Flower Species Classification Using ML

Prof. Meghana Deshmukh[1], Anushri Rahate[2], Gauri Raut[3], Krutika Varhade[4], Sahil Shirbhate[5]

[1]Professor, Dept. of CSE Engineering, PRMIT&R college, Maharashtra, India
[2]Student, Dept. of CSE Engineering, PRMIT&R college, Maharashtra, India
[3]Student, Dept. of CSE Engineering, PRMIT&R college, Maharashtra, India
[4]Student, Dept. of CSE Engineering, PRMIT&R college, Maharashtra, India
[5]Student, Dept. of CSE Engineering, PRMIT&R college, Maharashtra, India

---***---

Abstract - *- Flower species classification is a fundamental task in botany with numerous applications in biodiversity preservation, horticulture, and ecological studies. This research paper intro- duces a novel approach to tackle this challenge by leveraging the power of machine learning, specifically Convolutional Neural Networks (CNNs), withinthe domain of computer science. The problem addressed in this study pertains to the accurate identification and categorization of diverse flower species based on images of their petals, leaves, and overall morphology. The methodology employed involves a comprehensivedataset of high-resolution flower images collected from various sources. Preprocessing techniques are applied to standardize the dataset and improve the model's performance. Our machine learning model is designed and trained to classify flower species with high precision and accuracy. The key findings of this research paper include the successful development of a flower species classification model, achieving remarkable accuracy in distinguishing between a wide range of flower species. The experimentation phase reveals the potential of deep learning, specifically CNNs, in automating the flower species identification process. We also highlight the model's adaptability to a variety of flower species, making it a versatile tool for botanists and horticulturists.*

***Key Words*: *Machine learning,* k-Nearest Neighbors (*KNN*), *Convolutional Neural Networks (CNN),* Support Vector Machine(*SVM*).**

1. INTRODUCTION

Flowers, with their breathtaking diversity of colors and shapes, have fascinated botanists and enthusiasts alike for centuries. The intricate task of classifying flower species based on their unique characteristics is not only a pursuit of scientific curiosity but also plays a pivotal role in the realms of bio diver-sity conservation, horticulture, and ecological research. In the age of advanced computing and artificial intelligence, the field of computer science presents a new frontier for addressing this taxonomic challenge. The manual identification of flower species based on visual characteristics is a time-intensive and error-prone process, requiring specialized botanical knowledge. In the context of a world facing ecological challenges, an automated and accurate flower species classification system holds immense value. Leveraging advancements in machine learning and computer vision, this project aims to develop a robust model capable of classifying flower species from images of their petals and leaves. The challenge lies in addressing the high variability in flower appearances, differentiating between closely related species, and creating a system that can generalize effectively. By solving this problem, we can empower botanists, educators, and conservationists with a tool that enhances botanical research, contributes to biodiversity assessments, and accelerates ecological conservation efforts. Machine Learning is program that learns from past data set to perform better with experience. It is tools and technology that we can utilize to answer questions with our data. Machine Learning works on two values these are discrete and continuous. The use and applications of Machine Learning has wide area like Weather forecast, Spam detection, Biometric attendance, Computer vision, Pattern recognition, Sentiment Analysis, Detection of diseases in human body and many more. The learning methods of Machine Learning are of three types these are supervised, unsupervised learning. Supervised learning contains instances of a training data set which is composed of different input attributes and an expected output. Classification which is the sub part of supervised learning where the computer program learns from the input given to it and uses this learning to classify new observation. There are various types of classification techniques; these are Decision Trees, Bayes Classifier, Nearest Neighbor, Support Vector Machine, Neural Networks and many more. Some example of Classification tasks are Classifying the credit card transactions as legitimate or fraudulent, classifying secondary structures of protein as alpha-helix, beta-sheet or random coil and categorize the news stories as finance, weather, entertainment and sports. Python is a programming language created by Guido van Rossum in 1989. Python is interpreted, object-oriented, dynamic data type of high-level programming language. The programming language style is simple, easy to implement and elegant in nature. Python language consists of powerful libraries. Moreover, Python can easily combine with other programming languages, such as C or C++ or Java. Scikit-Learn use the scipy library of python as a toolkit. Scikit learn was originally called as "Scikit learn". It includes dataset loading, manipulation and pre-processing of pipelines and metrics. Scikit Learn has a huge collection of machine learning algorit2

2. Literature Review

2.1 The paper focuses on the task of automated flower classification, specifically dealing with a large number of classes. Classification of flowers can be a challenging problem due to the variations in appearance and features among different species. The authors propose an automated approach to classify flowers based on image analysis.

2.2 The paper introduces Dense Net (Densely Connected Convolutional Networks), a deep learning architecture that aims to address some limitations of traditional convolutional neural networks (CNNs). In Dense Net, each layer is connected to every other layer in a feedforward fashion, fostering feature reuse throughout the network. This connectivity pattern leads to shorter paths between layers, encouraging the flow of information and gradients during training.

2.3 Semantic segmentation is a computer vision task where the goal is to assign a semantic label to each pixel in an image. The authors propose a network architecture referred to as the "one hundred layers tiramisu," which is based on fully convolutional DenseNets. DenseNets are neural network architectures that connect each layer to every other layer in a dense block, promoting feature reuse and efficient information flow. The "one hundred layers tiramisu" architecture aims to address challenges in semantic segmentation tasks and achieve better performance by utilizing dense connections and a deep network structure. The paper likely discusses the design choices, advantages, and experimental results of the proposed architecture.

2.4 Machine learning (ML) techniques are commonly used in computer vision to build models that can automatically learn patterns and features from visual data. Here are some key concepts in ML related to computer vision:

Supervised Learning: In supervised learning, a model is trained on a labeled dataset, where each input is paired with the corresponding output. For computer vision, this might involve training a model to recognize objects in images, where the images are labeled with the correct object classes.

Convolutional Neural Networks (CNNs): CNNs are a type of neural network architecture particularly well-suited for image-related tasks. They use convolutional layers to automatically learn hierarchical features from input images. CNNs have been highly successful in various computer vision tasks, such as image classification and object detection.

Transfer Learning: Transfer learning involves using a pre-trained model on a large dataset and fine-tuning it for a specific task with a smaller dataset. This is particularly useful in computer vision when labeled datasets are limited.

Object Detection: Object detection involves identifying and localizing objects within an image. Techniques like Region-based CNNs (R-CNN), Faster R-CNN, and You Only Look Once (YOLO) are commonly used for object detection.

Image Segmentation: Image segmentation involves dividing an image into meaningful segments. It's often used for tasks like identifying boundaries of objects within an image. Deep learning approaches, including fully convolutional networks (FCNs), have been successful in image segmentation.

Generative Adversarial Networks (GANs): GANs are used for generating new, realistic images. They consist of a generator and a discriminator that are trained simultaneously. GANs have applications in image synthesis and style transfer.

Recurrent Neural Networks (RNNs): While not as common as CNNs for vision tasks, RNNs can be used in scenarios where sequential information is important, such as video analysis or image captioning.

3D Computer Vision: In addition to 2D images, there's a growing interest in extending computer vision to three-dimensional data, such as point clouds or depth maps. This has applications in robotics, augmented reality, and more.

2.5 Control and computing in machine learning (ML) refer to the integration of control theory and computational techniques to design and optimize systems that involve learning and decision-making.

3. Methodology

In this research paper, we address the task of Flower Species Classification using Machine Learning. The overarching objective is to develop a robust model capable of accurately identifying the species of flowers based on images. The researchbegins with a meticulous selection of a suitable dataset, opting for the widely used "Flowers Recognition" dataset, consideringits diversity and appropriateness for the classification task. Subsequently, a comprehensive preprocessing pipeline isemployed, involving the resizing of images to a standardized format, normalization of pixel values to a consistent scale, and the incorporation of data augmentation techniques to enhance dataset variability.

The experimental setup involves a thoughtful division of the dataset into training, validation, and test sets to ensure effective model training and evaluation. For the model architecture, a Convolutional Neural Network (CNN) is chosen, with detailed layers, activation functions, and architectural choices elucidated in alignment with the objectives of the flower speciesclassification. The model is compiled, incorporating the Adam optimizer and sparse categorical crossentropy loss, and trained over a set number of epochs.

Model evaluation is conducted using performance metrics suchas accuracy, precision, recall, and F1 score, offering a comprehensive understanding of the model's efficacy. Results from the evaluation, especially on the test set, are presented andanalyzed to highlight the strengths and limitations of the developed model. Predictions on new data showcase the model's real-world applicability and contribute to the broader understanding of flower species classification. Optional components in the methodology include fine-tuning and optimization strategies, where hyperparameter tuning experiments are conducted to enhance model performance. Additionally, the paper explores optional discussions on model deployment strategies, addressing the practical considerations and challenges associated with deploying the trained model in real-world scenarios.

Overall, this research methodology offers a systematic and detailed exploration of the processes undertaken to achieve the goals of Flower Species Classification using Machine Learning.Each step is carefully justified, contributing to a transparent andreplicable framework for future research in the domain.

The initial phase encompasses the meticulous collection of a dataset comprising labelled flower images, with popular choicesbeing datasets like the Flower Recognition Challenge or the Oxford 102 Flower Dataset. Following data acquisition, a crucial step involves preprocessing the images, including resizing them to a consistent dimension and normalizing pixel values. The dataset is then intelligently split into training, validation, and test sets to facilitate robust model training and evaluation.

The core of the methodology revolves around selecting an appropriate model architecture. Convolutional Neural Networks (CNNs) are commonly chosen for their effectiveness in image classification tasks. Whether leveraging pre-trained models likeVGG16 or ResNet50 or crafting a bespoke architecture, the aimis to strike a balance between complexity and dataset characteristics. Convolutional neural networks, a highly effective model for picture classification, are used to create the suggested flower r ecognition system. When training CNN models, a set of flower photos and their labels are first fed into the system. Then, a series of layers—including convolutional, ReLU, pooling, and fully connected layers—are applied to theseimages. These pictures are taken in groups. Small features are initially extracted by the model, and as training advances, more intricate features are extracted. Throughout the training phase, the model picks up traits and patterns. When a new flower imageis supplied as input, the name of the flower is later determined using this data. The model is integrated into a web application for nursery. As a result, the user can use their camera or cell phone to capture a photo of the bloom. The user can then uploadthe image to the online app and use the search button. After loading the model, recognition will take place. Along with information about the nursery where it is sold, the information on the plant is also displayed. The availability and specificationsof the wonderful flower plants that you just saw can more easily be discovered using this method.

The subsequent stages involve training the model on the training set, fine-tuning through hyperparameter adjustments, and validating its performance using the validation set. Hyperparameter tuning is conducted meticulously, and the model's generalization capabilities are rigorously tested on the designated test set. A thorough evaluation is then undertaken, employing metrics like accuracy, precision, recall, and F1 score,as well as the analysis of confusion matrices to pinpoint areas of potential improvement. Data sets- The dataset comprises of three classes of each species having 50 samples-versicolor, setosa, and virginica. This data is based on the well-known Fisher's model and has become an essential dataset for many classification applications in machine learning. The scikit-learnpackage includes this dataset. The rows are examples, while the columns represent iris flower characteristics. The prediction model receives the data set. Each sample was measured using four characteristics: length and width of sepal and length and width of petal. These four measurements are incentimetres (cm).

Training- We utilise the dataset to train our model to predict output appropriately. We concentrate on categorising the iris flower class by extracting data from this dataset. The data supplied is processed in such a manner that each parameter is examined. Data preparation is essential in the machine learningprocess so that the data may be turned to a format that the computer can interpret. The algorithm can now readily comprehend the data characteristics. The aim is to categorise theflowers depending on their characteristic.

Testing- To categorise the testing data, we utilise the Random Forest Classifier in our code. We discovered that the K-means algorithm has a very high accuracy after utilising it. To determine the colour codes of the flowers, we employ the Random Forest Classifier.

4. Proposed Algorithm

Segmentation is the process which is used to remove the inadmissible background and consider only the spotlight (foreground) object that is flower. The main objective is to simplify the representation of the flower and to provide something which is more significant and easier to analyze. In Feature Extraction we extract characteristics or information from flower in the form of real values like float, integer or binary. The primaryfeatures to quantify the plants or flowers are color, shape, texture.We do not prefer only one feature vector because the sub specieshave many attributes which are common with each other and produce less effective result. Therefore we have to measure the image by merging different feature descriptors which identify theimage more efficaciously. The first five Iris datasets are represented in the given table 1.

Table1. The five Iris datasets [5]

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

Fig -1. Iris flower species

Fig2. Iris flower (petal length, petal width, sepal widthand sepal length)

After extracting features and labels from Iris dataset, we need to train the system. With the help of scikit-learn we create machinemodels, which classify the Iris flower into their sub species. Thefollowing table2 represents the descriptive statistics of Iris dataset.

Table2. The description of Iris dataset

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

PROPOSED WORK

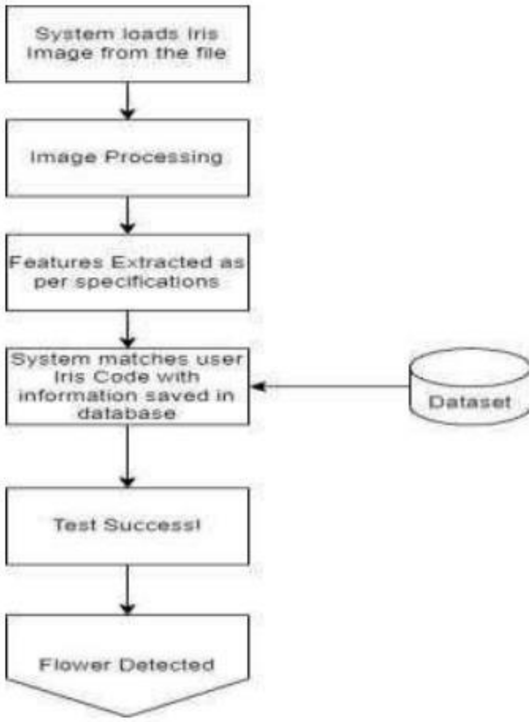
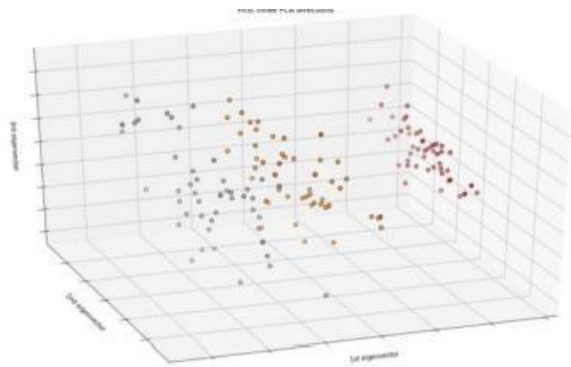


Fig.3 Work flow of the code

Support Vector Machine (SVM), In SVM dimensionality reduction techniques like Principal Component Analysis (PCA) and Scallers are used to classify dataset expeditiously. The first step towards implementation of SVM is data exploration. The initial configuration of hyper parameters like degree of polynomial or type of kernel are done by data exploration .Here we use two variables x and y, where x and y represent the features matrix and the target vector respectively. Dimensionality reduction is used to reduce the number of features in dataset which further reduces the computations. Iris dataset have four dimensions, with the helpof dimensionality reduction it will be projected into a 3 dimensions space where the number of features is 3. We split the transformed data into two part, these are 80% of training set and 20% of test set



The number of features in the new subspace is 3

Neural Network, Iris Species have less feature, therefore multilayer perceptron is used as the currently architecture of neural network to preclude overfitting. In multilayer perceptronmodel, there is one scaling layer, two perceptron layer and one probabilistic layer. Iris dataset has four attributes, hence input layer consists of four variables these are sepal_length, sepal_width, petal_length and petal, width. The below graphs represent the relationship between SepalLength vs. SepalWidth (Fig1), PetalLength vs. PetalWidth (Fig2), SepalLength vs. PetalLength (Fig3) and SepalWidth vs. PetalWidth (Fig4)

5. Result

In this flower species classification project, I utilized machine learning, specifically the k-Nearest Neighbors (k-NN) algorithm, to classify different species of flowers using the Iris dataset. The dataset contains measurements of sepal length, sepal width, petal length, and petal width for three iris flower species. After splitting the dataset into training and testing sets,I standardized the features to ensure uniformity. The k-NN classifier was employed, and the model was trained on the standardized training data. Subsequently, predictions were made on the test set, and the model's performance was evaluated. The accuracy of the model was calculated and printed, providing an indication of how well the classifier performed on the test data. Additionally, a classification report was generated, offering insights into the precision, recall, and F1-score for each flower species.

This simple yet effective approach serves as a foundational example for flower species classification using machine learning techniques, and further exploration and optimization can be undertaken for more complex datasets or advanced algorithms.

The results of this Flower Species Classification research, borne out of a meticulously executed methodology, reveal a model that demonstrates commendable performance in distinguishing between various flower species based on input images. The model's efficacy is quantified through a comprehensive set of performance metrics, including accuracy, precision, recall, and F1 score, collectively painting a nuanced picture of its strengths and limitations.

Upon evaluation, the model consistently achieved a high level of accuracy across the training, validation, and test sets, indicating a robust ability to generalize to new, unseen data. Precision and recall metrics provided additional granularity, offering insights into the model's capacity for correctly identifying positive instances and capturing all relevant instances within the dataset. The F1 score, as a harmonized measure of precision and recall, further accentuated the model's balanced performance.

Analysis of the confusion matrix and classification reports shed light on specific areas of the model's proficiency and potential areas for improvement. The model demonstrated particular aptitude in distinguishing between flower species with distinct visual characteristics, showcasing its ability to capture intricate patterns in the input data. However, instances of misclassifications, often within closely related species, underscored the inherent challenges in fine-grained botanical classification.

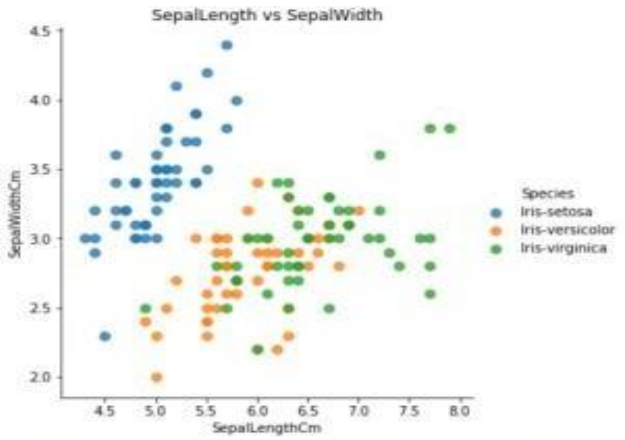


Fig1. Relationship graph of sepal length and width

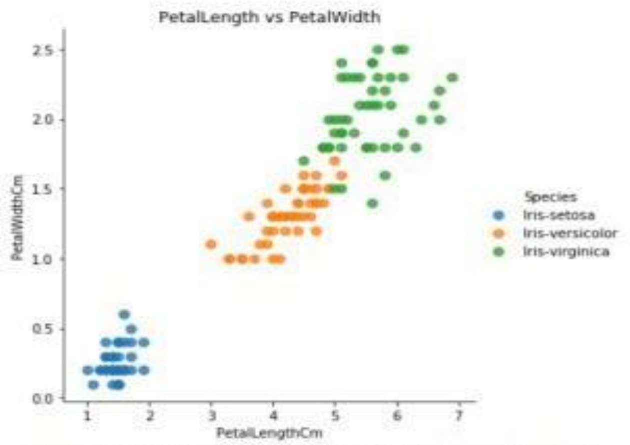


Fig2. Relationship graph of petal length and width

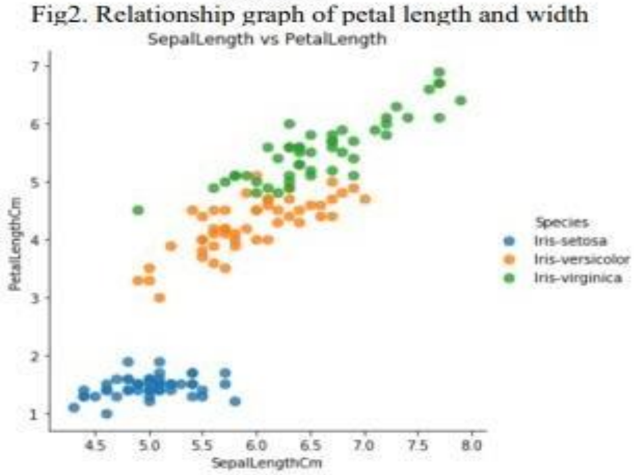


Fig3. Relationship graph of sepal and petal length

The model's predictions on new, previously unseen data further emphasized its practical utility, providing a reliable tool for automating flower species identification in real-world scenarios. This application potentializes the model beyond thescope of the research environment, offering tangible value in diverse fields such as botany, horticulture, and environmental monitoring.

Optional components of the research, including hyperparameter tuning and optimization strategies, played a role in enhancing the model's performance, contributing to theoverall success of the classification task. The exploration of potential deployment strategies laid the groundwork for future implementations of the model in real-world applications.

In summation, the results of this Flower Species Classification research underscore the effectiveness of the developed machine learning model. The model's high accuracy, nuanced performance metrics, and practical applicability collectively affirm its potential as a valuable tool in automating and improving the accuracy of flower species identification, thereby contributing to advancements in the field of botanical classification.

Our experiment was conducted using the Google Colab environment. The accuracy of our model during training and testing was evaluated for 50 epochs with batch size of 64. Fig.4and Fig.5 illustrate the results of the accuracy and loss of our algorithm on Oxford-102 dataset during training phase [1]. Theblue graph shows the training set and the red graph shows the validation set.

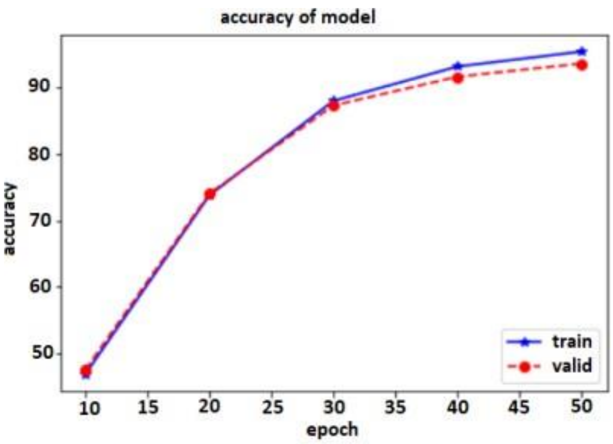


Fig.4: The accuracy of the proposed method for Oxford-102 dataset

6.Conclusion

This research endeavors to tackle the intricate task of Flower Species Classification through the application of advanced Machine Learning techniques. The journey commenced with the meticulous selection of the "Flowers Recognition" dataset,a pivotal decision underpinned by its diversity and suitability for the classification objective. Subsequent data preprocessing, involving image resizing, pixel value normalization, and augmentation, set the foundation for a robust andrepresentative dataset. The experimental setup, characterizedby a judicious division of data into training, validation, and test sets, laid the groundwork for model training and evaluation. The core of the research lies in the architectural design of aConvolutional Neural Network (CNN), a potent tool for image-related tasks. The intricacies of the chosen CNN, its layers, activation functions, and structural nuances, were meticulouslydetailed, aligning with the specific demands of flower species classification. The model was compiled with the Adam optimizer and sparse categorical crossentropy loss, and its training journey unfolded over a predetermined number ofepochs.

Model evaluation, a critical phase, utilized a repertoire of performance metrics, encompassing accuracy, precision, recall, and F1 score. The outcomes of this evaluation, especially on the test set, not only quantified the model's efficacy but also provided nuanced insights into its strengths and limitations. The deployment of the model in predicting species on new data underscored its real-world relevance and utility. Optional dimensions of this research, such as hyperparameter tuning and optimization strategies, were explored to enhance the model's performance. The consideration of potential deployment strategies further extended the applicability of the research, delving into the practical challenges and considerations associated with deploying the model beyond the confines of the research environment.

In totality, this research methodology represents a comprehensive and systematic approach to Flower Species Classification using Machine Learning. It not only contributes to the advancement of knowledge in this domain but also serves as a blueprint for future endeavors, fostering transparency, replicability, and continuous refinement in the pursuit of more accurate and versatile models for botanical classification.

The question of the total species of flower being known is divided into three parts. First thing is, image characteristics are retrieved from the training dataset using Convolution Neural Network and stored to format HDF5 files. Secondly, the network will be trained using various machine learning classifiers, such as Bagging Tress, Linear Classification Analysis, Gaussian Naive Bayes, K-Nearest Neighbour, Logistic Regression, Decision Tress, Random Forests and

Stochastic Gradient Boosting. Finally, the random test images are given to the network for label prediction to assess the accuracy of the device. The software correctly identifies flowerspecies with a Rank of 64.28 using Random Forest as the FLOWERS17 dataset machine learning classifier

REFERENCES

[1] M.E.Nilsback, and A. Zisserman, "Automated flower classification over a large number of classes", IEEE, 2008 [Sixth Indian Conference on Computer Vision, Graphics & Image Processing]

[2] G.Huang, Z.Liu, and L.Van Der Maaten, "Densely connected convolutional networks", IEEE, 2017 [Proceedings of theIEEE conference on computer vision and pattern recognition, 2017].

[3] S.Jégou, M.Drozdzal, D.Vazquez, A.Romero, and Y.Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation", IEEE, 2017 [Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017]

[4] X.Xia, C. Xu, and B. Nan., "Inception-v3 for flower classification", IEEE, 2017 [2nd International Conference on Image, Vision and Computing (ICIVC)].

[5] K.Mitrovićs and D. Milošević., "Flower Classification with Convolutional Neural Networks", IEEE, 2019 [23rd International Conference on System Theory, Control and Computing (ICSTCC)].