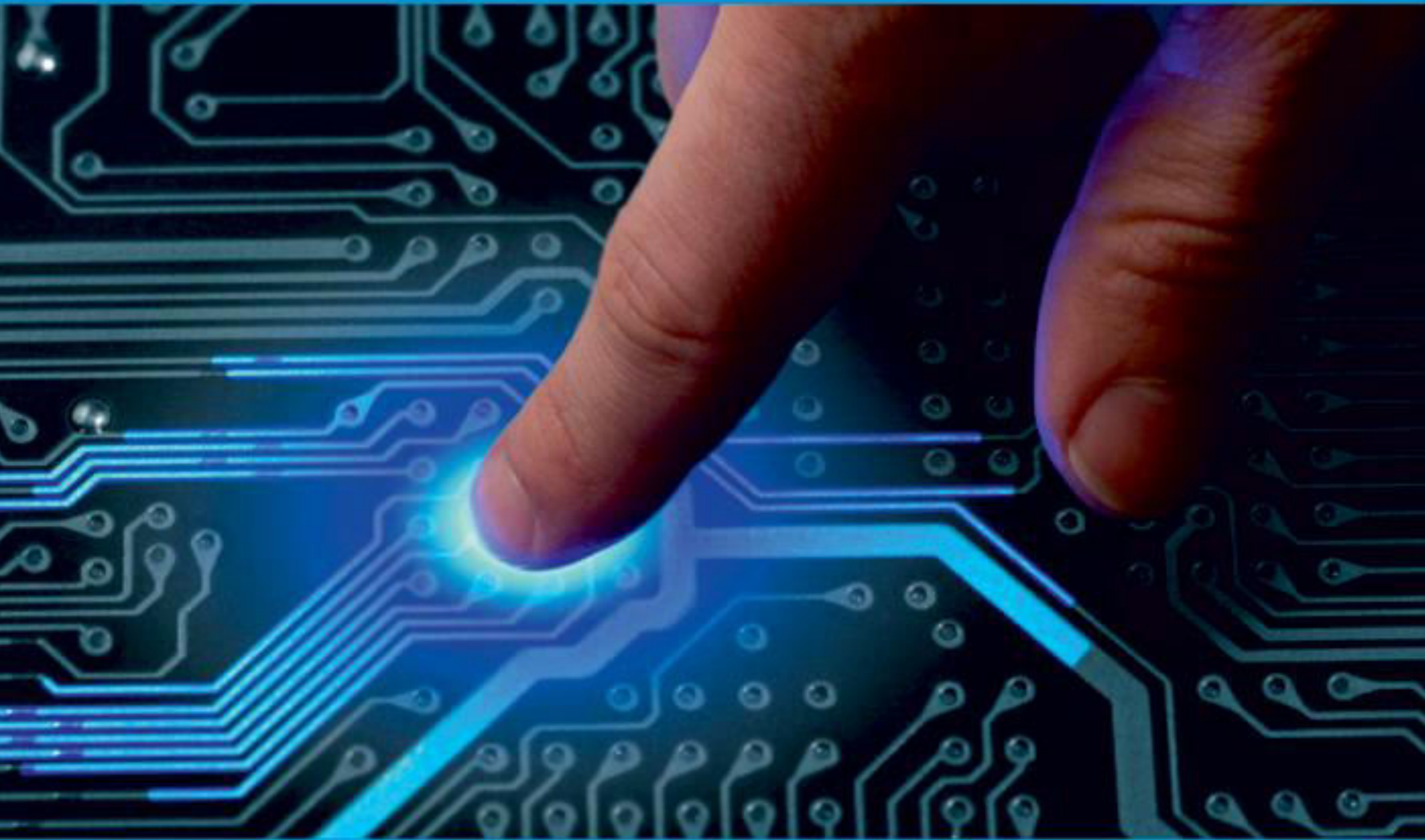
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006108 |

Plant Species Identification Using Machine Learning Techniques

Ms.S.Ranjani M.Tech, Musharaf Hussain.A, Santoshkumar.R, Thamaraikani.J

Assistant Professor, Dept. of Computer Science Engineering, Saranathan College of Engineering, Trichy, India

UG Student, Dept. of Computer Science Engineering, Saranathan College of Engineering, Trichy, India

ABSTRACT: -Plants are very crucial for life on Earth. There is a wide variety of plant species available, and the number is increasing every year. Species knowledge is a necessity of various groups of society like foresters, farmers, environmentalists, educators for different work areas. This makes species identification an interdisciplinary interest. This, however, requires expert knowledge and becomes a tedious and challenging task for the non-experts who have very little or no knowledge about the typical botanical terms. However, the advancements in the fields of machine learning and computer vision can help make this task comparatively easier. There is still no system developed that can identify all the plant species, but some efforts have been made. In this study, we also have made such an attempt. Plant identification usually involves four steps, i.e. **image acquisition, pre-processing, feature extraction,** and **classification.** In this study, images from Swedish leaf dataset have been used, which contains 1,125 images of 10 different species. This is followed by pre-processing using Gaussian filtering mechanism and then texture and color features have been extracted. Finally, classification has been done using Multiclass-support vector machine, which achieved accuracy of nearly 88.08%, which we aim to enhance further.

KEYWORDS: image acquisition, pre-processing, feature extraction, and classification.

I. INTRODUCTION

Plants are a fundamental part of life on our planet. They give us oxygen to breathe, food, medicine and plenty of other things which make our lives worth living. They are the backbone of all life. However, identifying the plants correctly is out of reach of an ordinary person as it requires specialized knowledge, and only the experts of botanical background are able to pull off this task. Moreover, even botanists do not have knowledge of all the existing plants in this world for there is an unlimited number of plant species. Hence, the task of plant identification is limited to a very small number of people. However, plant species knowledge is necessary for various purposes such as identifying a new or rare species, balancing of the ecosystem, medicinal purposes, agricultural industry, etc. To be able to achieve these objectives, automation of plant species identification is a necessity.

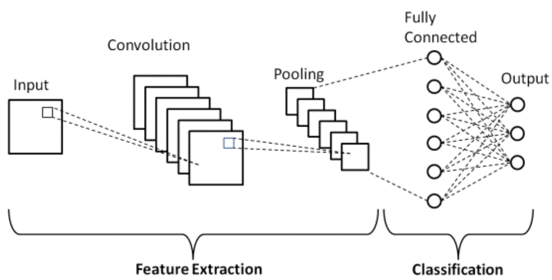


Fig1. CNN Layer Diagram

II. LITERATURE SURVEY

2.1 Title: Hyper parameter optimization of deep CNN classifier for plant species identification using artificial bee colony algorithm

Description: Tailoring a deep convolutional neural network (CNN) for an implementation is a tedious and time-consuming task especially in image identification. In this study, an optimization scheme based on artificial bee colony (ABC) algorithm so-called optimal deep CNN (ODC) classifier for hyperparameter optimization of deep CNN is

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006108 |

proposed for plant species identification. It is implemented to a ready-made leaf dataset namely Folio containing #637 images with 32 different plant species. The images had undergone various image preprocessing such as scaling, segmentation and augmentation so as to improve the efficacy of the ODC classifier. Therefore, the dataset is augmented from #637 to #15,288 leaf images whose #12,103 images is allocated for training phase and the remainder for testing the ODC. Moreover, a validation process on 20% of the training dataset is performed along with the training phase in both optimization and classification stages. The accuracy and loss performance of the ODC are examined over the training and validation results. A new approach for optimizing the parameters and weights of CNN through BO and BA will be explained. First, design of experiments technique will be conducted to investigate the significant factors affecting validation accuracy (VA). Then, BO will be used to find the optimal hyperparameters for the network by optimizing the significant factors in order to minimize the classification error on the validation set, which is the objective function.

2.2 Title: The specific DNA barcodes based on chloroplast genes for species identification of orchidaceous plants
Description:DNA barcoding is currently an effective and widely used tool that enables rapid and accurate identification of plant species. The Orchidaceae is the second largest family of flowering plants, with more than 700 genera and 20,000 species distributed nearly worldwide. The accurate identification of Orchids not only contributes to the safe utilization of these plants, but also it is essential to the protection and utilization of germplasm resources. In this study, the DNA barcoding of 4 chloroplast genes (matK, rbcL, ndhF and ycf1) were used to provide theoretical basis for species identification, germplasm conservation and innovative utilization of orchids. By comparing the nucleotide replacement saturation of the single or combined sequences among the 4 genes, we found that these sequences reached a saturation state and were suitable for phylogenetic relationship analysis. The phylogenetic analyses based on genetic distance indicated that ndhF and ycf1 sequences were competent to identification at genus and species level of orchids in a single gene. In the combined sequences, matK + ycf1 and ndhF + ycf1 were qualified for identification at the genera and species levels, suggesting the potential roles of ndhF, ycf1, matK + ycf1 and ndhF + ycf1 as candidate barcodes for orchids.
The mitochondrial gene COI has been widely used by taxonomists as a standard DNA barcode sequence for the identification of many animal species. However, the COI region is of limited use for identifying certain species and is not efficiently amplified by PCR in all animal taxa.

II. PROPOSED SYSTEM

In the project, the convolutional neural network based plant species prediction system has been implemented. In the proposed method, 80+ plant species image dataset has been used to build the model. The collected image dataset will be preprocessed by four kind of image preprocessing stages which are image resizing, Data augmentation, data normalization. For the detection of species, the convolutional neural network is utilized to extract the feature and performs the classification. The classification metrics such as accuracy, precision score, recall score and F1 score are calculated to evaluate the performance of the model. The developed model will be deployed with help of Gradio python packages.

ADVANTAGE

- Accuracy of the model will be increased.
- More than 80+ species will be used to train the model.
- Overall performance of the system will be increased.

III. IMPLEMENTATION

MODULE LIST

- Data collection
- Image preprocessing
- CNN development
- Performance evaluation

MODULE DESCRIPTION

DATA COLLECTION

The dataset are collected from kaggle. This Kaggle data collection has only numerical value and in this data can be used for multiple purpose. Kaggle supports a variety of dataset publication formats, but we strongly encourage dataset publishers to share their data in an accessible, non-proprietary format if possible. Not only are open, accessible data formats better supported on the platform, they are also easier to work with for more people regardless of their tools.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |
|| Volume 10, Issue 6, June 2022 ||
| DOI: 10.15680/IJIRCCE.2022.1006108 |

IMAGE PREPROCESSING

Image pre-processing is an important step as it helps to enhance the quality of image for further processing. This step is necessary as an image inherently contains noise and this may result in lower classification accuracy. It is performed to remove the noise that hampers the identification process and handle the degraded data. A series of operations are followed to improve the image of the leaf which include, converting the RGB image to grayscale, then from grayscale to binary, followed by smoothing, filtering etc. Pre-processing mechanism used in this paper contains noise handling along with resizing operation and image enhancement.

CNN DEVELOPMENT

The network we have build is a CNN (Convolutional Neural Network) and termed as it Cancer Net. This network performs the following operations. Use 3×3 CONV filters Stack these filters on top of each other Perform max-pooling Use depth wise separable convolution (more efficient, takes up less memory.Its consists of input layers, convolution layers, ReLu layer, maxpooling layers for extract the features of images of build model. Feature extraction train the model the build the model.

PERFORMANCE EVALUATION

The testing process is implemented by this function can split the model with a test set of 30% of the original data set. The input just specify the size of the input and is called D (see the code above X_ train shape). The dense layer is used where the real work happens; it takes the input and does a linear transformation to get an output of size 1. The linear transformation we want to apply is the sigmoid activation function so that in output we are in a range of 0 and 1. Loss per iteration, training loss, validating loss is implemented in module. Accuracy and sensitivity of the analyzed.

GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.

IV. SYSTEM ARCHITECTURE

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.

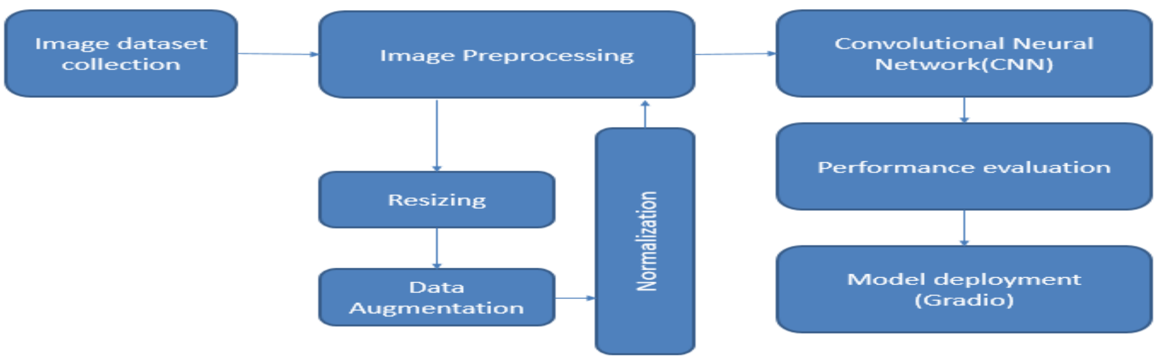


Fig 2. System Architecture

IMAGE PREPROCESSING

Each leaf image was duplicated twice to execute the three preprocessing strategies: **non-preprocessed**, **cropped**, and **segmented**. Non-preprocessed images were kept unaltered. Cropping was performed based on a bounding box enclosing the leaf .To facilitate an efficient segmentation, we developed a semi-automated

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006108 |

approach based on the GrabCut method. GrabCut is based on iterated graph cuts, and was considered accurate and time-effective for interactive image segmentation. The first iteration of GrabCut was initialized by a rectangle placed at the relevant image region, the focus area defined during image acquisition and available in an image's EXIF data. This rectangle should denote the potential foreground whereas the image corners were used as background seeds. The user was then allowed to iteratively refine the computed mask by adding markers denoting either foreground or background, if necessary. The total amount of markers was logged for every image. To speed up the segmentation process, every image was resized to a maximum of 400 px at the longest side while maintaining the aspect ratio. Finally, the binary mask depicting only the area of the leaf was resized to the original image size. The boundary of the upsized mask was then smoothed using a colored watershed variant after morphological erosion of the foreground and background labels, followed by automated cropping to that mask. Image acquisition and preprocessing require substantial manual effort depending on the image type and preprocessing strategy. We aim to quantify the effort for each combination in order to facilitate a systematic evaluation and a discussion of their resulting classification accuracy in relation to the necessary effort.

RESIZING

After handling noise, resizing operation has been performed. In our study, the images have been resized to [300 × 400]. Resizing is done using equation 2.

DATA AUGMENTATION

Data augmentation in data analysis are techniques used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data. It acts as a regularizer and helps reduce overfitting when training a machine learning model.

CONVOLUTIONAL NEURAL NETWORK(CNN):

Deep Learning – which has emerged as an effective tool for analyzing big data – uses complex algorithms and artificial neural networks to train machines/computers so that they can learn from experience, classify and recognize data/images just like a human brain does. Within Deep Learning, a Convolutional Neural Network or CNN is a type of artificial neural network, which is widely used for image/object recognition and classification. Deep Learning thus recognizes objects in an image by using a CNN. CNNs are playing a major role in diverse tasks/functions like image processing problems, computer vision tasks like localization and segmentation, video analysis, to recognize obstacles in self-driving cars, as well as speech recognition in natural language processing. As CNNs are playing a significant role in these fast-growing and emerging areas, they are very popular in Deep Learning.

V. PERFORMANCE EVALUATION

TRAINING GRAPH LOSS

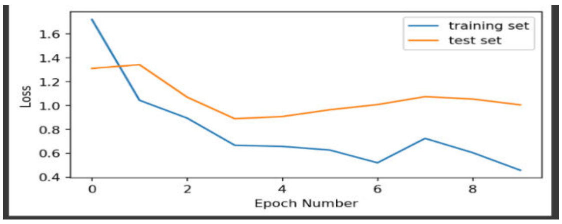


Fig 3. Traning Graph Loss

TRAINING GRAPH ACCURACY

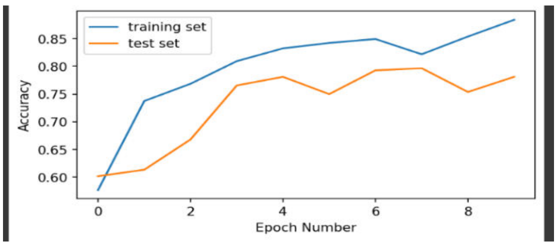


Fig 4. Training Graph Accuracy

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006108 |

VI. RESULT

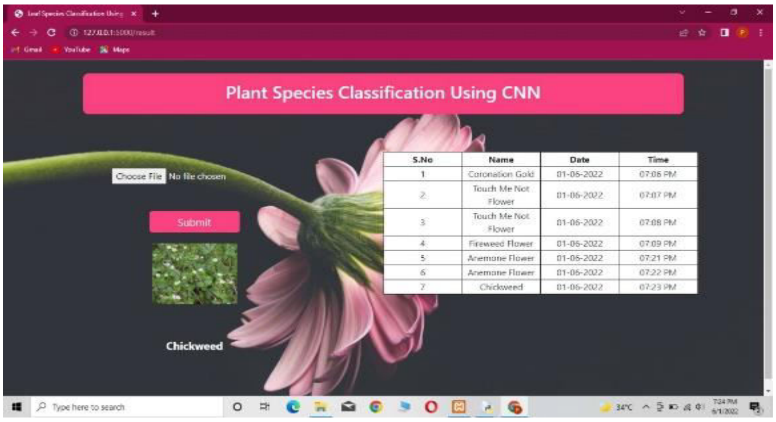


Fig 5. Result

VII. CONCLUSION

This paper has proposed an automatic plant species identification approach which is employed using computer vision and machine learning techniques to classify plant leaf images.
The study has been conducted in phases like image pre-processing, image segmentation, feature extraction and finally classification of the image.
A combination of texture and color features (5 and 4 respectively) were extracted and then SVM classifier was used for classification. The system was tested on Swedish dataset and attained an average accuracy of 88.08%.The model could automatically classify 15 different plant species. Texture and color feature space performed satisfactorily well in comparison to the methods which work only on morphological shape features.

REFERENCES

[1] S. H. Lee, C. S. Chan, P. Wilkin, P. Remagnino, "Deepplant: Plant identification with convolutional neural networks," in Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), pp. 452-456, 2015.
[2] J. Wäldchen, P. Mäder, "Plant species identification using computer vision techniques: A systematic literature review," Archives of Computational Methods in Engineering, vol. 25, no. 2, pp. 507-543, 2018.
[3] 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.
[4] D. Shreya, "How many plant species are there in the world? Scientists now have an answer," MONGABAY NEWS AND INSPIRATION FROM NATURE'S FRONTLINE, May 12, 2016.
[5] S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang, Q.-L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," In Proceedings of 2007 IEEE International Symposium on Signal Processing and Information Technology, pp. 11-16, 2007.
[6] N. Valliammal, S. N. Geethalakshmi, "Automatic recognition system using preferential image segmentation for leaf and flower images," Computer Science & Engineering, vol. 1, no. 4, pp. 13-25, 2011.
[7] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. Lopez, J. V. B. Soares, "Leafsnap: A computer vision system for automatic plant species identification," In Proceedings of European Conference on Computer Vision, pp. 502-516, 2012.
[8] J. Hossain, M. A. Amin, "Leaf shape identification based plant biometrics," In Proceedings of 13th International Conference on Computer and Information Technology (ICCIT), pp. 458-463, 2010.
[9] Z. Wang, B. Lu, Z. Chi, D. Feng, "Leaf image classification with shape context and sift descriptors," In Proceedings of 2011 International Conference on Digital Image Computing: Techniques and Applications, pp. 650-654, 2011.
[10] C. A. Priya, T. Balasaravanan, A. S. Thanamani. "An efficient leaf recognition algorithm for plant classification using support vector machine," In Proceedings of International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012), pp. 428-432, 2012.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006108 |

BIOGRAPHY

Musharaf Hussain.A Department of Computer Science And Engineering, From Saranathan College Of Engineering, Panjappur, Trichy.

Ms.S.Ranjani. M.Tech Assistant Professor of Computer Science And Engineering, From Saranathan College Of Engineering, Panjappur, Trichy.

Santosh Kumar.R Department Of Computer Science And Engineering, From Saranathan College Of Engineering, Panjappur, Trichy.

Thamaraikani.J Department Of Computer Science And Engineering, From Saranathan College Of Engineering, Panjappur, Trichy.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com