemble employs a weighted average or weighted voting of the contributing models, a further extension must be chosen in order to find out whether the machine learning model should require learning when and how much should each model be trusted when making predictions or use the prediction only for each model separately instead of the voting ensemble.

4) Power of machine: There are millions of plant species and cultivars in the world. For proper classification of them all, an enormously powerful computer and a thousand hours of training are required, much more than for training the ImageNet dataset, for example, which contains over 1 million images and over only 1000 classes.

B. *Future work.*

1) Tune the model: In future work, a more accurate and effective deep network for plant image classification could be developed through a better understanding of the effects of data, data augmentation, network depth/width, and other forms of regularization (i.e., dropout ratio, stochastic depth). In addition, the Stacking method with the ability to select the correct models for classification could be used. Dataset supervision is a key to improve the CNN robustness to data variation.

2) Image segmentation and image preprocessing: The table III depicts that the proposed method performed better in terms of overall accuracy. The prediction of cultivar name is better in finding the genus of that particular plant first. For further work, the part of the plant could be defined such as leaf, flower, bud, or bark and later the particular genus or cultivar name could be identified. For that, segmentation, a pattern, and object recognition along with image processing can be applied. In this step, for example, segmentation of images is done in order to separate the leaves from the background. According to Fei Hu (Hu et al., 2020), the accuracy of image segmentation is strongly influenced by the background soil, weeds, and fallen leaves. Hence, the classification effect is overwhelmingly dependent on the accuracy of image segmentation.

The traditional approach to classifying plants has been based on the visible physical characteristics of the plant. The traditional machine learning methods such as Artificial Neural network (ANN), Support Vector Machines (SVM), k-Nearest Neighbour (KNN) can extract handcrafted features such as the colour of the flower or size of the leaf, habit of bush, shape, and texture from raw images remodeled into an appropriate format. Further combining the Machine Learning

(ML) techniques with image classification by Convolution Neural Network (CNN) can improve the performance.

Moreover, the proposed method can be extended for use in other classification applications.

VIII. Conclusion.

The current research project intends to study an innovative approach to plant classification by comparing four CNN models. The multiple CNN architecture models were taught in two ways: beginning only by the name of the cultivar, then by the name of the cultivar after a preliminary determination of the genus. Later, it was assessed using images of flowers of two genus and 66 cultivars. All four CNN models were found to be capable to generate well an automated plant classification system for Genus from 89,1% for EfficientNet-B7 to 98,9% for Xception.

The latest (CNN) achieved the most success in the field of image recognition. CNN is a neural network which comprises of four layers namely Convolutional layer, Pooling layer, Activation function layer, and Fully connected layer. Combining and mixing these layers could give the new algorithms, such as DenseNet-201, Xception, Inception-ResNet-V2, or EfficientNet-B7.

The dataset of different plant flower was collected from various sources: private photo, internet open sources images and extracted private videos. All these images were color images. Images by data augmentation were added to reduce the imbalance of classes which gave outcomes result of 1183 images with 25 images in each class on average.

Four CNN algorithms are trained and compared on the basis of accuracy by The majority voting-based classifier ensemble and loss, using the transfer learning-based approaches. The training datasets were randomly split into multiple batches with the batch size set to 10. These batches were fed into the model with a stochastic order, once all batches were fed, one epoch was finished. Each model was trained for 20 epochs. In addition, Adaptive Moment Estimation (Adam) was used as the optimization algorithm for computing loss function. The learning rate of 0.001 and the activation function in each layer was ReLU. In total 16 models were saved: 4 CNN models for each of 4 classes - Genus, Cultivar, Dahlia, and Helleborus.

Overfitting of models were avoided by the Dropout method, the transfer learning method, the Final Batch Normalization Layer and Data augmentation.

The most important findings are: not all CNN models such as EfficientNet-B7 use their parameters with the same level of efficiency, which achieves 1.86% for Cultivar and 2.56% for Dahlia - two largest groups of classes and 20,16% for Helleborus classes with 89.1% for Genus class (7 and 2 classes in each group respectively). This result was unpredictably quite unclear.

Classification of the plants is crucially important. It helps to rationalize knowledge about plants and correctly identify the plant and which group it belongs to, which in turn allows for scientific research and utilizes it for practical use by many specialists. The suggested model will help the professional and non-professional to correctly classify the plant by photographing the flower.

agsm

References

Abai, Z. and Rajmalwar, N., 2019. 'Densenet models for tiny imagenet classification'.
URL: *https://arxiv.org/abs/1904.10429*

Alipour, N., Tarkhaneh, O., Awrangjeb, M. and Tian, H., 2021. Flower image classification using deep convolutional neural network, *in* '2021 7th International Conference on Web Research (ICWR)', IEEE, pp. 1–4.

Bates, D., 2015. 'The growing world of dahlias'.

Bianco, S., Cadene, R., Celona, L. and Napoletano, P., 2018. 'Benchmark analysis of representative deep neural network architectures', *IEEE Access* **6**, 64270–64277.

Brickell, C. D. e. a., 2009.

Chandra, T. B., Verma, K., Singh, B. K., Jain, D. and Netam, S. S., 2021. 'Coronavirus disease (covid-19) detection in chest x-ray images using majority voting based classifier ensemble', *Expert systems with applications* **165**, 113909.

Dua, D. and Graff, C., 2017. 'UCI machine learning repository'.
URL: *http://archive.ics.uci.edu/ml*

Guan, X., 2021. A novel method of plant leaf disease detection based on deep learning and convolutional neural network, *in* '2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)', IEEE, pp. 816–819.

Hawkins, D. M., 2004. 'The problem of overfitting', *Journal of chemical information and computer sciences* **44**(1), 1–12.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 770–778.

Hu, F., Yao, F. and Pu, C., 2020. Learning salient features for flower classification using convolutional neural network, *in* '2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS)', IEEE, pp. 476–479.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q., 2017. Densely connected convolutional networks, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 4700–4708.

Jaiswal, A., Gianchandani, N., Singh, D., Kumar, V. and Kaur, M., 2021. 'Classification of the covid-19 infected patients using densenet201 based deep transfer learning', *Journal of Biomolecular Structure and Dynamics* **39**(15), 5682–5689.

Kingma, D. P. and Ba, J., 2014. 'Adam: A method for stochastic optimization', *arXiv preprint arXiv:1412.6980* .

Kocaman, V., Shir, O. M. and Baeck, T., 2021. 'The unreasonable effectiveness of the final batch normalization layer'.
URL: *https://arxiv.org/abs/2109.09016*

Kumar Shukla, P., Kumar Shukla, P., Sharma, P., Rawat, P., Samar, J., Moriwal, R. and Kaur, M., 2020. 'Efficient prediction of drug–drug interaction using deep learning models', *IET Systems Biology* **14**(4), 211–216.

URL: *https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-syb.2019.0116*

Lee, J. W. and Chan Yoon, Y., 2019. Fine-grained plant identification using wide and deep learning model ¡sup¿1¡/sup¿, *in* '2019 International Conference on Platform Technology and Service (PlatCon)', pp. 1–5.

Lv, R., Li, Z., Zuo, J. and Liu, J., 2021. Flower classification and recognition based on significance test and transfer learning, *in* '2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)', pp. 649–652.

Pan, S. J. and Yang, Q., 2009. 'A survey on transfer learning', *IEEE Transactions on knowledge and data engineering* **22**(10), 1345–1359.

Qu, N., Li, Z., Zuo, J. and Chen, J., 2020. 'Fault detection on insulated overhead conductors based on dwt-lstm and partial discharge', *IEEE Access* **8**, 87060–87070.

RHS, 2015. 'Royal horticultural society'.

Ruder, S., 2016. 'An overview of gradient descent optimization algorithms'.
URL: *https://arxiv.org/abs/1609.04747*

Sharma, P., Hans, P. and Gupta, S. C., 2020. 'Classification of plant leaf diseases using machine learning and image preprocessing techniques', *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* pp. 480–484.

Smith, L. N., 2018. 'A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay'.
URL: *https://arxiv.org/abs/1803.09820*

Sokolova, M., Japkowicz, N. and Szpakowicz, S., 2006. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation, *in* 'Australasian joint conference on artificial intelligence', Springer, pp. 1015–1021.

Spencer, Roger; Cross, R. L. P., 2007. 'Plant names: a guide to botanical nomenclature. (3rd ed.)'.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. 'Dropout: a simple way to prevent neural networks from overfitting', *The journal of machine learning research* **15**(1), 1929–1958.

Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A., 2016. 'Inception-v4, inception-resnet and the impact of residual connections on learning'.
URL: *https://arxiv.org/abs/1602.07261*

Tan, M. and Le, Q. V., 2019. 'Efficientnet: Rethinking model scaling for convolutional neural networks'.
URL: *https://arxiv.org/abs/1905.11946*

Zhang, J., Zheng, B., Gao, A., Feng, X., Liang, D. and Long, X., 2021. 'A 3d densely connected convolution neural network with connection-wise attention mechanism for alzheimer's disease classification', *Magnetic Resonance Imaging* **78**, 119–126.
URL: *https://www.sciencedirect.com/science/article/pii/S0730725X21000138*

Zhu, W., Braun, B., Chiang, L. H. and Romagnoli, J. A., 2021. 'Investigation of transfer learning for image classification and impact on training sample size', *Chemometrics and Intelligent Laboratory Systems* **211**, 104269.
URL: *https://www.sciencedirect.com/science/article/pii/S016974392100*

APPENDIX

Fig. 18. Confusion Matrix for EfficientNetB7 model for cultivar.

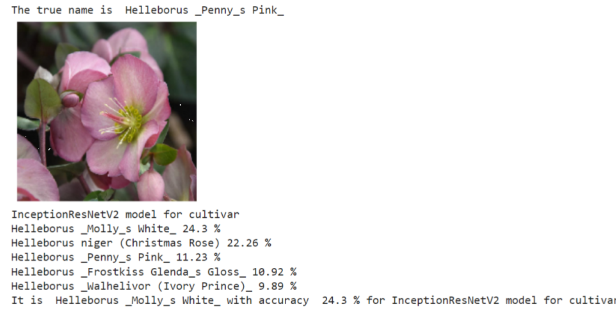


Fig. 19. Confusion Matrix for EfficientNetB7 model for Dahlia.

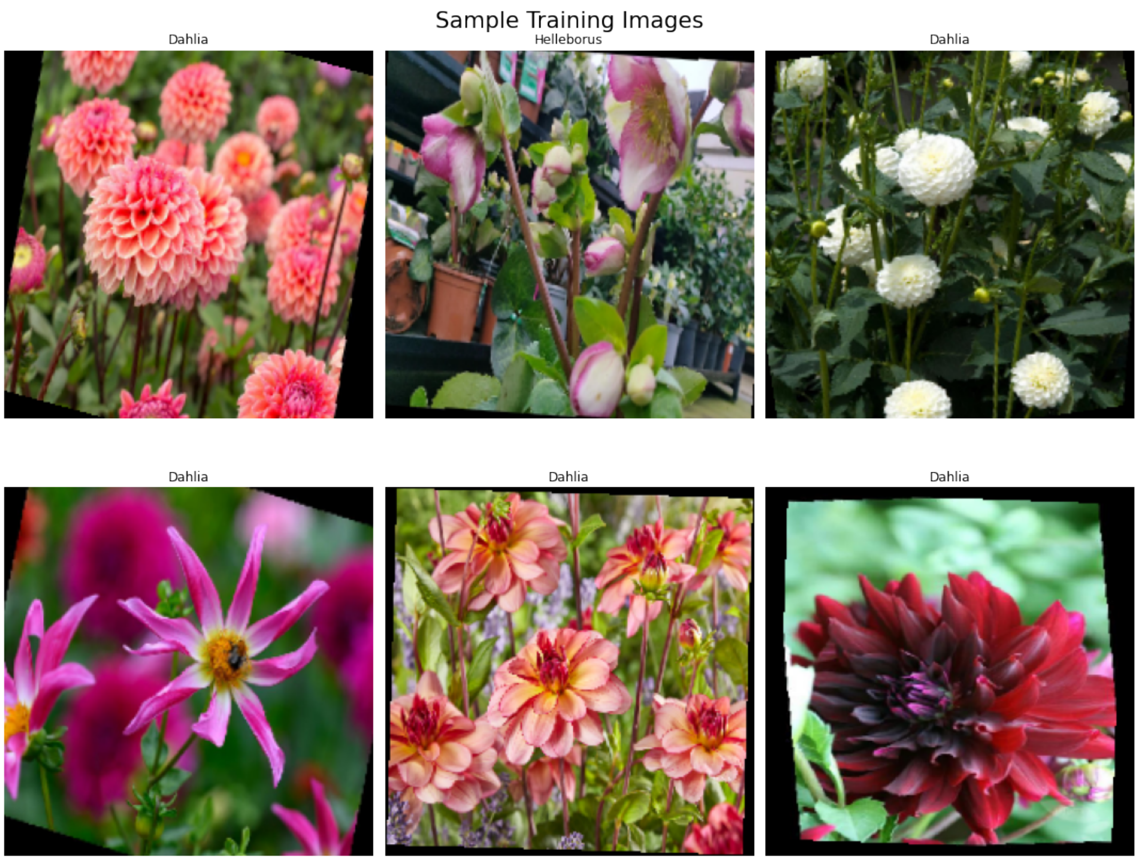


Fig. 20. Training set

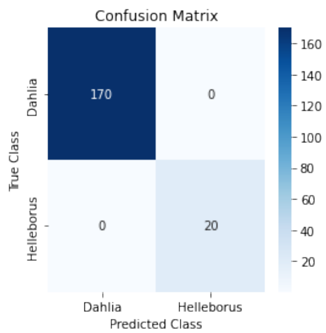


Fig. 21. Confusion Matrix for DenseNet201 model for genus.

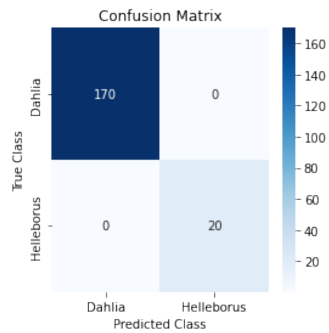


Fig. 22. Confusion Matrix for Xception model for genus.

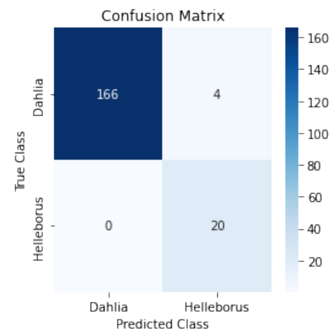


Fig. 23. Confusion Matrix for InceptionResNetV2 model for genus.

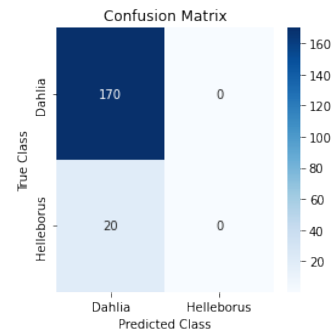


Fig. 24. Confusion Matrix for EfficientNetB7 model for genus.

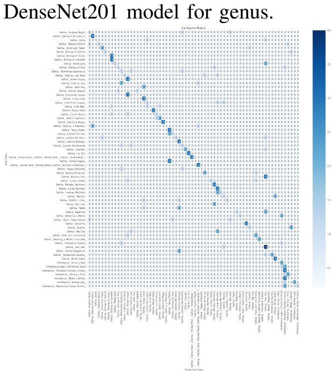


Fig. 25. Confusion Matrix for DenseNet201 model for cultivar.

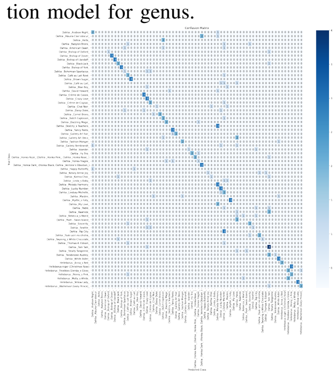


Fig. 26. Confusion Matrix for Xception model for cultivar.

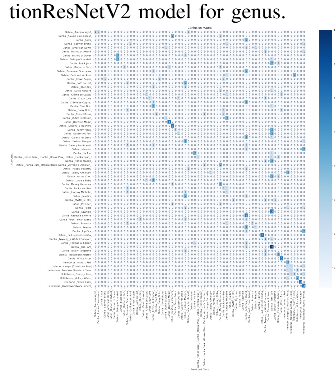


Fig. 27. Confusion Matrix for InceptionResNetV2 model for genus.

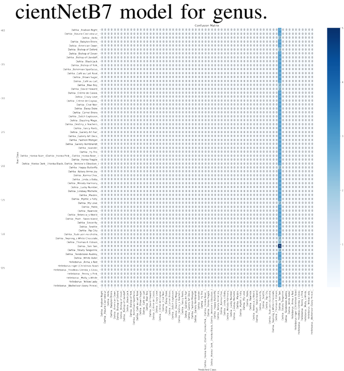


Fig. 28. Confusion Matrix for EfficientNetB7 model for cultivar.

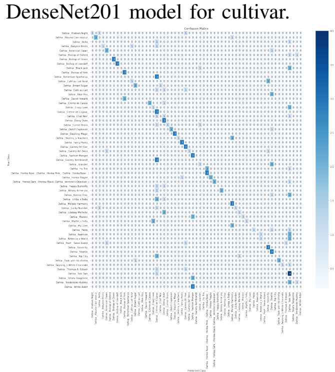


Fig. 29. Confusion Matrix for DenseNet201 model for Dahlia.

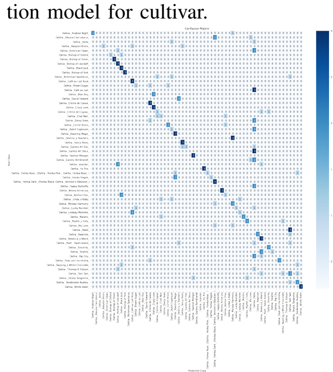


Fig. 30. Confusion Matrix for Xception model for Dahlia.

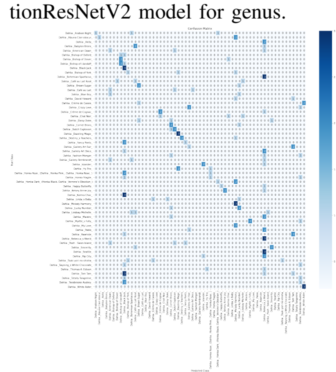


Fig. 31. Confusion Matrix for InceptionResNetV2 model for Dahlia.

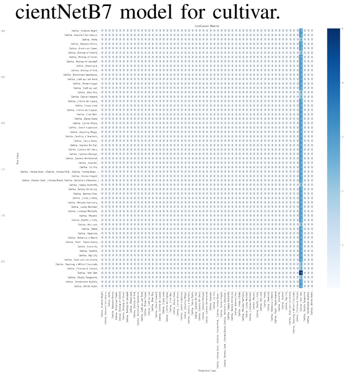


Fig. 32. Confusion Matrix for EfficientNetB7 model for Dahlia.

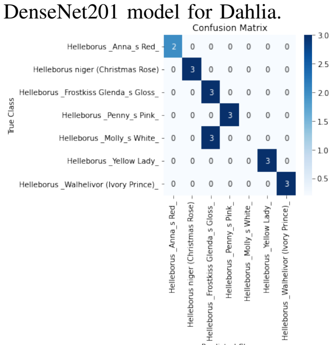


Fig. 33. Confusion Matrix for DenseNet201 model for Helleborus.

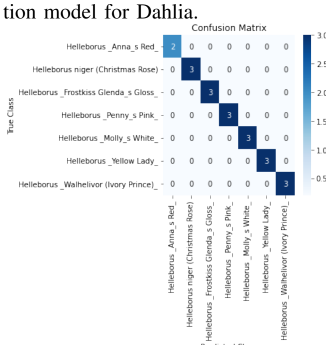


Fig. 34. Confusion Matrix for Xception model for Helleborus.

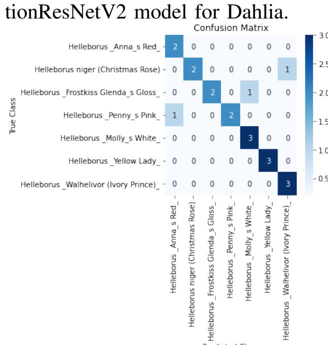


Fig. 35. Confusion Matrix for InceptionResNetV2 model for Helleborus.

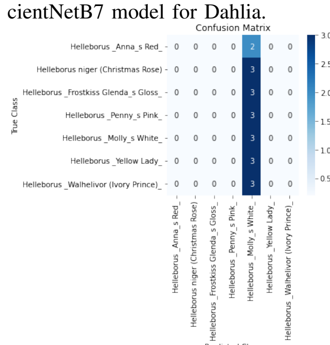


Fig. 36. Confusion Matrix for EfficientNetB7 model for Helleborus.

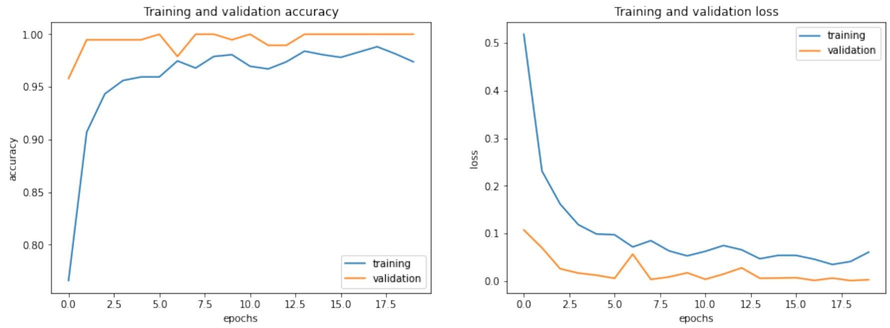


Fig. 37. Training and validation accuracy and loss for DenseNet201 model for genus.

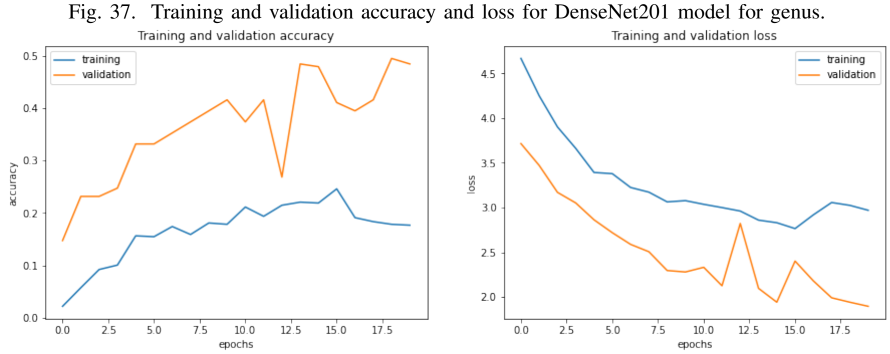


Fig. 38. Training and validation accuracy and loss for DenseNet201 model for cultivar.

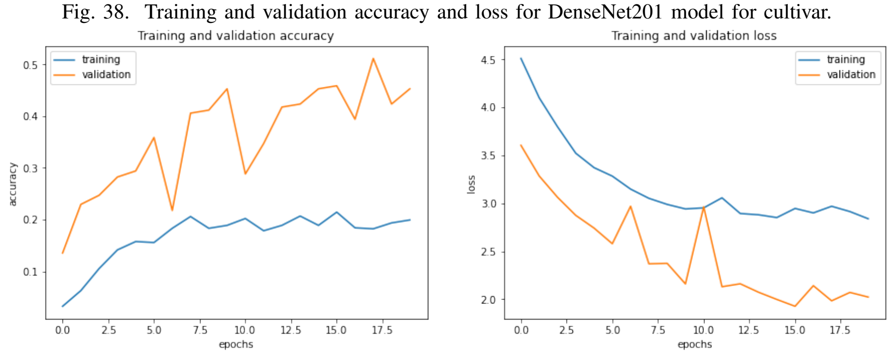


Fig. 39. Training and validation accuracy and loss for DenseNet201 model for Dahlia.

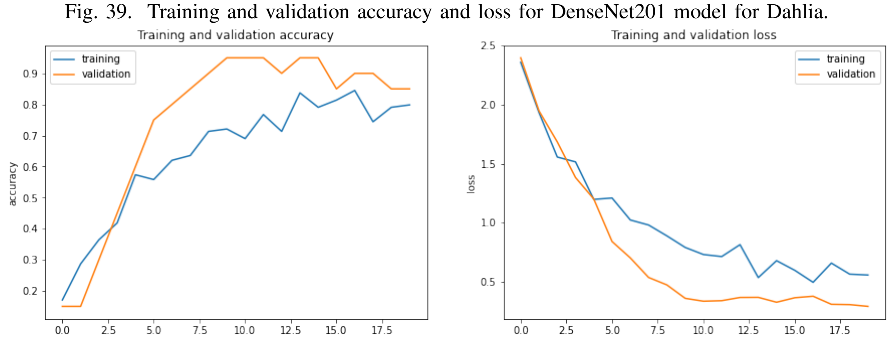


Fig. 40. Training and validation accuracy and loss for DenseNet201 model for Helleborus.

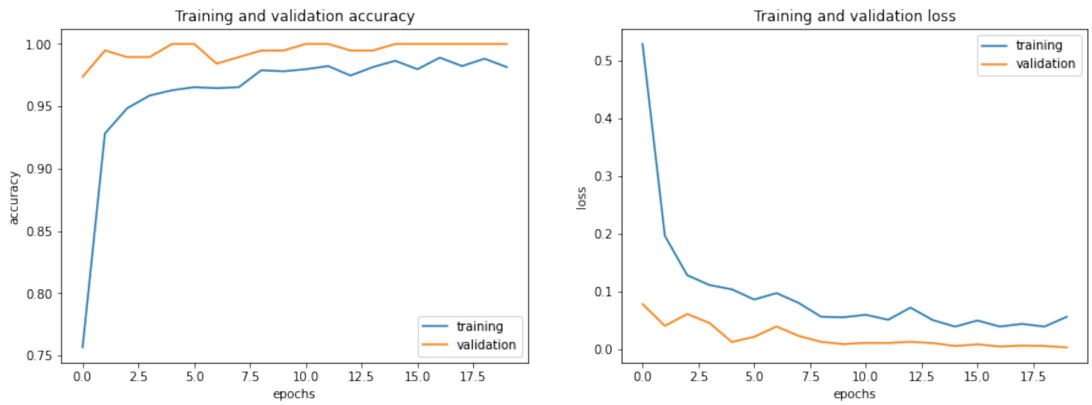


Fig. 41. Training and validation accuracy and loss for Xception model for genus.

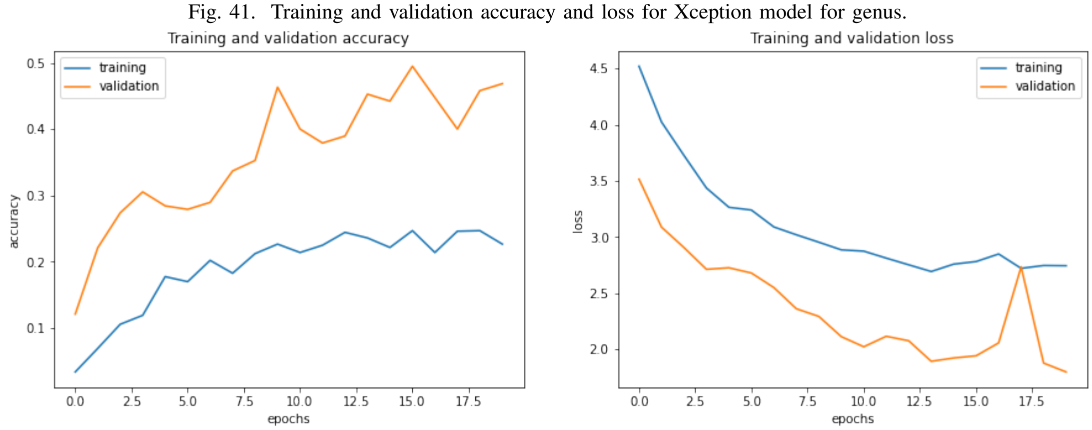


Fig. 42. Training and validation accuracy and loss for Xception model for cultivar.

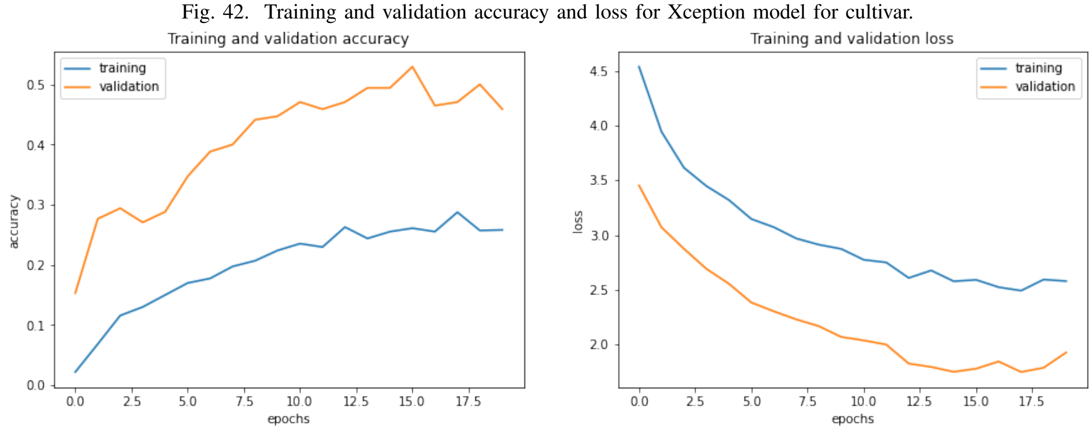


Fig. 43. Training and validation accuracy and loss for Xception model for Dahlia.

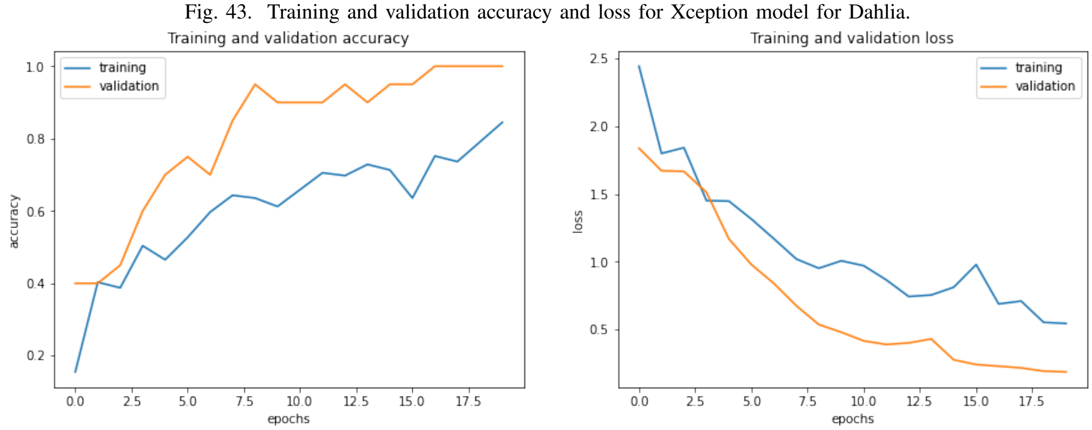


Fig. 44. Training and validation accuracy and loss for Xception model for Helleborus.

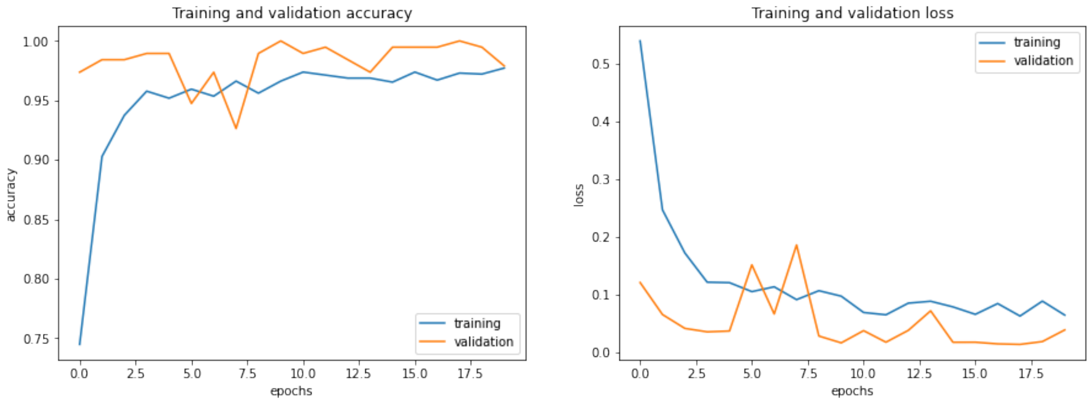


Fig. 45. Training and validation accuracy and loss for InceptionResNetV2 model for genus.

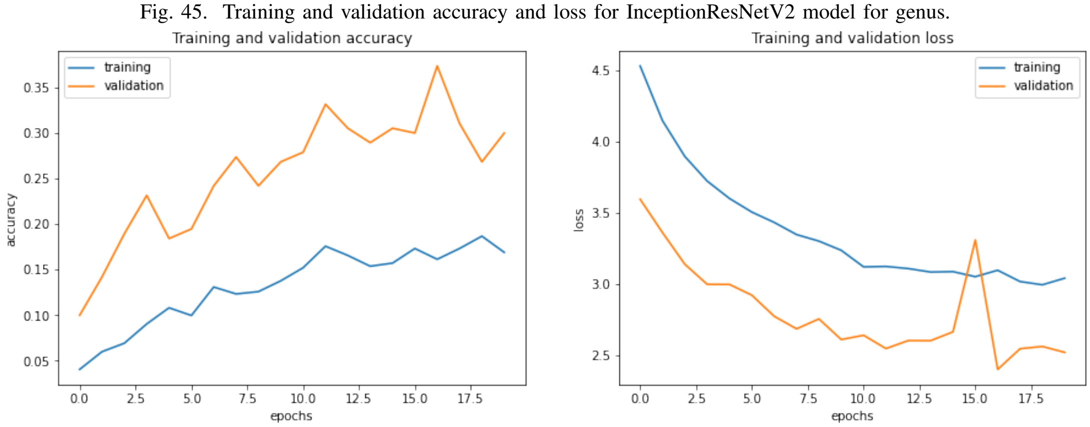


Fig. 46. Training and validation accuracy and loss for InceptionResNetV2 model for cultivar.

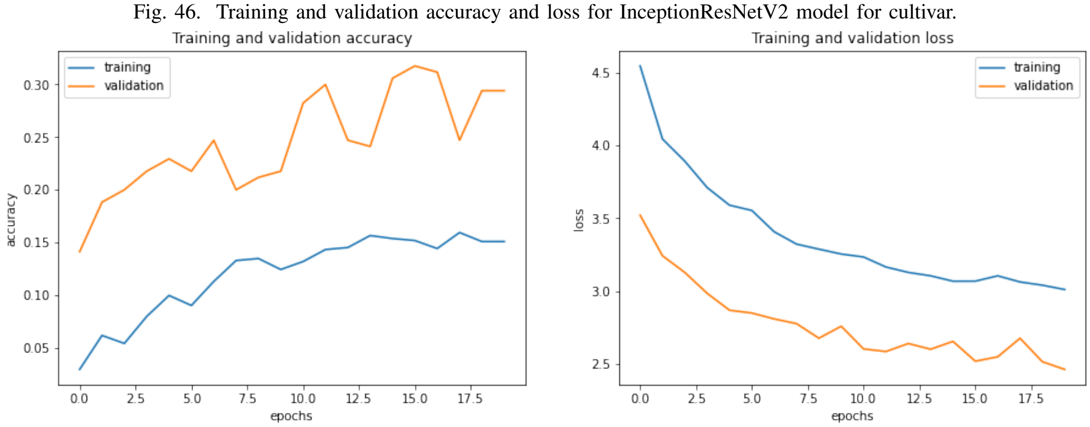


Fig. 47. Training and validation accuracy and loss for InceptionResNetV2 model for Dahlia.

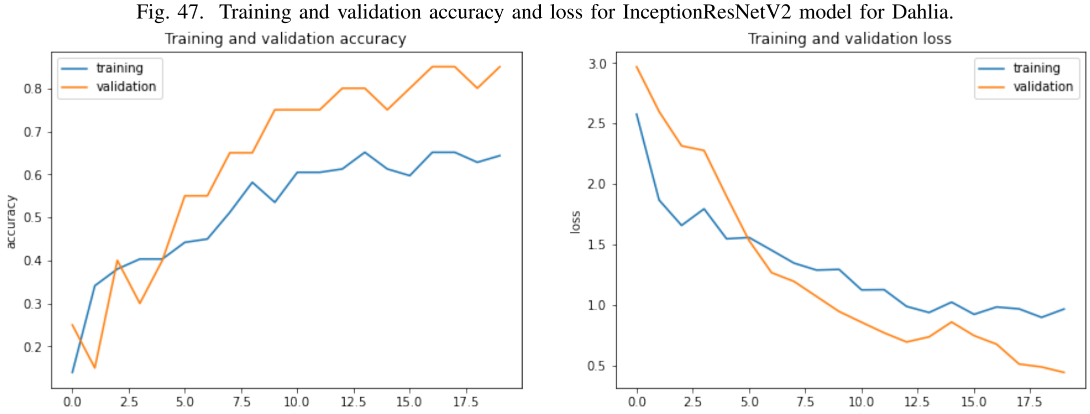


Fig. 48. Training and validation accuracy and loss for InceptionResNetV2 model for Helleborus.

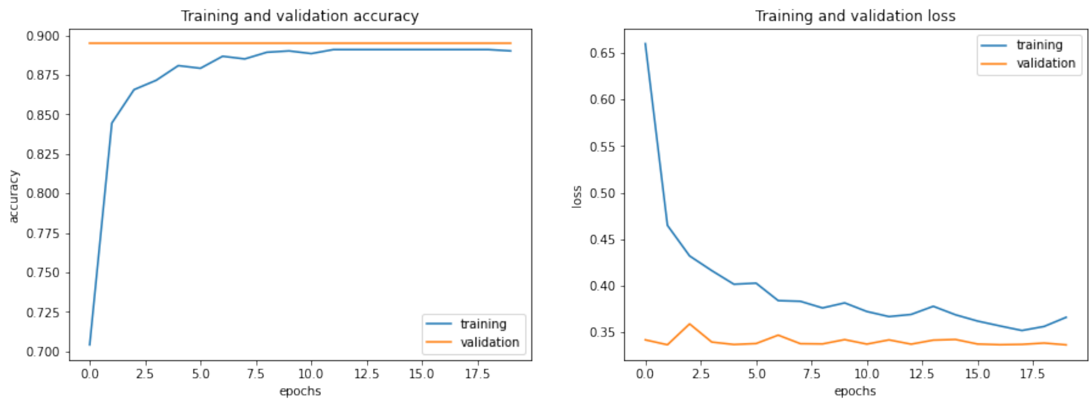


Fig. 49. Training and validation accuracy and loss for EfficientNetB7 model for genus.

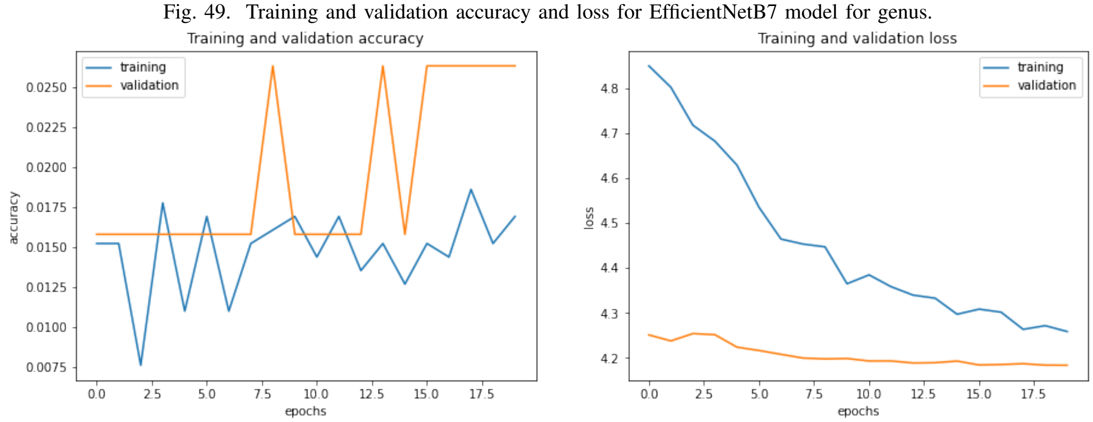


Fig. 50. Training and validation accuracy and loss for EfficientNetB7 model for cultivar.

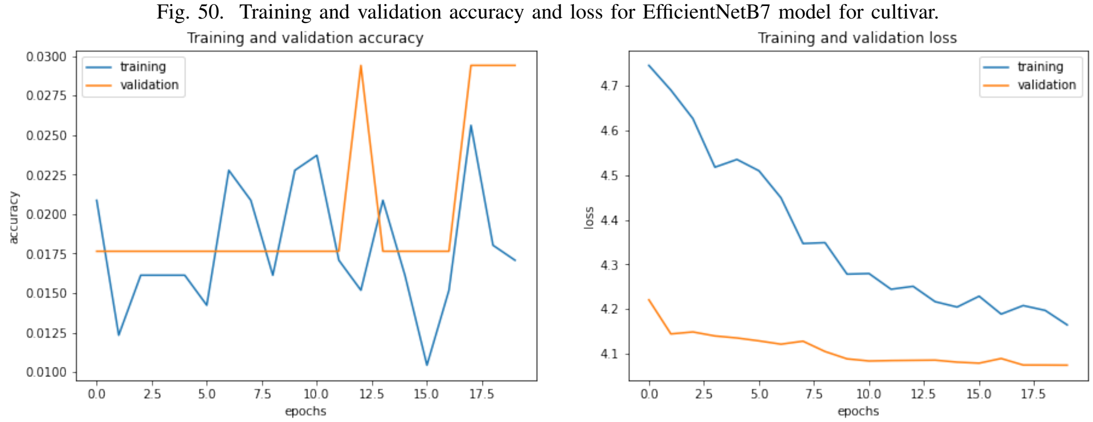


Fig. 51. Training and validation accuracy and loss for EfficientNetB7 model for Dahlia.

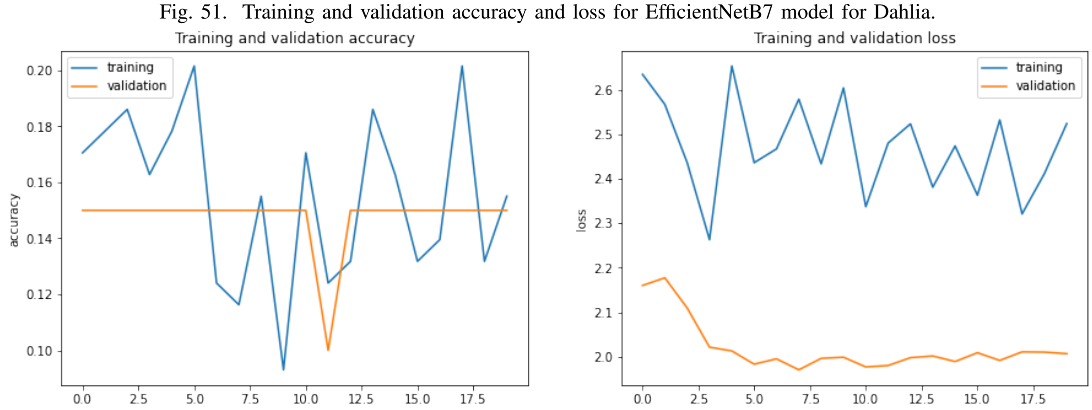


Fig. 52. Training and validation accuracy and loss for EfficientNetB7 model for Helleborus.

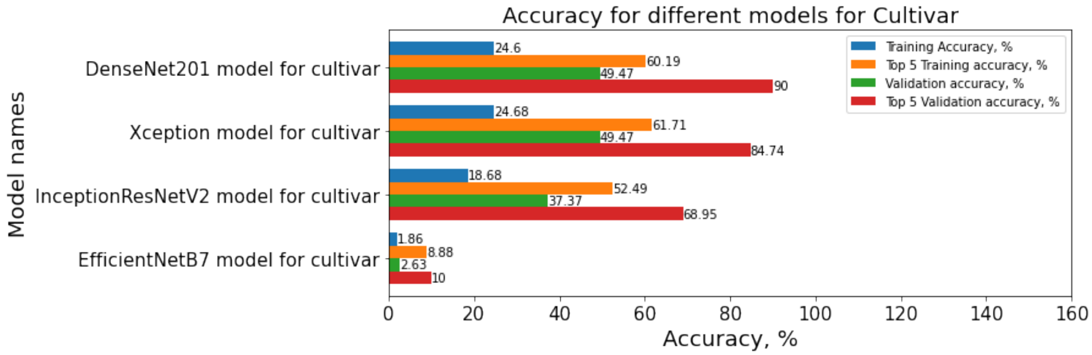


Fig. 53. Training and validation accuracy for cultivar.

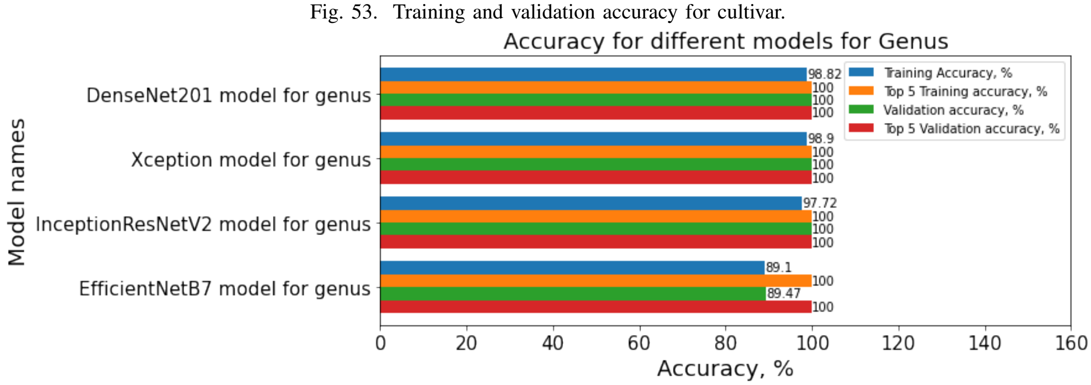


Fig. 54. Training and validation accuracy for genus.

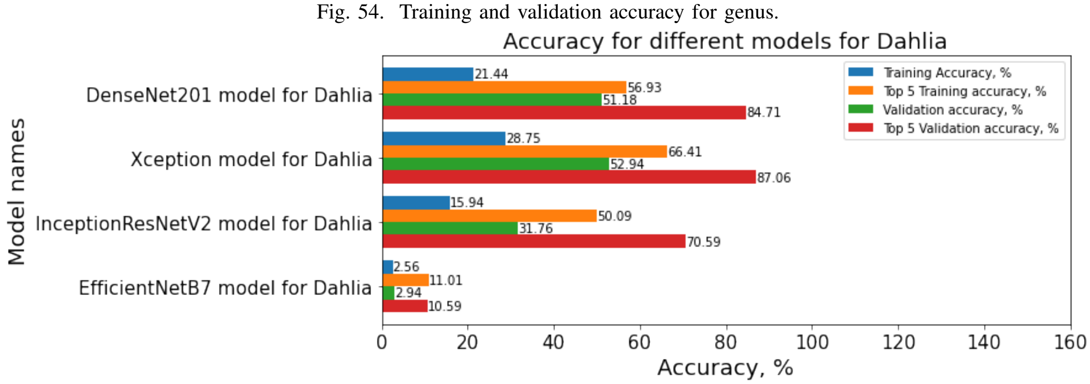


Fig. 55. Training and validation accuracy for Dahlia.

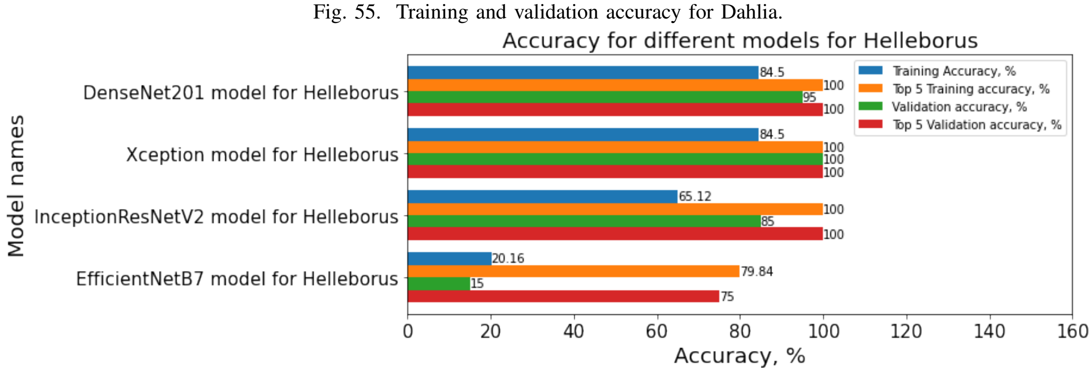


Fig. 56. Training and validation accuracy for Helleborus.