How deep learning identifies and learns aspects of plant for classification

V. Sowmya Devi[1], *Govu* Pooja[1], *Sande* Ramya[1], and *Nakka* Vishnu Priya[1]
[1] Department of CSE, Sreenidhi Institute of Science and Technology, India

Abstract— Identification of plant species is very crucial for preserving biodiversity, developing farm-based practices and ecosystem administration. By merging classical structural approach with innovative machine learning methods. We gathered data from diverse plants highlighting mainly on leaves, fruits, flowers and vegetables. Focusing on their dimension, outline and their features, and also environmental factors like climate and soil type. To extract optimal features which are missed by the traditional methods to overcome this we took high quality snapshots with advanced image processing methods. Convolutional Neural Network (CNN) of supervised learning algorithm used along with the Keras and Tensorflow the features were analysed and deployed in Streamlit. Accomplishing an accuracy of 95% on a dataset which was trained over 10,000 plant samples with conventional strategies. The outcomes boost the process and also improves identification by making reliable tool and helps for ecologists, botanists and agricultural scientists. This study achieves a robust solution for plant species identification by integrating advanced algorithms and traditional techniques. It encourages for conservation analysis and wildlife protection initiatives.

Keywords— Machine Learning, Classification, Convolutional Neural Network, Streamlit.

1. Introduction

Plants are very essential part in our life. Plants provides us fresh air to breath, medication and many on earth. It is a backbone for all our life. Plant species identification is very helpful to botanists to know the plant species and its uses by using some advanced knowledge algorithms with good accuracy [1]. Plant species are infinite that botanists also could not identify all species. They only identify some plant species. So identification of plant species plays an crucial role for agricultural activities and medical applications. Identification of plant species is very important in every one's life. Plant species are nearer to 3,90,000 in all over the globe and new species can come across every year [1]. Plant species identification plays an important role in botanists, environmental science and grasping biodiversity. Conservatively, botanists classify the plant species by looking structure of leaf, color and texture. It is a time-consuming process with less accuracy by the human errors. Around 250,000 flowering plant species have been identified worldwide. Many blossoming flowers are frequently found in gardens, parks, by the side of the road, and in many other places. Only taxonomists or botanists are frequently able to identify flowers [2]. Aesthetic and Emotional Value: Flowers are essential for improving the surroundings and boosting human emotions. They can be a source of joy and inspiration for both everyday life and artistic endeavors. Clinical data and medical research have demonstrated the therapeutic benefit of many flowers, which has greatly aided in the creation of natural treatments and healthcare procedures. In the field of flower classification and recognition, the introduction of Convolutional Neural Networks (CNNs) have greatly improved the effectiveness and precision of floral identification [3]. India is capable of producing a wide range of horticultural products. Its environment's adaptability is the reason for this. Ninety percent of horticultural products are fruits and vegetables. Other horticultural product categories include crops, spices, flowers, and fragrant plants. Among all horticultural products, 314.65 million tons of fruits and vegetables are produced [4]. India exports 161 million USD worth of fruits and vegetables. As far as exporting fruits and vegetables goes, it came in at number 14 worldwide. The top five countries in terms of fruit and vegetable exports are the United States, the Netherlands, France, Germany, and China [4]. Vegetables and fruits provide protection against a number of diseases and prevent the onset of problems like diabetes, high blood pressure, and heart disease. Despite the fact that fruits and vegetables can easily be added to a good diet, many individuals have a tendency to disregard their health due to their busy lives. Poor food consumption is the main cause for non-communicable diseases which is also stated in World Health Organization (WHO) [5]. Machine learning algorithms have come into view in recent years to know the limitations and errors in the traditional methods of identifying of plant species. By examining large datasets and uncovered patterns, we use machine learning algorithm to identify plant species efficiently with better accuracy. This machine learning techniques extract the color, texture and high-resolution images to process the images to classify plant species. We also see the essential factors such as soil, climate and growth level of plants for providing better accuracy. Among all machine learning techniques, we use Convolutional Neural Network (CNN) by extracting of images for images processing automatically for image-based classification task. CNN helps in processing the image classification by extracting images like leaves, flowers, fruits and vegetables as an input. The CNN model was developed by TensorFlow and Keras frameworks is designed to process the images for getting high accuracy. This CNN model provides efficiency, good precision and scalability for classifying plant species. CNN also minimize the human error rate and classify the plant species accurately. The modern tools like CNN helps in classifying the plant species more effectively to make work easier for botanists, agricultural scientists and researchers. Additionally, it supports biodiversity conservation, improves agricultural decision-making and helps in better understanding of plant ecosystems. With the help of this study, machine learning is placed as a powerful tool to identify the challenges which occur in classification of plant species and provides environmental and agricultural sustainability goals.

2. Related Work

Traditionally, classification of plant species depends on manual classification by skilled botanists, focusing on physical characteristics like leaves, flowers, fruits, vegetables and stems. Recent improvements in computer vision and deep learning have automated this process there by, improving the efficiency and accuracy. variations in lighting, image quality, and natural differences in plant appearance. Multi-modal approaches, like a CNN-based architecture for classification of plant species demonstrate that combining features from multiple plant components such as leaves, flowers, fruits and vegetables provides better results than using a single type. The multi-model approach effectively addresses the challenges which occur in the identification of plant species. Earlier methods depend on the texture and color descriptors using KNN classifiers, which results to 90% accuracy [11]. However, these approaches are restricted to outdated datasets and absence of robustness in modern methods. The proposed system supports CNN and Streamlit, by using an updated dataset and accomplishing an accuracy of more than 95%.

Table 1. Related work

Reference Number	Year	Features	Classifier	Accuracy	Limitations	Improved System
[1]	2021	Shape, texture, and color	SVM	93.26 %	Used traditional methods	Used modern algorithms
[2]	2019	Shape, texture, and color	MLK-SVM	76.92%	Only 14 different features	More than 15 different features
[3]	2024	Color	CNN	95%	Only used dataset of flowers	Used leaves, vegetables, fruits and flowers
[4]	2021	Color, Texture	C4.5 and KNN	94.63 %	Dataset used for the experiment purpose is very old.	We took a new experimental model like CNN.

[6]	2020	Leaf veins	SVM	82.67 %	Tested only on few species	Used CNN model with streamlit framework. Achieved 96% accuracy with 52 species.
[7]	2023	Shape, color of leaf.	Tensor flow, Keras, Machine Learning, Convolution Neural Networks.	91%	Only applied for identifying leaf of a plant	Combination of leaf, fruit, vegetables and flower with more accuracy.
[8]	2019	Statistical and co-occurrence features derived from the Wavelet transformed sub-bands	SVM	94.3 %	Tested only on 2612 variety of images.	Tested on more than 10,000 images**.**
[9]	2018	Color, Shape and Texture	ANN	81.9 %	Partial data that include all flowers species with 20 images for each.	Full data include all flower species more than 20 images.
[10]	2021	Shape, texture, and color.	SVM and MLP	85.82 %	Less preprocessing techniques.	More preprocessing technique

						to improve accuracy.
[11]	2021	Stem, leaves, color	Resnet model	95%	Has less accuracy for the algorith m they used.	CNN, TensorFlo w has more accuracy level.
[12]	2020	Shape, texture, and color	SVM, PCA, K-Means	90%	Dataset cannot be lobed properly	Used CNN to increase dataset quality.
[13]	2023	Shape, color	CNN	94%	Sensitivi ty to imbalanc ed data, overfitti ng, and computa tional complex ity.	Data can be easy And understan dable

3. Proposed Model

The suggested model identifies plant species using a Convolutional Neural Network (CNN), which is well-known for gleaning complex information from picture data. With the help of Keras and TensorFlow, it can identify different plant species by analyzing high-resolution photos of leaves, flowers, fruits, and vegetables and analyzing characteristics like size and shape. To ensure resilience, the model takes into account environmental fluctuations and is trained on more than 10,000 labeled samples. The generalizability and quality of the images were enhanced using preprocessing methods such as data augmentation. Outperforming conventional procedures, it achieves an accuracy of above 95%. It is implemented with Streamlit since it offers an intuitive interface for real-time plant identification, promoting agricultural applications and biodiversity. Our contributions to the proposed model are:

1. We have tested over 52 plant species where as in [6] they have tested only on 6 plant species.
2. We used modern methods such as CNN, Tensorflow, Keras and Streamlit having accuracy of more than 95% where as in [1] they have only used traditional methods like SVM which is having less accuracy.
3. We have classified plants based on leaf, vegetable, fruit and flowers characteristics, where as in [7][23] they have classified the plants by taking only leaf characteristics.

4. We have tested on more than 10,000 images for the plant classification, where as in [8][24] they have tested on 2612 images only.
5. We have used the dataset that consists of combination on leaves, flowers, fruits and vegetables, where as in [3][22] they have used only the flower dataset.
6. The performance of the proposed model has achieved an accuracy of more than 95% in classification of the plant species with the help of the images of leaves, flowers, fruits and vegetables[25].

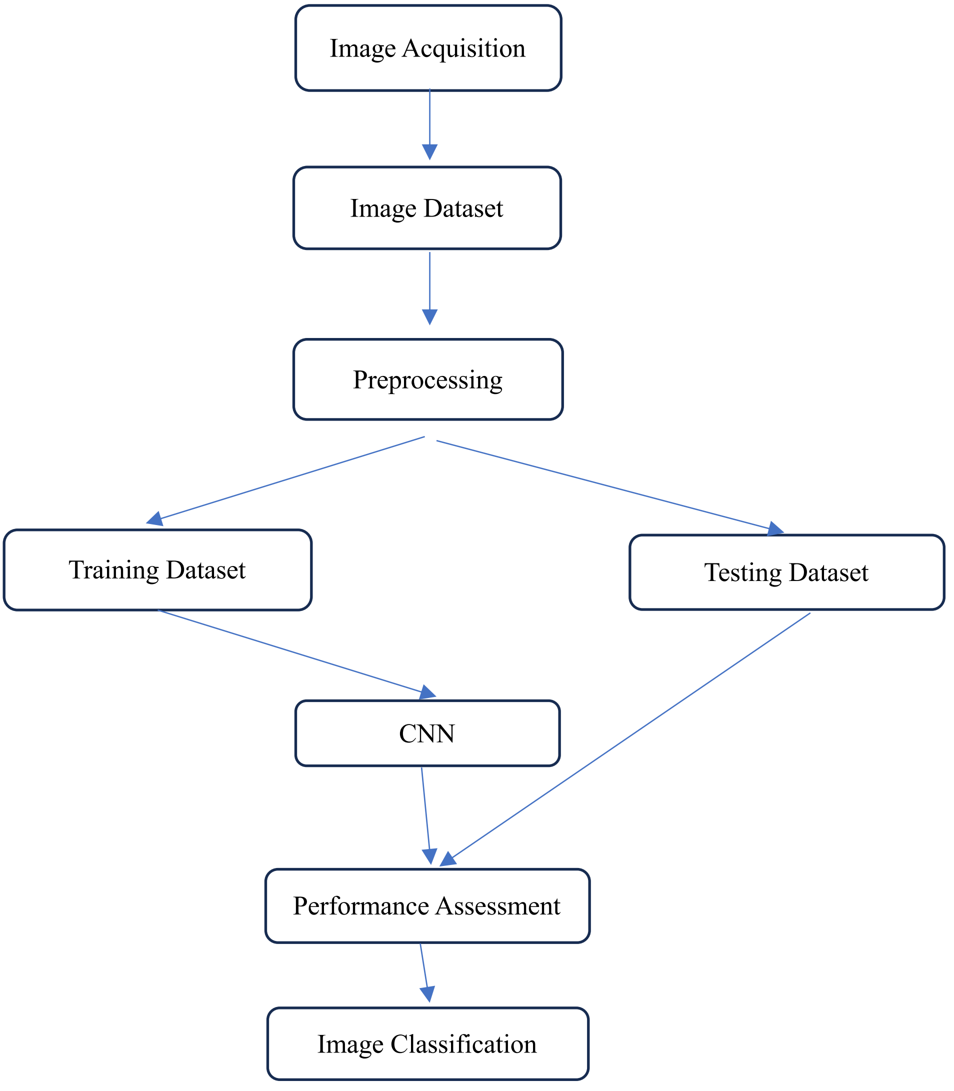

Fig.1. Algorithm

From the Fig. 1, the algorithm used in the proposed model is described as-
1. Image Acquisition: In this initial stage, pictures are collected from a variety of sources, including cameras, datasets, and internet sources. The model uses these pictures as input.

2. Image dataset: This could mean "Image Data Preparation" or "Image Data Retrieval." In this step, the photos for the classification assignment are gathered, arranged, and maybe labelled.

3. Preprocessing: This step entails preprocessing the obtained image data to ensure that it is in the right format for the model. Typical preprocessing steps include resizing images to a uniform size. Scaling pixel values from 0 to 1 or -1 to 1 normalizes them. Augmentation techniques like flipping, zooming, and rotating are employed to make the dataset appear larger than it actually is. The training dataset is the subset of images used to train the model. The CNN's labeled images allow it to learn traits that relate to the targetlabels.

4. Testing Dataset: The model's performance is assessed using a different testing dataset following training. To guarantee an objective evaluation of the model's capacity for generalization, dataset shouldn't overlap with the training dataset.

5. CNN (Convolutional Neural Network): The picture classification task is carried out using a CNN. After extracting the features from images using convolutional layers, it classifies the images using fully connected layers.

The challenges involved in classification of plant species; Convolutional Neural Networks (CNNs) exhibits a remarkable efficiency. CNNs are deep learning models that are ideal for image classification tasks like recognising plant species from photographs since they are made to automatically understand spatial hierarchies in picture data. CNN consist of-

Convolutional Layers: on identify characteristics like edges, textures, and colours, these layers apply filters on the images. When recognising plant species by their leaf form, texture, or colour patterns, these characteristics are essential.

Pooling Layers: These make the model more effective by reducing the spatial dimensions of the feature maps while keeping crucial information.

Fully Connected Layers: To generate predictions, the network flattens the data after convolution and pooling, then sends it via fully connected layers.

Output Layer: For classification tasks, this layer usually employs a softmax function, in which the model produces a probability distribution across various plant species.

6. Performance Evaluation: Metrics including accuracy, precision, recall, F1 score, etc. on the testing dataset are used to evaluate the model's performance once it has been trained. This stage aids in assessing the model's ability to generalise to fresh, untested images.

7. Image Classification: The last step involves classifying fresh, unseen images using the trained CNN. The model gives the image a class label based on the learnt features.

4. Results and Discussion

1. Model Performance: The Convolutional Neural Network (CNN) architecture implemented in **TensorFlow** and **Keras** demonstrated high efficiency in classifying plant species based on their visual characteristics (leaves, flowers, and fruits/vegetables). The model was trained fine-tuned with machine learning, and achieved 95% accuracy of the classifications.

2. Dataset and Preprocessing: The dataset comprised **52 classes** of plant species, with 70% used for training, 20% for validation, and 10% for testing. The leaf images are taken from [14] which consists of 8 classes of plant species namely Apple, Berry, Fig, etc., flower images are taken from [15] which consists of 9 classes namely Rose, Daisy, Lotus, etc., fruit and vegetable images are taken from [16] which consists of 36 classes of plant species namely Kiwi, Orange, Carrot, Cauliflower, etc. Images are normalized for optimal CNN performance. Data augmentation techniques, such as rotation, flipping, and zooming, were applied to enhance the dataset's diversity.

3. Visualization of Model Learning: The model's learning and decision-making process were interpreted by analyzing the feature maps generated within the CNN layers. These analyses revealed that the model focuses on critical features such as leaf veins, petal shapes, and fruit textures for classification [17][18].
4. Streamlit Web Application: A user-friendly interface was developed using **Streamlit** to allow users to upload images for real-time classification. The application provides-the predicted plant species with confidence percentages and some details about the image. Additional information, such as the **scientific name** and **flowering season**, etc. retrieved from a pre-defined database.
5. Performance Metrics
The tables given below summarizes the classification performance for each plant type.

Table 2: Performance metrics

Class	Precision		Recall		F1-Score	
	CNN	SVM	CNN	SVM	CNN	SVM
Leaf	0.93	0.84	0.92	0.83	0.92	0.83
Flower	0.96	0.72	0.95	0.72	0.95	0.72
Fruit	0.91	0.87	0.90	0.85	0.90	0.75
Vegetable	0.91	0.52	0.90	0.52	0.90	0.43

The Table 2 gives the performance metrics of Classification using CNN model with Streamlit and Support Vector Machines (SVM) [6] [2] [20] [21].

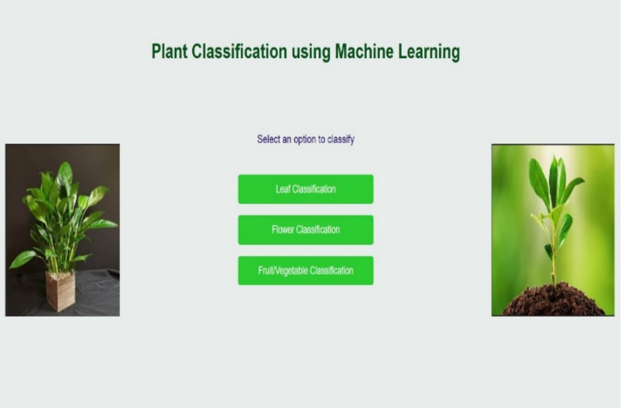

Fig. 2. Main Page

The Fig. 2 represents the main page of the Plant Classification. Here, we can select any option given such as, Leaf Classification, Flower Classification, Fruit/Vegetable Classification which helps to identify the plants with the help of leaf, flower, fruit and vegetable images.

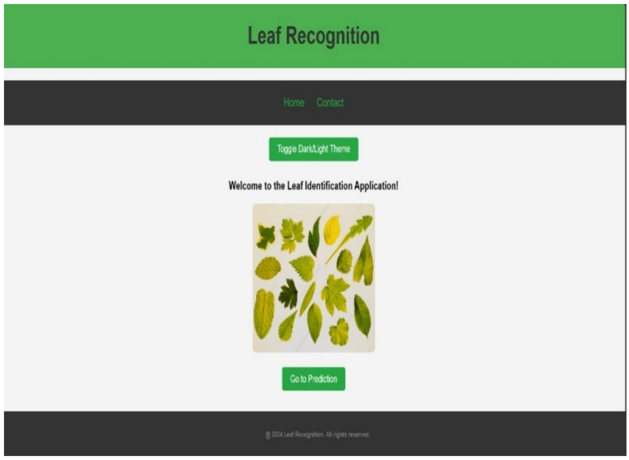

Fig.3. Leaf Recognition

If we select Plant Classification from Fig.2, it redirects to the Leaf Recognition as shown in Fig.3. We can click on "Go to Prediction" which helps in identifying the plants with the leaf images.

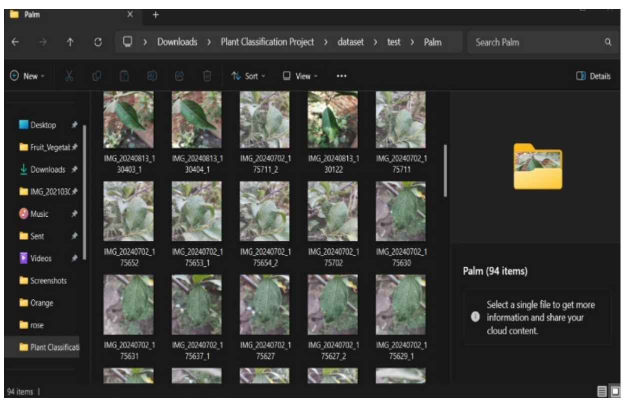

Fig.4. Leaf Dataset

The Fig.4 consists of the images of leaves which are used for predicting the plants.

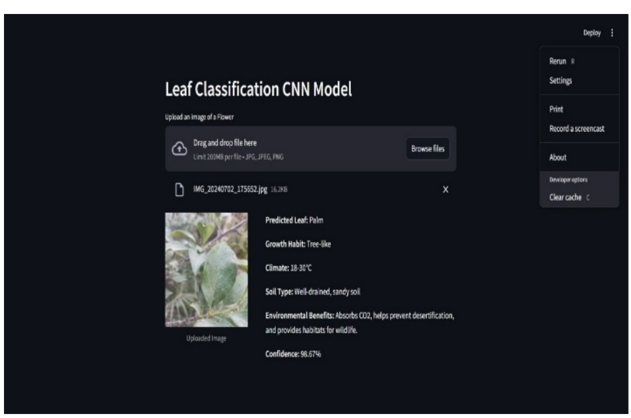

Fig.5. Leaf Classification

If we click on "Go to Prediction" in Fig. 3, it redirects to Leaf Classification as shown in Fig. 5 where we can upload the images of leaves which helps in identifying the plants.

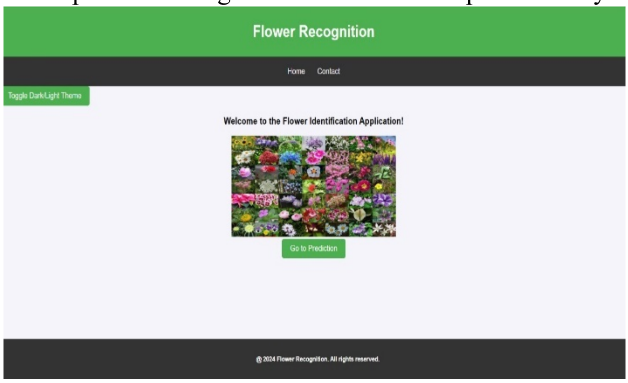

Fig.6. Flower Recognition

If we select Flower Classification from Fig.2, it redirects to the Flower Recognition as shown in Fig. 6. We can click on "Go to Prediction" which helps in identifying the plants with the flower images.

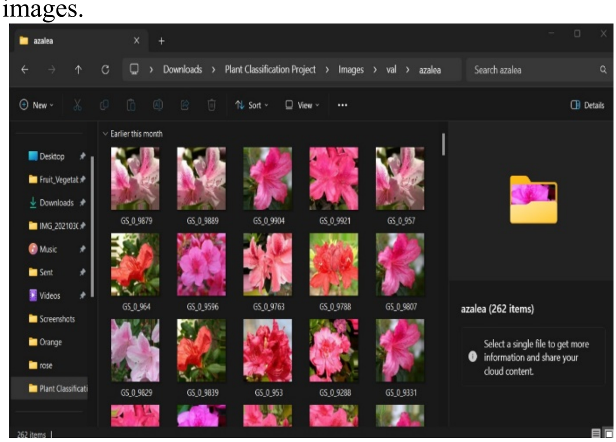

Fig.7. Flower Dataset

The Fig.7 consists of the images of flowers which are used for predicting the plants.

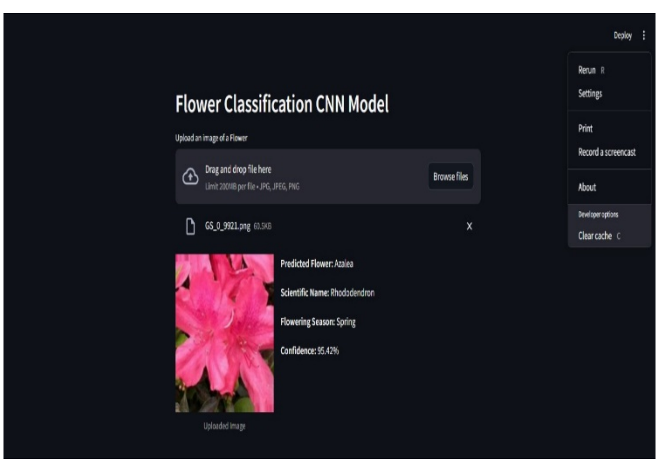

Fig.8. Flower Classification

If we click on "Go to Prediction" in Fig. 6, it redirects to Flower Classification as shown in Fig. 8 where we can upload the images of flowers which helps in identifying the plants.

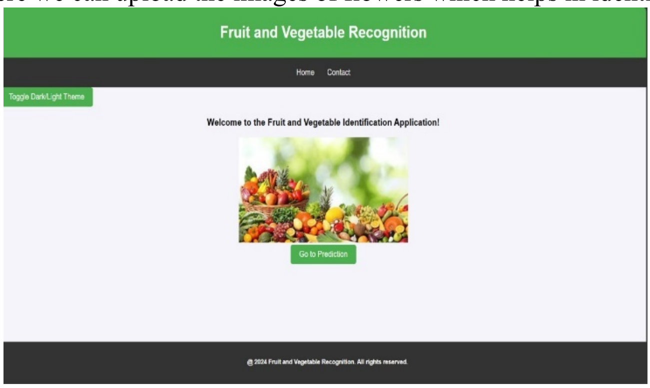

Fig.9. Fruit and Vegetable Recognition

If we select Fruit/Vegetable Classification from Fig. 2, it redirects to the Fruit/Vegetable Recognition as shown in Fig. 9. We can click on "Go to Prediction" which helps in identifying the plants with the fruit/vegetable images.

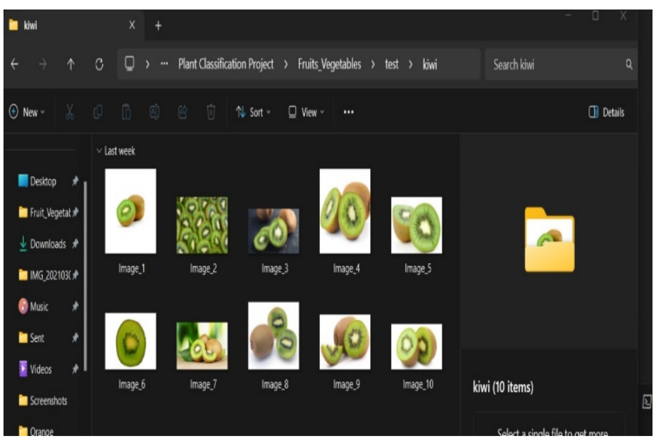

Fig.10. Fruit Dataset

The Fig.10 consists of the images of fruits which are used for predicting the plants.

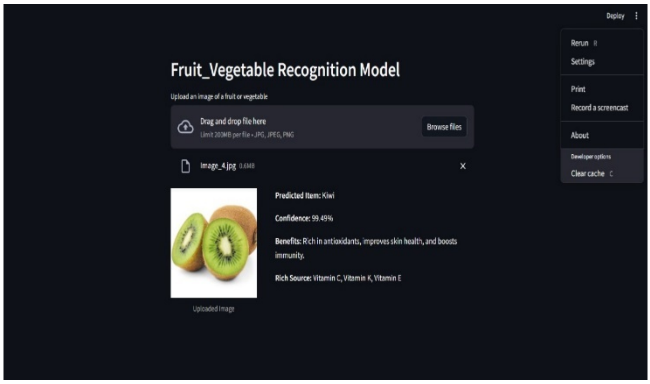

Fig.11. Fruit Recognition

If we click on "Go to Prediction" in Fig. 9, it redirects to Fruit_Vegetable Recognition as shown in Fig. 11 where we can upload the images of fruits which helps in identifying the plants.

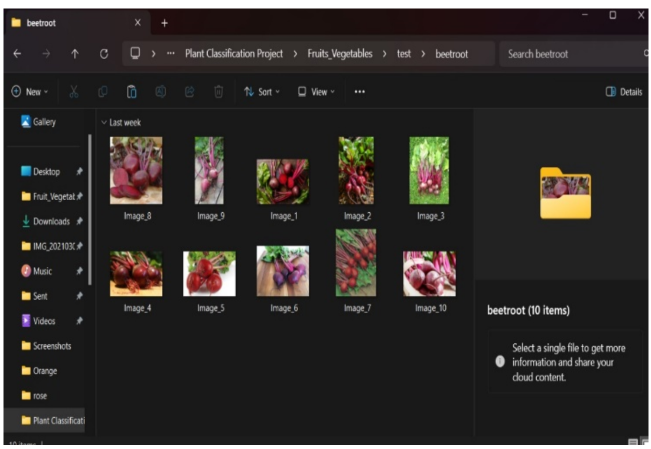

Fig.12. Vegetable Dataset

The Fig.12 consists of the images of vegetables which are used for predicting the plants.

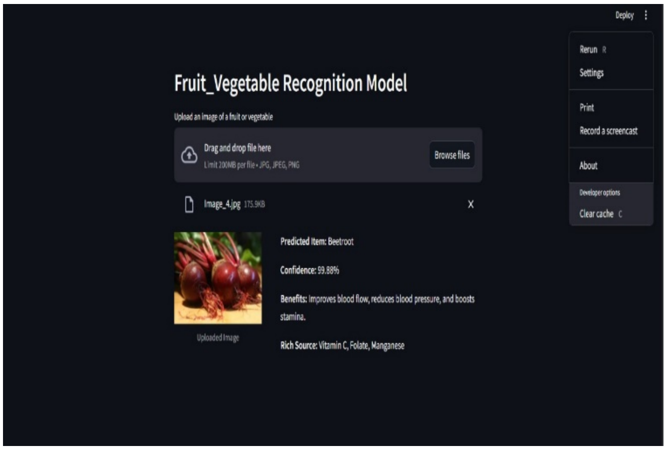

Fig.13. Vegetable Recognition

If we click on "Go to Prediction" in Fig. 9, it redirects to Fruit_Vegetable Recognition as shown in Fig. 13 where we can upload the images of vegetables which helps in identifying the plants.

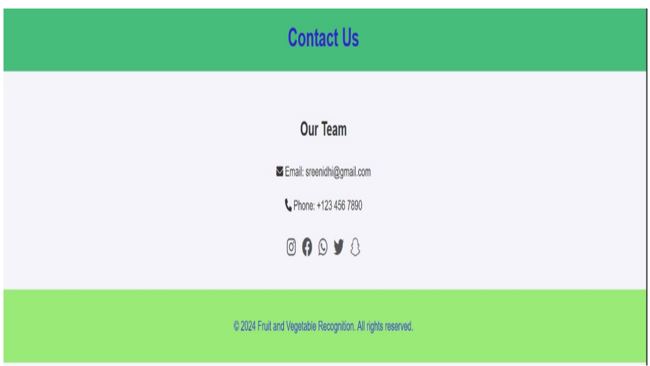

Fig.14. Contact

If we click on "Contact" in Fig. 3, Fig. 6, Fig. 9, it redirects to the Contact Us page as shown in Fig. 14.

5. Conclusion

The study of plant species identification provides vast learning and exhibit the Machine learning techniques more potentially. Proposed model uses modern techniques like Convolutional Neural Network (CNN) which takes features of plant like leaf, vegetables, flowers and fruits and characteristics such as color, shape and size. The proposed system also helps to focus on main characteristics such as soil and climate type which enhance more accuracy of identifying plant species. The Convolutional Neural Network (CNN) gives us more flexibility and accuracy of identifying plants based on performing on characteristic tasks. In this model we use over 10,000 datasets of plants and achieved more than 95% accuracy level with good precision of identifying of plant species using Convolutional Neural Network (CNN). The proposed model as better accuracy when compared to the traditional techniques which was labor work which has less efficiency in identifying plant species. And we used Tensorflow, keras framework which identifies the plant species based on the real-time characteristics of identifying plant species. This model techniques helps to identify plant species more accurately which helps and support for Botanists and Ecologists. Performance metrics using Convolutional Neural Network (CNN) integrated with Sreamlit which have higher precision, recall, and F1 score, in Table-2, compared to the Support Vector Machine (SVM) model. This feature helps in highlighting the identification of plant species which uses Convolutional Neural Network (CNN). Streamlit is used very much in identifying images and gives accurate results. Additionally, it is an application for very botanist, ecologists and for agricultural purpose. It provides more accuracy and efficiency of identification of plant species which uses modern techniques rather than the traditional methods like labor work and tools. The Convolutional Neural Network (CNN) provides more accuracy and efficiency of plant species identification.

6. Future Scope

Future work for the proposed which includes improving the datasets with more geographical plants and endangered in more species datasets with capturing more features and characteristics of plant species which enhance robustness of identifying plant species. Also, we can use more modern techniques which gives more accuracy and efficiency to identify plant species. The model also may misclassification of species which has same features between hibiscus and rose sometimes. Environmental Factors such as soil, climate and conditions also lead to edge cluttering of identifying plant species which leads to the accuracy. Sometimes class imbalance in the datasets which may have less sample features may affect the precision of datasets of the classes. Developing an application on devices like mobile in real time helps more for identifying the plant species for botanists and ecologists. Continuous learning of data may adopt to retraining datasets when they are improved. Integration with Geographic Information System (GIS) tools may support biodiversity mapping and conservation, applications in precision may help to botanists and agricultures. As well, crowdsourced data collection, Augmented Reality (AR)-based educational tools also may further improves scalability, accessibility and flexibility in ecologists, botanists, agriculture and biodiversity of plant species identification.

References

1. Plant Species Identification based on Plant Leaf Using Computer Vision and Machine Learning Techniques, Kaur Surleen, Kaur Prabhpreet, J Multimed Inf Syst 2019;6(2):49-60, https://doi.org/10.33851/JMIS.2019.6.2.49
2. Patel, Sanskruti & Patel, Isha. (2019). Flower Identification and Classification using Computer Vision and Machine Learning Techniques. International Journal of Engineering and Advanced Technology. 8. 277-285. 10.35940/ijeat.E7555.088619.

3. Li, Y.; Lv, Y.; Ding, Y.; Zhu, H.; Gao, H.; Zheng, L. Research on a Flower Recognition Method Based on Masked Autoencoders. Horticulturae **2024**, 10, 517. https://doi.org/10.3390/horticulturae10050517
4. Tripathi, Mukesh & Maktedar, Dhananjay. (2021). Detection of various categories of fruits and vegetables through various descriptors using machine learning techniques. International Journal of Computational Intelligence Studies. 10. 36. 10.1504/IJCISTUDIES.2021.10036519.
5. Appadoo, Anish & Gopaul, Yashna & Pudaruth, Sameerchand. (2023). FruVegy: An Android App for the Automatic Identification of Fruits and Vegetables using Computer Vision and Machine Learning. International Journal of Computing and Digital Systems. 13. 169-178. 10.12785/ijcds/130114.
6. Ambarwari, Agus & Adrian, Qadhli Jafar & Herdiyeni, Yeni & Hermadi, Irman. (2020). Plant species identification based on leaf venation features using SVM. TELKOMNIKA Indonesian Journal of Electrical Engineering. 18. 726-732. 10.12928/TELKOMNIKA.v18i2.14062.
7. H, Soumya & Joshi, Sampada & N, Hemanth. (2024). Review on Poisonous Plants Detection Using Machine Learning. International Journal of Advanced Research in Science, Communication and Technology. 186-195. 10.48175/IJARSCT-15328.
8. Sanchez Galan F, Javier & Henry-Royo, Anel & Jo, Kang-Hyun & Caceres-Hernandez, Danilo. (2022). Automatic Feature Detection and Classification for Watermelon (Citrillus lanatus). 1-7. 10.1109/IWIS56333.2022.9920868.
9. Almogdady, Huthaifa & Manaseer, Saher & Hiary, Hazem. (2018). A Flower Recognition System Based On Image Processing And Neural Networks. International Journal of Scientific and Technology Research. 7.
10. L. S, J. S and P. B. R, "Optimized Plant Species Classification through MobileNet-Enhanced Hybrid Models," 2024 5th International Conference for Emerging Technology (INCET), Belgaum, India, 2024, pp. 1-6, doi: 10.1109/INCET61516.2024.10593020.
11. N. Hemalatha, P. Sukhetha and R. Sukumar, "Classification of Fruits and Vegetables using Machine and Deep Learning Approach," 2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT), Pune, India, 2022, pp. 1-4, doi: 10.1109/TQCEBT54229.2022.10041654.
12. Swu, Vikaho & Kharir, Ibaphyrnaishisha & Bora, Dibya. (2020). Identification of Different Plants through Image Processing Using Different Machine Learning Algorithms. 43. 172-179.
13. Nguyễn, Nhân & Le, Van & Le, Thi & Hai, Vu & Pantuwong, Natapon & Yagi, Yasushi. (2016). Flower species identification using deep convolutional neural networks.
14. https://www.kaggle.com/datasets/mahmoudshaheen1134/plant-leaf-image-dataset
15. https://www.kaggle.com/datasets/sammymilitante/flowerclass
16. https://www.kaggle.com/datasets/kritikseth/fruit-and-vegetable-image-recognition
17. Nguyễn, Nhân & Le, Van & Le, Thi & Hai, Vu & Pantuwong, Natapon & Yagi, Yasushi. (2016). Flower species identification using deep convolutional neural networks.
18. Jianghao Chen, Yiming Huo, Junyu Li. Recognition and classification of flower species based on artificial intelligence. Academic Journal of Computing &

Information Science (2021), Vol. 4, Issue 8: 78-82. https://doi.org/10.25236/AJCIS.2021.040815.
19. Yao, Siyang. (2024). Machine Learning in Agriculture: KNN-Based Classification of Fruits and Vegetables. Science and Technology of Engineering, Chemistry and Environmental Protection. 1. 10.61173/tszzmw78.
20. Behera, Santi & Rath, Amiya & Sethy, Prabira. (2020). Fruit Recognition using Support Vector Machine based on Deep Features. Karbala International Journal of Modern Science. 6. 16. 10.33640/2405-609X.1675.
21. Waluyo Poetro, B. S., Maria ., Zein, H., Najwaini, E., & Zulfikar, D. H. (2024). Advancements in Agricultural Automation: SVM Classifier with Hu Moments for Vegetable Identification. Indonesian Journal of Data and Science, 5(1), 15-22. https://doi.org/10.56705/ijodas.v5i1.123
22. Gulzar, Yonis, et al. "Adaptability of deep learning: datasets and strategies in fruit classification." BIO Web of Conferences. Vol. 85. EDP Sciences, 2024.
23. Mulugeta, A. K., Sharma, D. P., & Mesfin, A. H. (2024). Deep learning for medicinal plant species classification and recognition: a systematic review. Frontiers in Plant Science, 14, 1286088.
24. Sajitha, P., et al. "A review on machine learning and deep learning image-based plant disease classification for industrial farming systems." Journal of Industrial Information Integration (2024): 100572.
25. Wang, W., Zhu, A., Wei, H., & Yu, L. (2024). A novel method for vegetable and fruit classification based on using diffusion maps and machine learning. Current Research in Food Science, 8, 100737.