iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44997

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

CNN-Based Machine Learning Model to Identify the Different Variety of Hibiscus Rosa-Sinensis Species Using Plant Leaf

Jeetumoni Barman[1], Nayana Das[2], Sumi Das[3], Diganta Barman[4]
[1, 2]*Department of Computer Application, Girijananda Chowdhury Institute of Management and Technology, Guwahati, Assam, India*
[3]*Department of Botany, Dakshin Kamrup College, Mirza, Assam, India*
[4]*State Programmer, Rashtriya Madhyamik Shiksha Abhiyan, Assam, India*

Abstract: ***Plants are one of the most essential living organisms in our planet. Plants are useful for human beings in many different ways. Plants are beneficial both for ornamental and medicinal purpose. So, identifying the correct variety of plant belonging to the same species is very important. Botanist or experts in botanical knowledge can accomplish the task very easily. But everyone is not a botanist nor can acquire expert botanical knowledge to identify them accurately. So, an alternative approach is needed to identify the plant variety. Leaves play an important role in their identification among the six basic parts of plants viz. roots, stems, leaves, flowers, fruits, and seeds. Among all the plant species available, Hibiscus rosa-sinensis (commonly known as China rose) is a plant which proves to be beneficial both as an ornamental and medicinal plant and are available in different varieties. But identifying the correct variety of Hibiscus rosa-sinensis with naked eye before the flower blooms on it is not an easy task. In this paper, we propose a CNN based Machine learning model to identify the different variety of Hibiscus rosa-sinensis species using plant leaf***

Keywords: Convolutional Neural Network (CNN), Machine Learning, Hibiscus rosa-sinensis, leaf data-set, Angiosperms, Digital Morphological Feature (DMF)

I. INTRODUCTION

Plants play a vital role in our day-to-day life for their medicinal and ornamental benefits. The most fascinating part of a plant is a flower. But for a common man without much botanical knowledge cannot identify a plant of same species and guess the colour of the flower it blooms until flowers bloom on the plant. *Hibiscus rosa-sinensis* is a plant species which when blooms, gives flowers of different variety. To identify a plant which can blooms a desired variety of flower, a machine learning model is proposed which will take plant leaves as input and predict the flower it can bloom.

In recent years, CNNs has dominantly occupied its position in the field of computer vision. The reason for its popularity is the availability of efficient and Graphical Processing Units and availability of sufficient image data for training the CNN based model. Figure 1 depict a comparative analysis, how a leaf image is analyzed by a CNN model and by a botanist

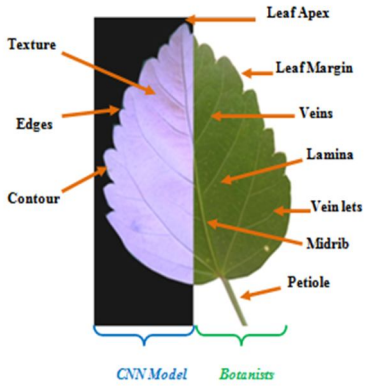


Fig. 1: Analysis of a Leaf by CNN model Vs Botanists

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

II. LITERATURE REVIEW

Flowers play a vital role of any plant mainly because of its ornamental values. Some flowers are useful for medicinal purpose and thus recognition of flowers is boon for industries like pharmaceutical and cosmetics. Thus, identification of plant species by introducing machine learning approach becomes necessary. Parvathy S N et al. [1] proposed a Convolution Neural Network (CNN) model to identify a flower by feeding an image of flower to be recognized using mobile camera.

The genus *Hibiscus* is in the family Malvaceae in the major group Angiosperms. The statistics of genus *Hibiscus* Species has 154 accepted species which is published in The Plant List. [2] *Hibiscus rosa-sinensis* is one of the accepted species of genus *Hibiscus* and the proposed paper use to identify different variety of *Hibiscus rosa-sinensis* plant.

Pallavi Shetty et al. [3] proposed a CNN based approach to identify distinct species of *Hibiscus* Plant namely *Hibiscus rosa sinesis*, *Hibiscus sabdarifaa*, *Hibiscus mutabilis*, *Hibiscus schizopetalus*, *Hibiscus syriacus*, *Hibiscus trionum*, *Hibiscus esculentus* using 500 distinct leaf image of *Hibiscus*, which gives an accuracy of 0.920 and loss value 0.1905

ArunPriya C. et al. [4] proposed an approach involving three phase viz. preprocessing phase, features extraction phase and classification phase which involves transforming the image into gray scale and boundary enhancement derives the common Digital Morphological Feature (DMF) from five fundamental features and Support Vector Machine (SVM) classification for efficient leaf recognition.

Plants play a vital role in the field of medicine. So, identification of plant species is very important. Adams Begue et al. [5] proposed a machine learning techniques for automatic recognition of medicinal plants using leaves from twenty-four different medicinal plant species.

Sivaranjani, C. et al. [6] proposed a machine learning model for real-time identification of medicinal plants species by the analysis of leaf images obtained directly from their habitat and irrespective of lighting conditions.

III. PROPOSED METHOD

The proposed model for identifying different varieties of *Hibiscus rosa-sinensis* is given in steps below:

A. Step I: Image acquisition:

This step involves the collection of leaf images of all the plants of different varieties of *Hibiscus rosa-sinensis* that is to be trained and which later can be used to identify the different varieties of *Hibiscus rosa-sinensis* species based on the knowledge acquired during the training phase. J. Wäldchen et al. [7] suggest that the images that is acquired in datasets can be categorized into three categories namely *scans*, *pseudo-scans* and *photos*. The leaf images obtained through scanning and photography in front of a simple background comes in *scans* and *pseudo-scans* categories respectively while photographs captured on natural backgrounds falls under *photo* category. The dataset used in the proposed model falls under the photo category as the leaf photographs are captured in natural environment. A dataset of five different varieties of *Hibiscus rosa-sinensis* is acquired. The dataset contain above 500 images of *Hibiscus rosa-sinensis* leaves. Each variety of *Hibiscus rosa-sinensis* contain above 100 images.

B. Step II: Image-preprocessing:

The images collected in step I are may inherently contain noise and which may lead to lower accuracy rate and thus may hamper the identification process. So, the images are pre-processed to enhance the quality of the images for future processing. During this step, the leaf images are rescaled, converted grayscale and augmented.

C. Step III: Labeling of Images:

Different labels are generated for each variety of *Hibiscus rosa-sinensis* in the dataset that contain the detail information about different varieties of *Hibiscus rosa-sinensis* as the red variety, white variety, light pink variety, orange variety, dark pink variety.

D. Step IV: Training Phase:

A CNN based model has been used to train the dataset. CNN is a model of Deep Learning that is used to classify the features from the input image. A CNN Architecture is shown in fig. 2. A CNN model has a number of layers which includes- Input Layer, Convolutional Layer, ReLU Activation Layer, Pooling Layer, Flatten Layer, Full Connected Layer, Soft-Max Layer and Output Layer.

1) Input Layer: This layer contain the leaf image in RGB of size 180 x 180 pixel in jpg format

©IJRASET: All Rights are Reserved

2) *Convolutional Layer:* This layer is the most significant layer of CNN where feature maps are created by applying numbers of filters to a training dataset. In this layer, a sequence of mathematical operation is performed and thus extracts the feature map of the original image. The output of the mathematical operation is computed by taking the scalar product of image variables with the filter or kernel values
3) *ReLU Activation Layer:* The activation layer consists of activation functions like ReLU, Softmax etc. Rectified Linear Unit (ReLU) is a non-linear activation layer, which replaces negative numbers of the filtered images with zeros. Mathematically it can defined as

$$f(x) = \max(0, x)$$

4) *Pooling Layer:* The pooling layer reduces the number of computational load and pass it to next layer for further processing. The max pooling operation is performed where a two-dimensional filter is slides over each channel of feature map and select the maximum element from the region of feature map. The pooling layer is normally placed between convolution layers
5) *Full Connected Layer:* The fully connected layers take the output from other layer and flattened it into vectors and pass it to next layer. The fully connected layer applies weights to predict the correct label.
6) *Output Layer:* This is the last layer, which gives the probabilistic results are given for each label of the leaf dataset and the label with highest probabilistic value gives the class. Softmax activation function mainly deals with output layer of the classifier. It determines the probabilistic outputs for each class of labels between 0 and 1, It basically determines the presence of certain object in a given image and if the object is found, then the probabilistic value is set to '1' otherwise set to '0'. Mathematically it can be defined as

$$softmax\,(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Where, z_i = *is the input vector*
e^{z_i} = *standard exponential function for input vector*
K= number of classes
e^{z_j} = *standard exponential function for output vector.*

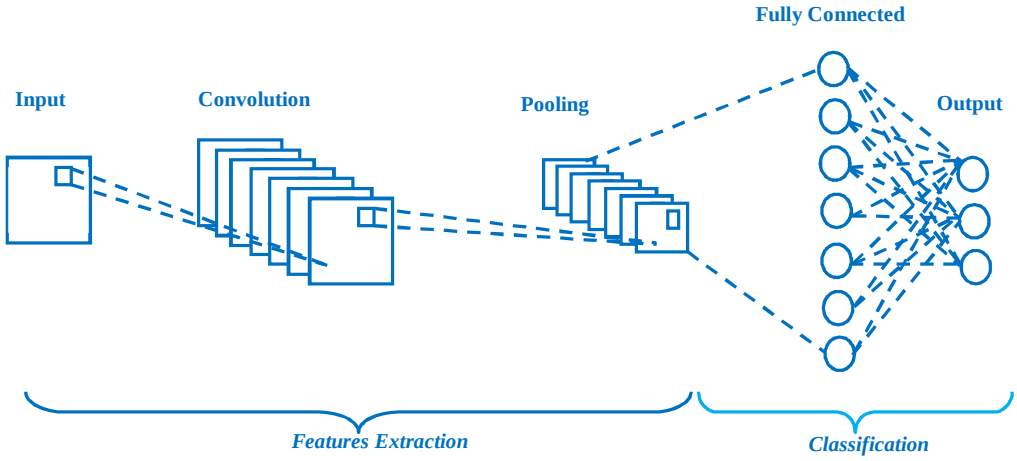


Fig. 2: Convolution Neural Network

E. Step V: Validation Phase

In validation phase, the model is fine tuned using its weights. The loss function used here is categorical cross entropy and optimizer used is Adam

F. Step VI: Prediction Phase

After the validation phase, the model is ready to take images for prediction. In Prediction phase, the model takes input of a leaf image of *Hibiscus rosa-sinensis* variety and predicts the class to which the leaf belongs

IV. RESULT AND DISCUSSION

The proposed model is tested on a sample dataset of nearly 600 images of *Hibiscus rosa-sinensis* leaves. The dataset includes five classes of *Hibiscus rosa-sinensis* (red variety, white variety, light pink variety, orange variety, dark pink variety).Each variety of *Hibiscus rosa-sinensis* contain more than 100 images. The dataset created has been divided into two parts. 80% of the dataset are used for training the model and 20% of the dataset are used for testing. During training the model, the leaf dataset is passed through CNN model with batch size 10 and epochs 20 which gives a result of accuracy of 91.85 %. The graph plot of training and validation accuracy is shown in Fig. 3 and graph plot of training and validation loss is shown in Fig. 4

V. CONCLUSION

This paper present a CNN based model for identification of five different variety of *Hibiscus rosa-sinensis* namely the red variety, the white variety, the pink variety, the orange variety and the dark pink variety using the respective leaf images. A dataset of more than 600 images of *Hibiscus* rosa-sinensis leaves has been used to train the model. Each variety of *Hibiscus rosa-sinensis* contains more than 100 leaf images. The proposed model is considered to be valid as gives an accuracy result of more than 90%. The model can further enhanced by training the model with leaf images of more different varieties of *Hibiscus rosa-sinensis* leaf which was not considered in the proposed model.

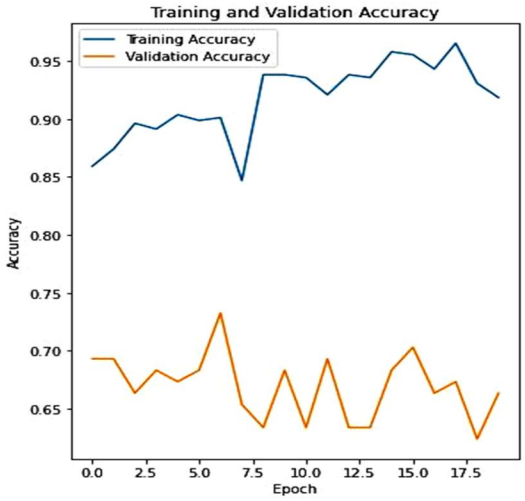


Fig 3: Training and Validation Accuracy Graph

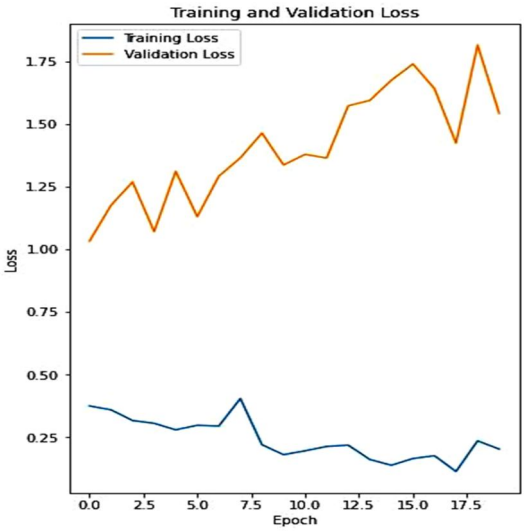


Fig. 4: Training and Validation Loss Graph

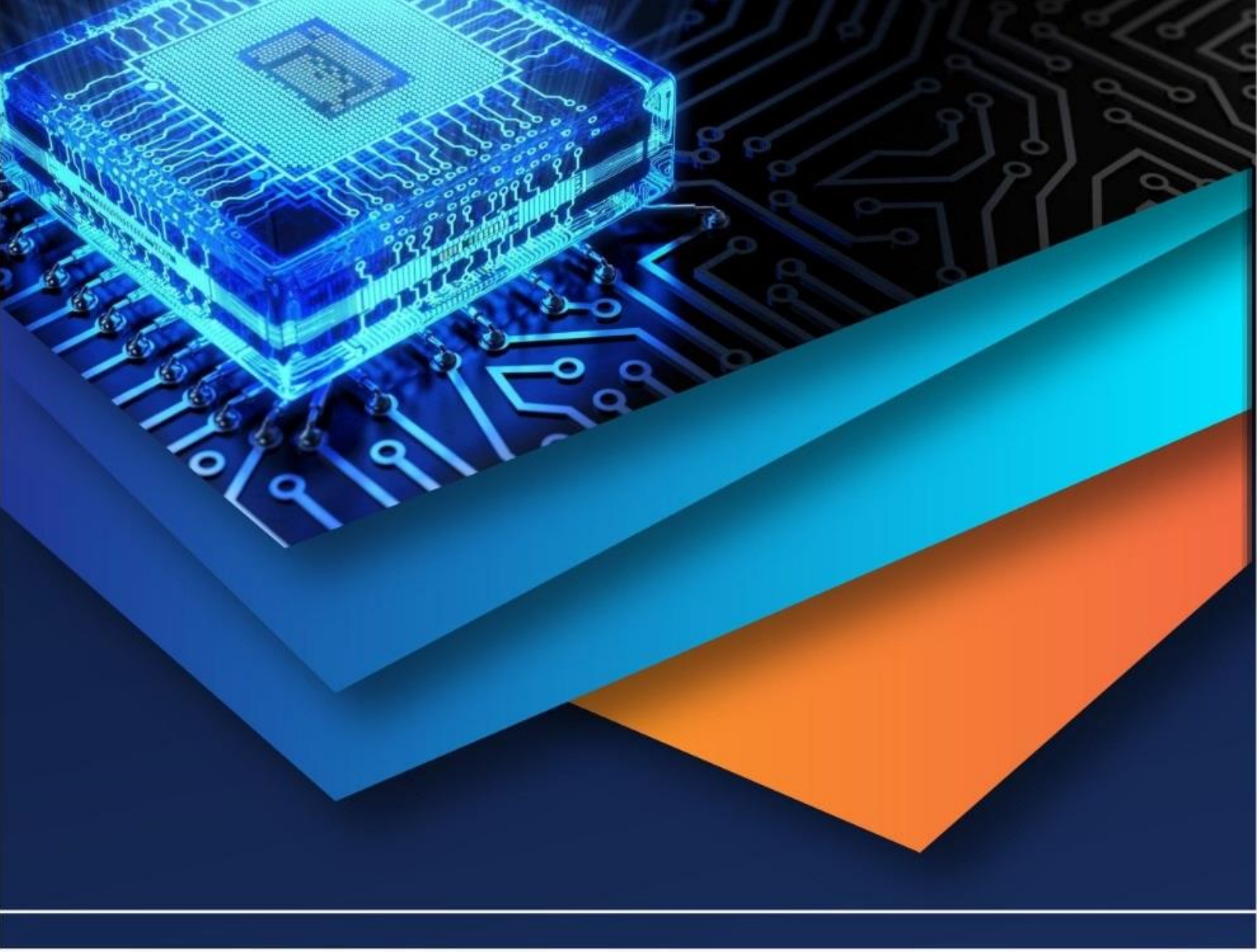
VI. CONFLICT OF INTEREST

All the authors have declared that no conflict of interest exists

REFERENCES

[1] Parvathy S N, N. Vrinda Rao. (2020). “Flower Recognition System Using CNN”, International Research Journal of Engineering and Technology (IRJET), Volume: 07 (Issue: 06), Pages 6609-6611.
[2] The Plant List (2010). Version 1. Published on the Internet; http://www.theplantlist.org/ (accessed 1st January)
[3] PALLAVI SHETTY, D. B. (2021). CNN Based Approach to Identify Hibiscus Plant Species. Iconic Research and Engineering Journals (IRE Journals) , Volume: 04 (Issue: 12), Page 22 -26
[4] ArunPriya C., Balasaravanan T., Antony Selvadoss Thanamani, “An Efficient Leaf Recognition Algorithm for Plant Classification Using Support Vector Machine”, Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, 2012, pp. 428-432.
[5] Sivaranjani, C. et al. (2019) Real-time identification of medicinal plants using machine learning techniques", in ICCIDS 2019 - 2nd International Conference on Computational Intelligence in Data Science, Proceedings. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ICCIDS.2019.8862126
[6] J. Wäldchen, M. Rzanny, M. Seeland, P. Mäder, "Automated plant species identification—Trends and future directions," PLoS computational biology, vol.14, no. 4, pp. 1-19, 2018

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)