HERBAL LEAF RECOGNITION USING MASK-REGION CONVOLUTIONAL NEURAL NETWORK (MASK R-CNN)

Article · June 2022

3 authors:

Laiali Almazaydeh
Abu Dhabi University

Reyad Salameen
Al-Hussein Bin Talal University

Khaled Elleithy
University of Bridgeport

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HERBAL LEAF RECOGNITION USING MASK-REGION CONVOLUTIONAL NEURAL NETWORK (MASK R-CNN)

LAIALI ALMAZAYDEH[1], REYAD ALSALAMEEN[2], KHALED ELLEITHY[3]

[1,2]Al-Hussein Bin Talal University, Department of Software Engineering, Jordan

[3]University of Bridgeport, Department of Computer Science & Engineering, USA

E-mail: [1]laiali.almazaydeh@ahu.edu.jo, [2]reyad.m.salameen@ahu.edu.jo, [3]elleithy@bridgeport.edu

ABSTRACT

Recent rapid technological advancements in pattern recognition and computer vision have led to great results on a wide range of applications. One of these applications is herbal plant species identification, as the proper automated system for the recognition of herbal plants is required for botanists to study therapeutic and nutritional uses of herbs. In literature, many studies have adopted classical machine learning approaches while some studies have adopted deep learning approaches, via the leaf images. For this work, we use the latest state-of-the-art framework, namely Mask R-CNN, to build such a classification system to identify a medicinal plant. In this paper, we demonstrate the development of the classification system using Mask R-CNN and its backbone: region proposal network, RoI Pooling, RoI Align, and the network head: classification & detection, segmentation. The trained model achieved average accuracy of 95.7% for the identification of 30 medicinal plant species loaded from the Mendely Dataset. The model output is obtained as bounding box for object detection, mask, and class indicating a plant species.

Keywords: *CNN, Deep learning, Mask R-CNN, PPIR, RPN*

1. INTRODUCTION

In 2019, the UN identified the need for agricultural output to double by 2050 [1], [2], therefore, in order to meet this need, a method called "plant phenotyping" is required for decision support in agriculture and also for the farmers to select plants with best properties in their prevailing environment, hence the farmers will be able to grow healthy plants in a sustainable way. The plant phenotyping was mostly achieved by a visual inspection of the plants which was limited by the throughputs and the accuracy as well, therefore, an automated plant phenotyping is needed to characterize a plant's physiological and biological traits by analyzing color images.

Plant phenotyping image recognition (PPIR) technology are applied to diagnose plant diseases, analyze growth rate, identify plant species, etc. [3]. However, plant species identification is a challenging task, even for experienced botanists, due to the great diversity of species [4], [5]. Among these species are medicinal plants, which include various types of plants used for therapeutic and nutritional purposes. Medicinal leaves are important part for plant identification as they appear for a long interval and are available in abundance, in contrast, the floral parts appear for a short interval.

Nowadays, the modern technology is used to observe plants in a systematic scientific way, whereas the latest non-invasive imaging and centering technologies supported by robotics, drones, and Artificial Intelligence (AI) are being available. In the field of AI, neural networks and deep learning are encompassed in the tasks of pattern recognition and classification [6]. A new type of neural network architecture called Mask Region Convolutional Neural Networks (Mask R-CNN) developed on top of Faster R-CNN, while Faster R-CNN is updated version of Fast R-CNN and Fast R-CNN is updated version of R-CNN.

The Region-CNN (R-CNN) proposed by Girshick et al. [7] is one of the state-of-the-art of the CNN based deep learning object detection approaches. One drawback with the original R-CNN framework is using selective search algorithm to extract the region of interest (final object locations), specifically for each image, up to 2000 regions will be extracted, so this process makes R-CNN slow and computationally expensive. Another drawback is the memory requirement because of the need to have multiple classifiers for each class of objects. Therefore, the speed and memory issues related to R-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CNN resolved to some extent in Fast R-CNN with the removal of selective search algorithm and some architectural changes.

The proposed Fast R-CNN proposed by Girshick [8] solves previously mentioned drawbacks by applying training in an end-to-end manner instead of training different modules separately. All modules used in the object detector, like convolution layers for feature extraction, Support Vector Machine (SVM) for classification and box regressor for bounding box regression are trained as a single network. As a result, the proposed Fast R-CNN model is 9x faster than R-CNN model.

The Faster R-CNN model proposed by Ren et al. [9] is an enhanced version of Fast R-CNN model, and it is much faster than the other versions of R-CNN, as it is 250x faster than R-CNN. Faster R-CNN consists of detector which is Fast R-CNN and region proposal which is RPN as a unified network. Therefore, the Mask R-CNN proposed by He et al. [10] is a consequence of adding a branch to Faster R-CNN that output the object mask.

In this paper, we use the latest state-of-the-art framework, namely Mask R-CNN, to build such a classification system to identify a medicinal plant. The model output is obtained as bounding box for object detection, mask, and class indicating a plant species.

This paper is organized as follows. Section 2 presents an overview of existing medicinal plants classification approaches. Section 3 is the development of the classification system using Mask R-CNN and its backbone: region proposal network, RoI Pooling, RoI Align, and the network head: classification & detection, segmentation. Section 4 discusses about the application evaluation and results. Section 5 summarizes how the research objectives are being achieved.

2. RELATED WORKS

In the past few years, some object detection and recognition studies are oriented toward the plant phenotyping analysis via leaf images. In this section, several related works, which are based on different technology for plant species identification, are reviewed in general.

A study [11] was utilized the Probabilistic Neural Network (PNN) to classify 32 kinds of common plants in China, 12 commonly used digital features of leaves were extracted, including morphological ones and geometrical ones. Principal Component Analysis (PCA) was used to orthogonalize these 12 features into 5 principal components to form the input vectors of the PNN. Experimental result indicated that the average accuracy of this algorithm is greater than 90%. Using a different dataset with 24 common species in Mauritius and a different classifier, Begue et al. [12] achieved almost a similar level of accuracy with the random forest classifier.

Gu et al. [13] used a combination of Gaussian interpolation and wavelet transform to extract the skeleton of the whole leaf. Then the classifiers, a radial basis probabilistic neural network (RBPNN), a k-nearest neighbor (k-NN), and a nearest neighbor (1-NN) were used to perform the recognition of twenty species images to a achieve a classification accuracy of 85.4%, 91.1% and 93.1% respectively. Other study [14] used k-NN algorithm to classify Thai herb leaf based on seven leaf features and the algorithm achieved an accuracy of 93.29%.

AlAsadi et al. [15] used two approached of classifiers; SVM and Neural Network feed-forward on two different datasets. A set of leaf features based on texture and shape were extracted for the plant recognition and accuracy of 91% was obtained.

Obviously, the greater part of the studies referenced above have focused on the recognition with derived features which are generally used with “classical” machine learning approaches. However, most of these derived features are low-level image representation, which is affected by background and noise, so it is difficult to be employed in practical applications [16]. Recently, "newer" approaches based on deep learning for the recognition with learned features which are automatically obtained by training the dataset have attracted the attention of many researchers to use the latest deep learning models.

In [16], Vo et al. used the Convolutional neural networks (CNNs) to classify the dataset containing 10 herbal plant species in Vietnam. A VGG16-based deep learning model comprising of 5 residual building blocks was utilized to extract the deep convolutional features from the images. The trained CNN achieved a recognition rate of 93.6%. Also, other studies have increasingly replaced the traditional AI process by CNN, like the study in [17] for Chines-herbal identification, and the study in [18] for weed species identification, with the accuracy reaching 71% and 86.2% respectively.

In [19], Mookdarsanit used Fast R-CNN for model architecture and feature extraction framework to identify 11 well-known Thai herbs, the trained Fast R-CNN achieved accuracy higher than 80%.
The work presented in this paper is inspired by one of the latest deep learning models, which is Mask

Journal of Theoretical and Applied Information Technology
15th June 2022. Vol.100. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

R-CNN to provide robust recognition system of herbal plants.

3. MATERIALS AND METHODS

3.1 Dataset

The dataset of medicinal herbs used is available from Mendeley in [20]. Mendeley offers free access to the comprehensive medicinal leaf dataset, which we used to train and test our approach. The dataset contains 1800 images of thirty species, each species consists of 60 to 100 high-quality images. This dataset provides benchmark data that can be used by researchers to develop and enhance their AI used models for detection, recognition the species and its diseases.

The instruments used to capture and print the images are mobile camera model (Samsung s9+) and (Canon inkjet) printer, then the images were stored in JPG format to provide organized database. In addition, a segmentation process to facilitate image recognition carried out on Mendeley dataset.

Fig. 1 shows the collection of the 30 species with their scientific names.

3.2 Annotation

Annotation of the images is the primary step, in light of the fact that the training data set used in Mask R-CNN must be labeled. The (Makesense.AI.) [21] image annotator was used to annotate the images; this tool is a web-based lightweight annotation tool. Each image in the training data was annotated manually by plotting bounding box on the shape of the region which specifying exact points of the leaf border, then label that image by giving name which is the short form of the scientific name of the leaf class, ex. AG which represents Alpinia Galanga. This process generated "json" file corresponding to each image. The "json" file contains the information regarding the leaf shape coordinates, image width, image height, label name, etc. These information per image is necessary for the training purpose to detect and recognize the class of the leaf object.

Here we chose 23 images per each species, and as the dataset contains 30 species of leaf, accordingly, 690 images were annotated. Fig. 2 demonstrates the annotation process for the image of Alpinia Galanga plant.

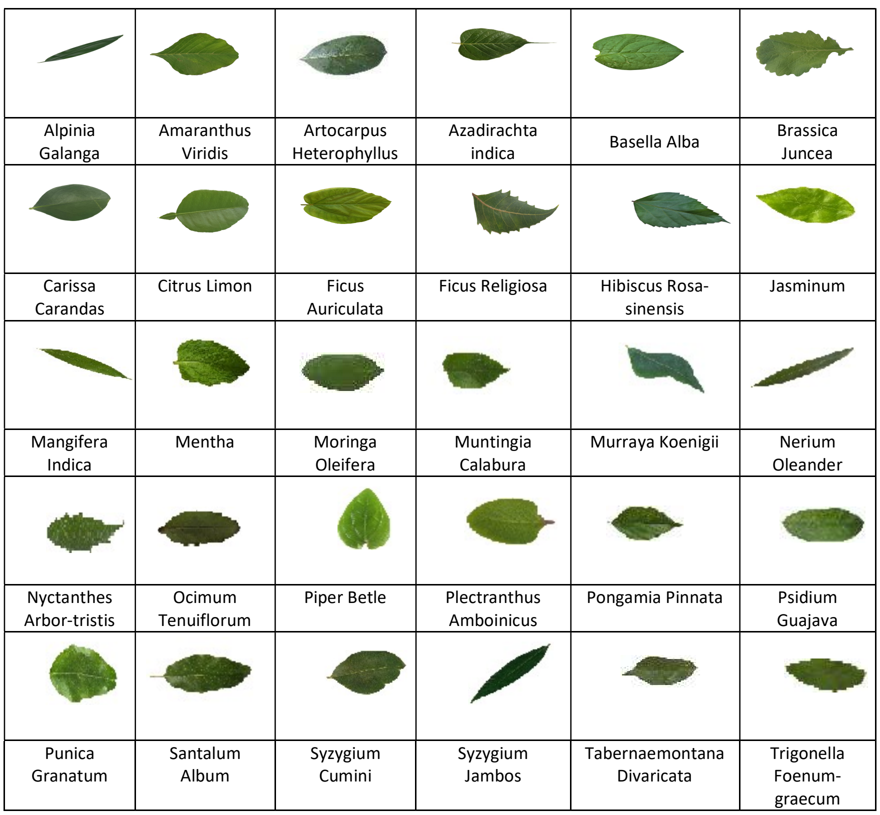

Figure 1: The collection of the 30 species

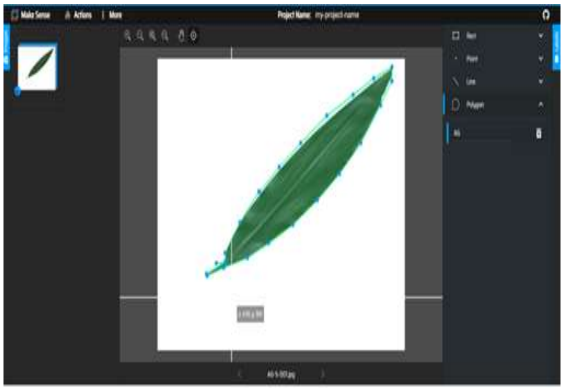

Figure 2: The image of Alpinia Galanga plant during the annotation process

3.3 MASK R-CNN

A new type of neural network architecture called Mask Region Convolutional Neural Networks (Mask R-CNN) was developed by a group of Facebook AI researchers in 2017 [10], in order to facilitate some key challenging research problems such as instance segmentation in computer vision field. Instance segmentation is the process of detecting and delineating each distinct object of interest in an image, and so instance segmentation is a combination of two subproblems, first is object detection which is the problem of finding and classifying a variable number of objects in an image, and second is semantic segmentation which is a problem of understanding of an image at the pixel level where an object class assigned to each pixel in an image. Therefore, the developed architecture can return an object class, a mask, and the bounding box for each detected object in an image.

Basically, Mask R-CNN is a combination of the Faster R-CNN with Fully Convolutional Networks (FCNs) which run on each Region of Interests (RoIs). The Mask R-CNN is similar to the faster R-CNN, but additionally it output the object mask, where FCNs are used to predict the mask from each RoI align which preserves the spatial orientation of features and leads to no loss of information. As shown in fig. 3, the architecture behind Mask R-CNN, mainly consists of two stages [10]. The first is a Region Proposal Network (RPN), "which anticipates object proposal bounding boxes based on anchor boxes", then feed them into the second stage. The second stage is an R-CNN detector, which "refines these proposals, classifies them, and computes the pixel-level segmentation for these proposals".

In the following, according to the Mask R-CNN architecture, we explain the main steps in more details, along with an example of herb leaf image, to follow change at each step.

3.3.1 Convolution Neural Network (CNN)

The input image is passed through a pre-trained CNN to generate feature maps. CNN [22] has multiple convolutional layers to extract the features at different levels, ranging from low level features such as lines specified by early layers and in the later layers successfully higher-level features were extracted, which are combinations of low-level features, such as features that comprise multiple lines to specify shapes.

So thereby in our work, we used the backbone network Resnet101 [23], [24] for feature extraction, which is a proven architecture with less error rate, and this was improved by adding Feature Pyramid Network (FPN) [25] to present the object better at multiple scales. The model was trained over the dataset, which is roughly around 690 images, that were reloaded from the "json" file.

In addition, image data augmentation was used to artificially expand the training dataset in order to improve the ability of the model to generalize. For augmentation, the Python "imgaug" module is used to improve convergence. All augmenters were applied to a random 50% of the images in a batch and executed in random order. The following parameters for resizing and scaling are considered:

```
IMAGE_RESIZE_MODE = "square"
IMAGE_MIN_DIM = 800
IMAGE_MAX_DIM = 1024
IMAGE_CHANNEL_COUNT = 3
```

Fig. 4 shows some network layers output at different stages of Mask R-CNN model.

3.3.2 Region Proposal Network (RPN)

From the backbone feature maps, the RPN discover the location of the object, so the task of RPN starts with the use of anchor boxes. Anchor boxes are basically a specified collection of bounding boxes with a specific height and width, once anchor boxes is generated, next is to calculate Intersection over Union (IoU) with the ground truth boxes (referred to as Non-maximum Suppression) , so the anchor box with the higher IoU will be labeled as foreground class where the value of IoU is greater than or equal to 0.5, and the area where IoU is less than to 0.5 will be labeled as background class [26], so the task of RPN is to predict foreground and background anchor boxes and finally the anchor

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

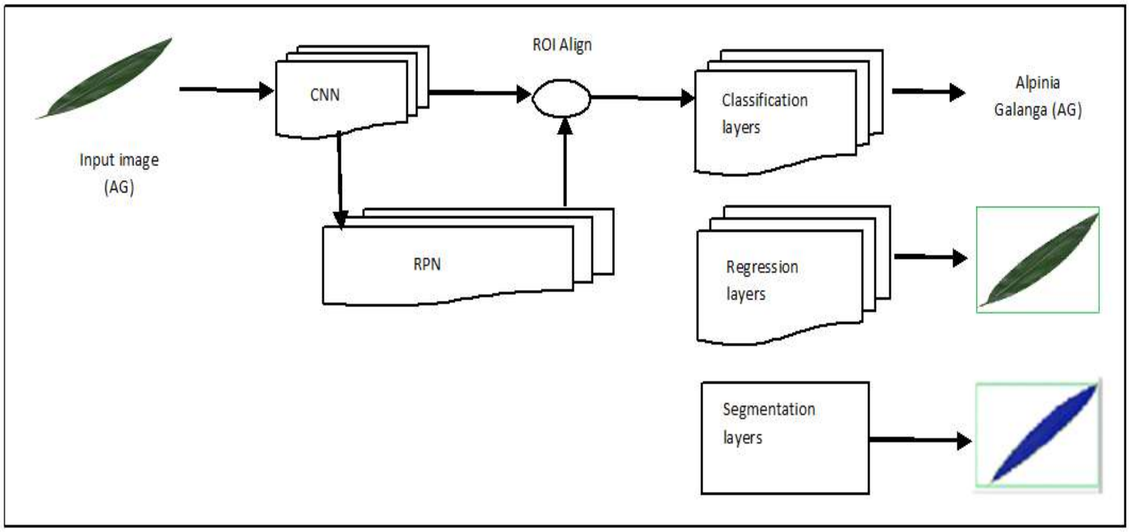

Figure 3: The architecture of the Mask R-CNN

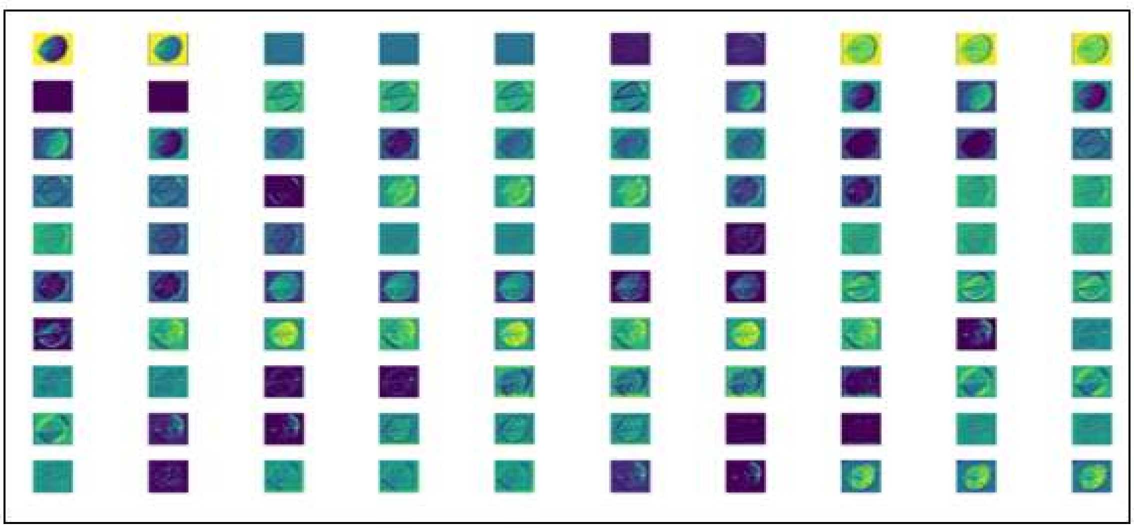

Figure 4: Resnet 101 last layers feature maps

boxes which are labeled as foreground class will pass to the next stage.

Fig. 5 shows refined anchors after non-max suppression.

3.3.3 Region of Interest (ROI) Align

This is the second stage which is also known as ROI pooling. There are two inputs that ROI pooling received, one is the output anchor boxes of different sizes from RPN, and the other input is different size of feature maps from pre-trained network. Therefore, the functionality here is to reduce all the feature maps to same size for each region proposal.

3.3.4 Bounding Box Regressor and ROI Classifier

This step runs on the RoI proposed by RPN and generates two outputs for each RoI, the class of the object in the RoI like Jasminum, Mentha, etc., and refined bounding box according to the size and location of the object.

3.3.5 Segmentation Masks

Conceptually, the Mask R-CNN is similar to the faster R-CNN, except that Mask R-CNN additionally outputs an object mask using pixel to pixel alignment, and this mask is a binary mask

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

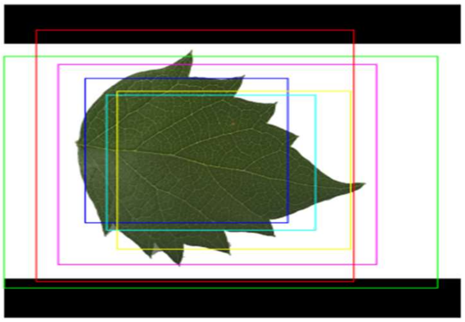

Figure 5: Illustration of anchor boxes

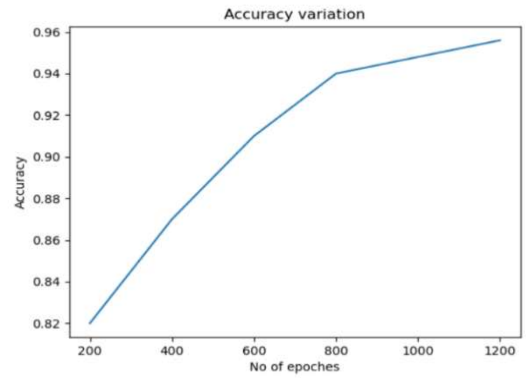

Figure 6: Number of epochs versus accuracy

outputted for each ROI based on IoU values. Not much overhead is incurred when computing this mask as it is another parallel branch with the bounding box creation and classification. The mask prediction branch is fully convolutional because convolution layers retain spatial orientation while such information is essential for location specific tasks like constructing an M × M object mask.

4. RESULTS

We evaluated the effectiveness of the Mask R-CNN architecture on Mendeley dataset using different medicinal plants images available in this dataset. To each species of medicinal leaf, 23 images were used to train the model. The experiment was performed under Windows 10 operating system using Python 3 [27].

The results of fine-tuning the Mask R-CNN model can extract image features using ResNet [24] of depth 101 layers and this is improved by adding FPN, also the model generate ROIs by RPN, predict the object class using softmax classifier, generate precise mask, and this is combined with bounding box.

In this experiment, 30 species of medicinal plants are used for training as well as testing, images was split into training and testing sets with 690 and 1085 images samples, respectively.

Drastic change in accuracy is observed as the model is trained up to 1200 epochs as shown in fig. 6. The trained model achieved average accuracy of 95.7% for the identification of medicinal plant species and 100% for the segmentation masks of the leaf, thereby, this provides a solid technical basis in identifying plant species. The experiment results are listed in Table. 1, and outputs samples are visualized in fig. 7, this figure shows that Mask R-CNN ran successfully as each processed image is segmented with mask and bounding box, and it is classified according to its object class.

Table 1: The experiment results

Parameter	Quantity	Percentage
Total No: of classes	30	-
Total images for training	690	-
Total images for testing	1085	-
No: of images Correctly identified	1038	95.67%
No: of images wrongly identified	47	4.33%
No: of images correct mask detected area	1085	100%

The model is able to predict all the dataset species, so that the recognition Accuracy per plant species is shown in fig. 8.

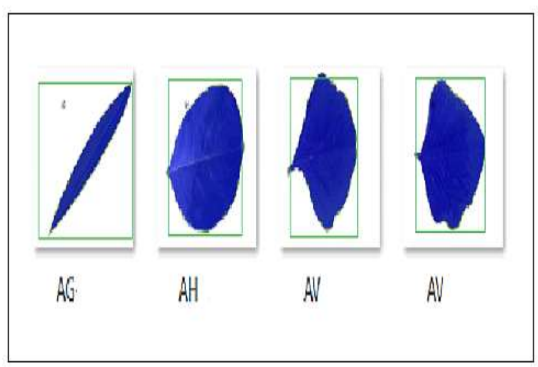

Figure 7: The output samples

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

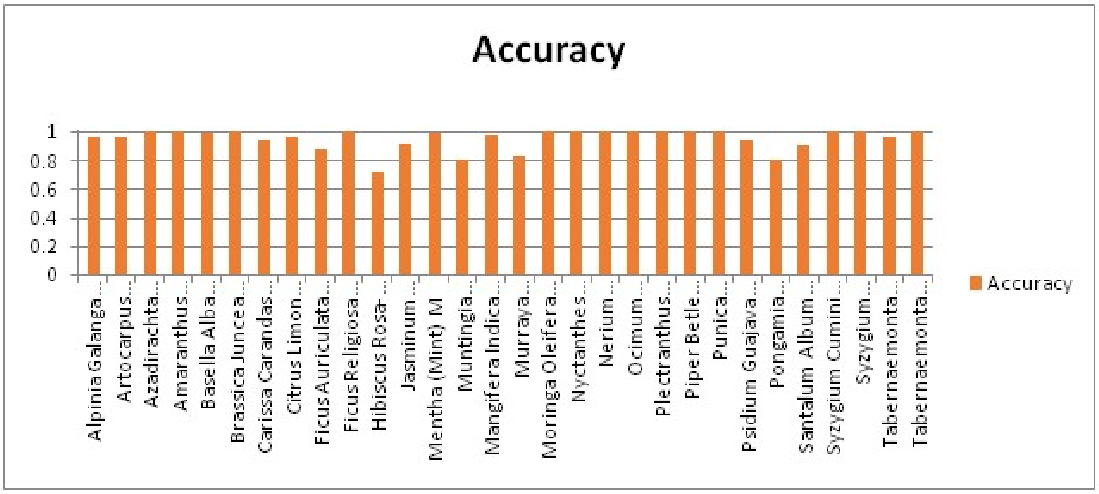

Figure 8: The classification accuracy of different species

4. CONCLUSION

This paper presents Mask R-CNN approach, which provides robust recognition system of herbal plants. The trained model achieved average accuracy of 95.7% for the identification of 30 medicinal plant species loaded from the Mendely Dataset. The accuracy of Mask R-CNN outperforms all other classical machine learning approaches and other deep learning approaches that only use leaf image features. As a future work, we plan to incorporate this work into PPIR technology that involve various operations like plant diseases diagnosis and growth rate analysis.

REFERENCES:

[1] United Nations, Department of Economic and Social Affairs, Population Division (2019), "World Population Prospects 2019: Highlights". Accessed on: June. 4, 2021. [Online]. Available: www.unpopulation.org.

[2] D. Tilman, C. Balzer, J. Hill, and B. Befort, "Global food demand and the sustainable intensification of agriculture," Proceedings of the National Academy of Sciences of the United States of America, vol. 108, no. 50, pp. 20260-20264, Dec 2011.

[3] J. Xiong, D. Yu, S. Liu, L. Shu, X. Wang, and Z. Liu, "A Review of Plant Phenotypic Image Recognition Technology Based on Deep Learning," Electronics, vol. 10, no. 81, pp. 1-19, 2021.

[4] K. Yang, W. Zhong, and F. Li, "Leaf Segmentation and Classification with a Complicated Background Using Deep Learning," Agronomy, vol. 10, no. 1721, pp. 1-12, 2020.

[5] L. Li, Q. Zhang, and D. Huang, "A Review of Imaging Techniques for Plant Phenotyping," Sensors, vol. 14, no. 11, pp. 20078-20111, Nov 2014.

[6] Y. Bengio, and Y. LeCun, "Scaling learning algorithms towards AI," Large-scale kernel machines, vol. 34, no. 5, pp. 1-41, 2007.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik., "Rich feature hierarchies for accurate object detection and semantic segmentation," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587, 2014.

[8] R. Girshick, "Fast r-cnn," In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448, 2015.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, vol. 28, pp. 91-99, 2015.

[10] K. He, G. Gkioxari, P. Dollar, R. Girshick, "Mask R-CNN," In proceedings of the IEEE international conference on computer vision, pp. 2961-2969, 2017.

Journal of Theoretical and Applied Information Technology
15th June 2022. Vol.100. No 11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[11] G. Wu, S. Bao, Y. Xu, X. Wang, F. Chang, and L. Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network," IEEE international symposium on signal processing and information technology, pp. 11-16, 2007.

[12] A. Begue, V. Kowlessur, F. Mahmoodally, U. Singh, and S. Pudaruth, "Automatic Recognition of Medicinal Plants using Machine Learning Techniques," International journal of Advanced Computer Science and Applications, vol. 8, no. 4, pp. 166-175, 2017.

[13] X. Gu, j. Du, and X. Wangt, "Leaf Recognition Based on the Combination of Wavelet Transform and Gaussian Interpolation," In international conference on intelligent computing, pp. 253-262, 2005.

[14] C. Pornpanomchai, R. Supolgaj, P. Tanasap, and C. Chaiyod, "Thai Herb Leaf Image Recognition System (THLIRS)", Agriculture and Natural Resources, vol. 45, no. 3, pp. 551-562, 2011.

[15] A. AlAsadi, E. Anduljalil, and A. Khaleel," Leaf Recognition based on Neural Network Feed-Forward and Support Vector Machine Classifiers," International Journal of Computer Science and Mobile Computing, vol. 6, no. 1, pp. 92-99, 2017.

[16] A. Vo, H. Dang, B. Nguyen, and V. Pham, "Vietnamese Herbal Plant Recognition Using Deep Convolutional Features," International Journal of Machine Learning and Computing, vol. 9, no. 3, pp. 363-367, 2019.

[17] X. Sun, and H. Qian, "Chinese Herbal Medicine Image Recognition and Retrieval by Convolutional Neural Network," PLoS ONE, vol. 11, no. 6, pp. 1-119, 2016.

[18] M. Dyrmann, H. Karstoft, and H. Midtiby," Plant species classification using deep convolutional neural network, Biosystems e ngineering, vol. 151, pp. 72-80, 2016.

[19] L. Mookdarsanit, and P. Mookdarsanit, "Thai Herb Identification with Medicinal Properties Using Convolutional Neural Network," Suan Sunandha Science and Technology Journal, vol. 6, no, 2, pp. 34-40, 2019.

[20] S. Roopashree, J. Anitha, "Medicinal Leaf Dataset," Mendeley Data, V1. Accessed on: May. 4, 2021. [Online]. Available: https://data.mendeley.com/datasets/nnytj2v3n5/1.

[21] Makesense.AI. Accessed on: May. 4, 2021. [Online]. Available: https://www.makesense.ai/

[22] B. Traore, B. Foguem, and F. Tangara, "Deep convolution neural network for image recognition," Ecological Informatics, vol. 48, pp. 257-268, 2018.

[23] Y. Chen, T. Yang, X. Zhang, G. Meng, X. Xiaoy, and J. Sun, "Detnas: Backbone search for object detection," Advances in Neural Information Processing Systems, vol. 32, pp. 6642-6652, 2019.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[25] T. Lin, P. Dollar, and R. Girshick, "Feature Pyramid Networks for Object Detection," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117-2125, 2017.

[26] Musyarofah, V. Schmidt, and M. Kada, "Object detection of aerial image using mask-region convolutional neural network (MASK R-CNN)," IOP Conference Series: Earth and Environmental Science, vol. 500, no. 1, pp. 1-10, 2020.

[27]https://www.python.org/downloads/release/python-395/