RARE MEDICINAL PLANT LEAVE IMAGE CLASSIFICATION USING RESNET-50 BASED ON TRANSFER LEARNING

M. Usha Rani[1], C.S. Udaya[2]
[1]Professor, Department of Computer Science, SPMVV, Tirupati (A.P.), India,
[2]Lecturer, Department of Computer Science, Smt. N.P.S Govt. College(W), Chittoor (A.P.), India,

Abstract

Now-a-days our lifestyle becoming techno-savvy, we are living away from the nature, but we cannot go away from the nature because we are the part of nature. As Medicinal plants are natural products, comparatively they are free from side effects, safe, eco-friendly and they available local. So Medicinal plants are very rich resources of ingredients, which can be used to prepare drugs in pharmacy. Other than this, plants play an important role in human culture, some plants have nutrition values, so therapists recommend using those plants but Medicinal plants identification is very greatest challenge. Only professionals and skilled persons can able to identity, but normal people are not able to identify them properly. Pre-trained Convolutional Neural Network (CNN) model gives excellent result in solving feature extraction, identification and classification of images. In this paper Transfer learning based CNN was used to identify and classify the plants, because it enables to create a deep Convolutional Neural Network. Here we used ResNet50 network as a pre-trained model. Other than this, several Transfer Learning architectures used to compare with ResNet50. This model gave 98.42% of training accuracy.

I. INTRODUCTION

Medicinal plants have been used in healthcare since time immemorial. In ancient times all people lived naturally, spending time in nature, people who eat only natural food. So they are less prone to illness. Naturalness is imbued, but in present times people have also forgotten the fact that nature is one. Its consequence is a wide variety of diseases. Many medicines are being developed to get rid of these diseases. There are many types of side effects caused by chemicals. Therefore, all are looking for natural medicines, so the use of medicinal plants in medicine has increased. The reason is plants are natural, the side effect of which is low, it cost is also low and also these plants are available is our local places. No matter how busy we are and forget about nature, there is no going away from it, because we are not separated from nature we are part of it. That is why again everyone is begging again everyone is begging into naturalness, therefore increased use of natural plant based medicines.

Classification of medicinal species is a very big challenge. The plants are all identical to look at so it is difficult to spot them. Using one plant instead of other by mistake can lead to many problems. So it is very important to find the right plant, So if we properly identify the medicinal plants then may get many benefits to those who work on Pharmaceutical industries, Ecological people, Forest dwellers, in the drug Laboratories, in the Glossary, to the Common people and also for Government. We can easily identify the plants manually, but it needs lot of experience and training. Only those who have good experience, good knowledge and experts in this area of work can do this. However, it takes lot of time also and there is a problem that is if they do not have proper idea, then they can identify wrongly, which leads certain side effects and may leads

to serious problems. So we need technology, with its help we can easily identify the medicinal plants.

In this case CNN can be used to identify Medicinal plants. CNN is a part of Artificial Intelligence (AI), which can be used for image processing. It is a CNN powerful image processing technique that uses Deep Learning to perform both generative and descriptive tasks. So it is a Machine Learning technique from the deep learning domain. Neural Network is a system of hardware and/of software. CNN have its own neurons arranged similar to the neural architecture in the frontal lobe of human brain to cover the entire field of visual. CNN used multilayer perception; it is designed in such a manner to reduce precession requirements. The CNN contains so many layers such as input layer, output layer and multiple hidden layers such as, Pooling Layers, Convolutional Layer, Normalization Layer and Fully Connected Layers. Large collection of images is required to classification, from the large collection of images, the CNN extracts the rich collection of features but training a CNN with that large collection is not that much an easy task. Therefore we can use pertained CNN such as ResNet-50, GoogleNet, AlexNex, VGG16, VGG19 etc., to leverage the power of CNN, so that it saves much time and effort for feature extraction. This paper deals with ResNet-50, mean Residual neural network. Also known as Residential network, Residual flow network, Residential Energy Services Network.

ResNet-50 is a pre-trained Convolutional Neural Network, 50 layers deep. We can load more than a millions images; it can classify images into 1000 different categories. This network has learned rich feature representation for a large collection of images.

Remaining paper arranges as follows: Section II contains the related previous work done by the different authors, Section III contains the methodology, Section IV contains experiment results, analysis and finally Section V contains conclusion.

II. RELATED WORK

Image processing has been discussing since few years. Many research works has been done in image classification and identification, they developed many architectures by using machine learning techniques and gave their results on medicinal plant identification and classification.

Ishrat Zahan et al. used Transfer Learning based CNN model to classify the plants based on the leaves, so that can identify plant diseases [1]. ResNet50 network was utilized, it is a popular CNN architecture and also experimented with few other popular pre-trained models 99.8% of training accuracy in identifying plant diseases. Muhammd Hammad et al. compared different Deep Learning Architectures in terms of various performance metrics [2]. Through the experiments author find out that Light Gradient Boosting Machine (GBM) is the best one than the other classifiers because it reduces the feature dimensions in feature extraction process and also in classifier process, got 93.6% of accuracy.

Liwen Gao et al. prepared a method for segmenting images accurately for those having the complex background images of Medicinal Plant leaves [3]. This method directly takes photos of plant leaves for identification achieved high 96.97% of accuracy. Soon Jye Kho et al. used ANN and SVM to identify automate plant, also used image processing technique by recognizing the patters, developed baseline automated system efficiently and achieved 83.3% accuracy [4].

Xiaolong Zhu et al. used CNN method for recognition of plant leaf, discussed about Faster Region CNN, they reviewed that proposed system achieves highest accuracy than the remaining classifiers [5].

C.H. Arun et al. presented an automated medicinal plant leaf identification system [6]. By using Statistical, The Grey Tone Spatial Dependency Matrix and Local Binary Pattern Colour texture feature is identified. Tried various classifier algorithms among all those algorithms from Grey Tone Spatial Dependency Matrices (GTSDM) and Local Binary Pattern (LBP), authors yielded the better accuracy of 98.7% accuracy. But authors used only colour texture feature and taken only 5 species. Yu Sun et al. discussed that plants can be identified in natural environment by using Deep Learning [7]. The combination of Snakes a technique with cellular neural networks used for leaf segmentation and for classification used a 26 deep learning layer contains of 8 residual building blocks was designed and achieved 91.78% of accuracy. Wang-Su et al. used CNN for Plant Leaf Recognition [8]. In this model, authors divided 3,767 training images in to eight types, and 100 test images were selected randomly. Achieved 94% of accuracy, even when 30% of the images were damaged, then also got high accuracy by using CNN.

Mostafa et al. presented a paper in that they have done Plant identification through optimizing the parameters of transfer learning by using deep neural networks [9]. Used deep CNN to identify the plant species. GoogleNet and VGG Net obtained higher performances compared to AlexNet. Achieved 80.18% of accuracy, through the combination of the predictions of better classifiers can improve performance of overall system. Adams Begue et al. presented that Machine Learning Techniques can be used to recognize Medical Plants automatically [10]. Random forest classifier using 10 fold cross validation technique was applied for classification and achieved 90.1% of accuracy. Jana Waldchen et al. discussed that by using Computer Vision techniques, one can identify the plant species. It is a Systematic Literature Review [11]. It is a review, its main aim is to identify different analysis and comparison of computer vision approaches for plant species identification, which is very helpful to the researcher, to recognize the best classifier among all classification methods according to the problem.

Nian Liu et al. presented a leaf identification method by using multi-feature fusion and also used Improved deep belief networks [12]. Before classification should extracted the features like texture and shape, through and also used proportion integration differentiation control to decrease the reconstruction error, these two are used for classification and got 93.9% of accuracy.

Aimen Aakifet al. presented a plant leaves classification by using Automatic classification, through three different stages performed plant identification [13]. First one is pre-processing technique, second one is extracting the features and final stage is classification. Different types of features such as shape defining features, Fourier description and morphological features were extracted then applied Artificial Neural Network for classification, achieved more 96% of accuracy. Guillermo et al. presented a plant identification method using vein morphological patterns with deep learning approach [14]. Here the authors proposed CNN for classification based on leaf vein patterns, Applied Hit or Miss Transform (HMT) to extract vein morphological patterns. Next binary image is cropped; next a set of features extracted by using LEAF GUI achieved accuracy of 96.9%.

In this paper, we developed a deep learning model by using Transfer Learning methodology such as ResNet-50 architecture, and collected pre-trained weights. This proposed model has so many benefits such as with minimum cost features can be extracted and have the quality of fine-tuning to improve the accuracy. The database in our problem includes 10 different species as well as 765 images to classify. In the following sections can see how the classification done.

III. PROPOSED METHODOLOGY

The procedure of CNN for Medicinal plant image classification was presented below in figure 1. First the deep CNN model divides the fed images into train and test sets, pre-trained ResNet-50 architecture loaded, done augmentation, extracted and finally done classification. Measured the accuracy of the trained classifier, then new image can be predicted by algorithm depends on classification and then performed test along with a vast results analysis.

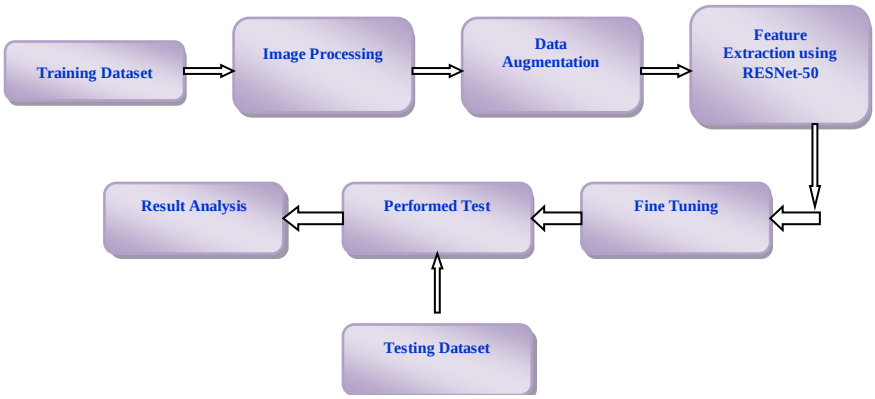

Fig. 1: Work flow for Classification

1. **Training Dataset:** Our proposed system contains the 30 categories of plants as dataset, these are all rare medicinal plants collected by took photographs by using smart phones well as from Kaggle dataset and from google then our proposed model trained by different classes of leaves for classification.

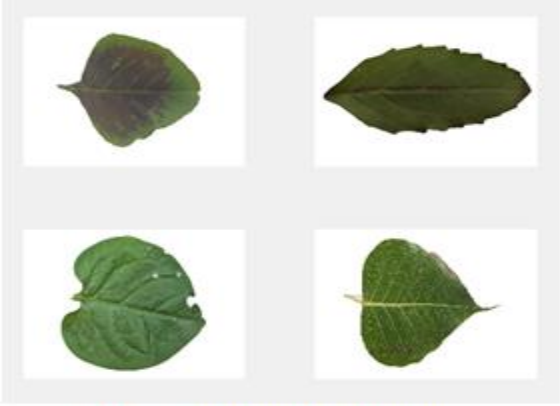

Fig 2: Sample of loaded image

2. **Image Processing and Labelling:** Load the path of the data set, using ImageDatastore, because it operated on image file location. From the path of the dataset, class labels were assigned automatically. The CounEachlabemethod counts the number of images is in each

category. Because each category contains unique number of images, should adjust them, so balance all the images in the training set.

3. **Augmentation process:** The first layer of CNN is input layer. CNN has different types of methods, each method has a different input size requirement, and in this model image input size is 224x224x3. Intermediate layers are a series of Convolutional layers also called hidden layers such as Max pooling layer, feature extraction layer, the last layer is output layer, classification will happened there. In our proposed model classification layer has 1000 different classes from millions of images. It Splits 30% of the images for training data and remaining for validation data by randomly. Proposed model processes only RGB images, so has to convert all the images into this format, to avoid re-saving all the images, it uses augmentedImageDatastore to resize and converts gray scale images into RGB. This augmentation can also used for training the network also for additional data augmentation.

4. **Feature Extraction using ResNet-50:** From the deep learning domain, one of the powerful machine learning techniques is CNN. CNN learns rich collection of features, but CNN requires a large collection of images to train data; these features often outperform features such as HOG, LBP or SURF. Training with large collection of images is not that much easy. However, there is an easy way. We can use pertained CNN to improve the power of CNN. If we use feature extractor of pre-trained CNN, then it saves time and effort efficiently. In this proposed system we used pre-trained CNN Resnet-50 for data processing and Medicinal plants leaf images as data set.

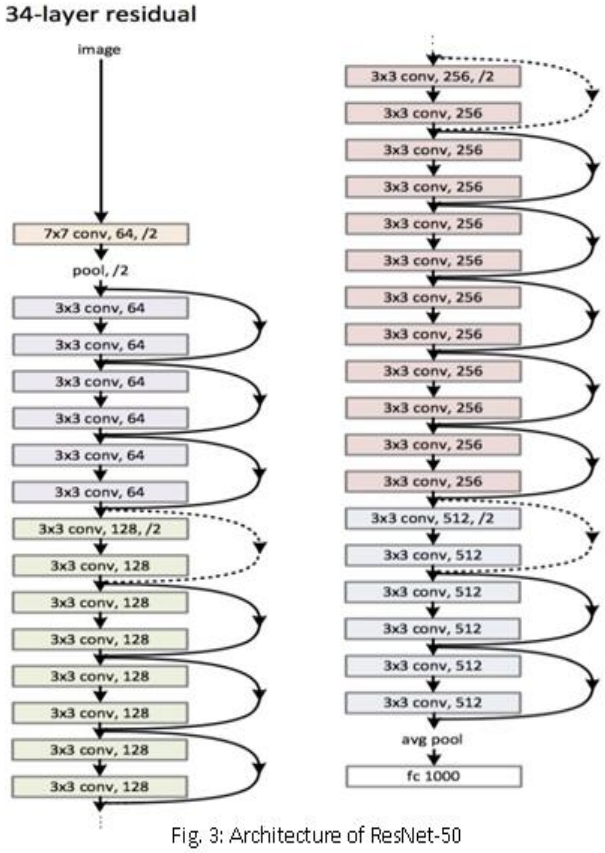

Fig. 3: Architecture of ResNet-50

Image classification using CNN is a very interesting subject. Once you learn how to classify images using CNN, you can use them for diverse collection of images, such as to identify brain tumor or cancer cell or skin diseases. Using CNN in MATLAB for Image classification is not a straight forward approach. But Resnet-50 we used here has made it simple.

There are so many popular pre-trained networks such as AlexNet, GoogleNet, VGG16, VGG19, ResNet-50. Most of these can train 1.2 minions of Images from Dataset into 1000 categories. ResNet-50 is one of such models is CNN contains 50 layers deep. You can load a pre-trained version of the network train more than a million images from the Medicinal Plant leaves database. The pre-trained network can classify images into 1000 object categories, such as Cassia Occidentalis leaves, Achyrantes aspera leaves, boerhavia diffusa leaves, bacopa monnieri leaves. As a result, the network has learned rich feature representations for a wide range of images. The architecture of a ResNet-50 model can be given in the figure 3.

5. **Performance Test:** CNN classifier can be used to train the multiclass classifier. Stochastic Gradient Descent Solver is used to training, it performs fast. High dimensional CNN feature vector can set Learners parameter to Linear in the fitcecoc() function to speed up the training, repeat the same procedure for test set also. Then features can be passed to the classifier to measure the accuracy of the trained classifier. Load new image to test the classification, read test images, Extract features using CNN, make prediction using classifier.

6. **Fine Tuning:** Our proposed method is used to increase the efficiency of the function. Adjustment process is so difficult, if we made small manipulation will also affects a lot in the training process with respect to the computation speed and time. The process of fine tuning should be repeat again and again, and then the accuracy will improve. The accuracy of ResNet-50 is given below.

Table 1: Parameter for fine tuning

Parameter	Value
Validation accuracy	98.9%
Training Finished	Reached final iteration
Epoch	8 of 8
Iteration	248 of 248
Iteration per epoch	31
Frequency	30 iterations
Patience	info
Learning rate schedule	Constant
Learning rate	0.01

IV. RESULT AND DISCUSSION

Proposed model stated training with the training dataset from the augmentation. The validation has been performed to generalize the model. Figure 4 shows the accuracy percentage of the training and Validation phases of the images. Figure 5 shows the loss percentage of the training and validation phases of the images with respect to the iteration (Epoch).

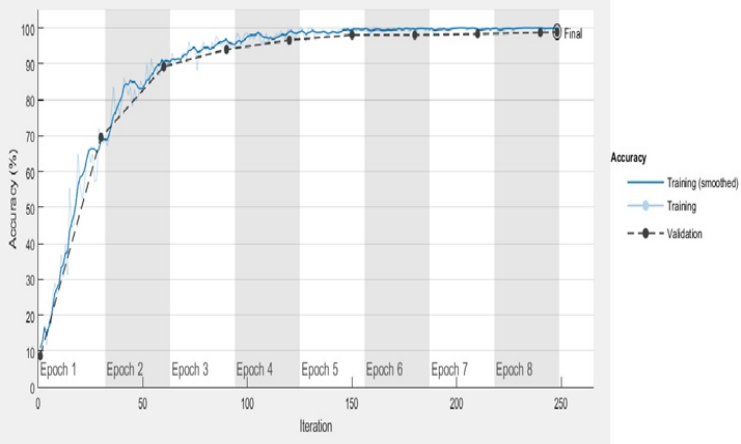

Fig 4: Measurement of Accuracy in Training and validation phase

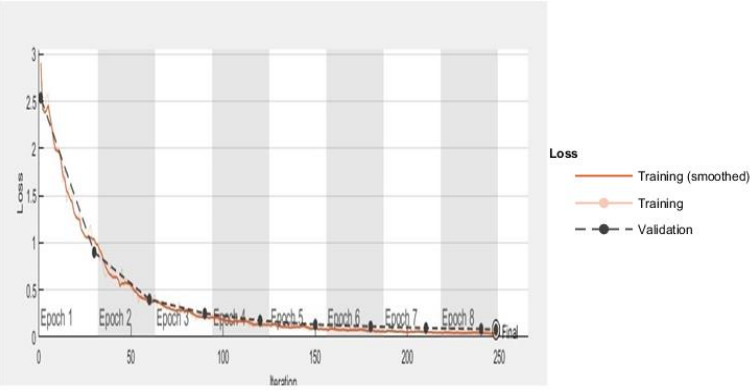

Fig 5: Loss percentage in training and validation phase

In this proposed model, we resized the images by 224x224x3, loaded the ResNet-50, it uses only RGB colours. Spit the dataset into training and test data, 30% for training and remaining for test data, performed Augmentation for resize the images and covert gray scale images into RGB. A fast Stochastic Gradient Descent solver is used for training by setting the fitcecoc function's 'Learners' parameter to 'Linear', batch size was equal to 32. Then test features can be passed to the classifier to measure the accuracy of the trained classifier. Then applied the trained classifier to the new images and make the prediction using the classifier. We have also done experiments with other popular pre-trained models namely AlexNet, VGG16 and VGG19. The figures 7, 8, and 9 are showing the validation and loss percentage of these models. Among all of them, our proposed model achieved better accuracy as well as low loss.

Table 2: Accuracy of the different models

Pre-Trained models	Training accuracy	Validation accuracy	Time used	Loss
AlexNet	96.58	97.9	818.3815 sec	1.37
VGG19	97.11	98.7	700.2429 Sec	1.64
VGG16	97.63	98.4	355.7076 Sec	0.79
ResNet50	**98.42**	**98.9**	**206.6693 sec.**	**0.49**

Algorithm of the proposed system
Input: Leaf Images dataset, new leaf image
Output: Classification of leaf images, identifies the name of the new loaded leaf image
Step 1: Leaf images are taken in different folders
Step 2: Divide those images as train and test set
Step 3: Resnet-50 can be loaded using the resnet50 function from Neural Network Toolbox
Step 4: Extract the features of the images
Step 5: Model obtained after training is fed with the test set
Step 6: Evaluate the results obtained some parameters like accuracy
Step 7: Load any new image, it extracts the features and makes prediction, shows the name of the leaf also obtain the accuracy.

Table 3: Accuracy of the images of different models

S.No.	Alex Net	VGG19	VGG16	Resnet-50
Image 1	99.2	97.62	97.89	99.4
Image 2	98.6	97.36	98.15	97.9
Image 3	98.9	98.15	96.84	100
Image 4	98.9	98.68	97.63	99.5
Image 5	97.6	97.10	96.84	99.2
Image 6	96.3	96.62	98.68	96.7
Image 7	97.3	97.37	98.16	98.3
Image 8	98.6	95.53	95.78	99.4
Image 9	98.4	97.30	97.10	97.5
Image 10	97.11	97.10	97.36	97.8

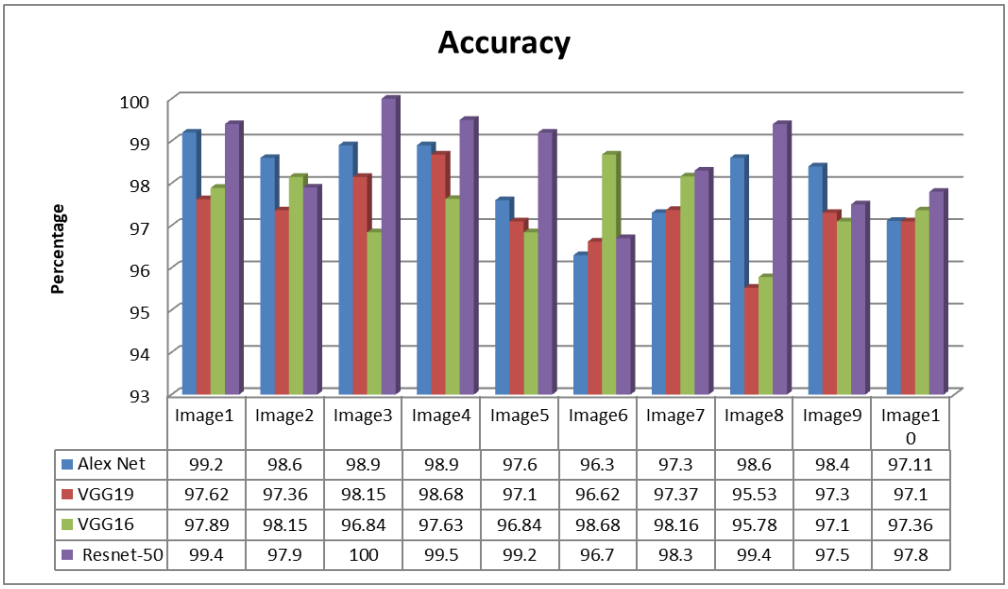

Fig 6: Accuracy Computation

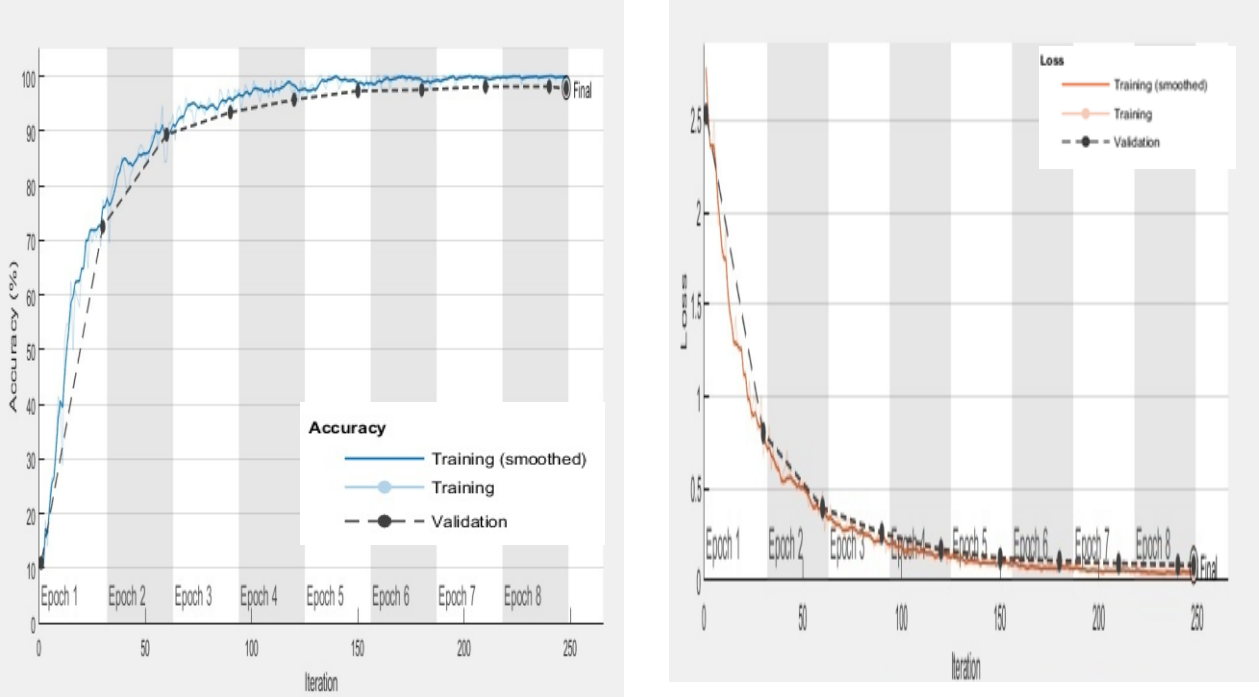

Fig 7: AlexNet Model Validation and Loss

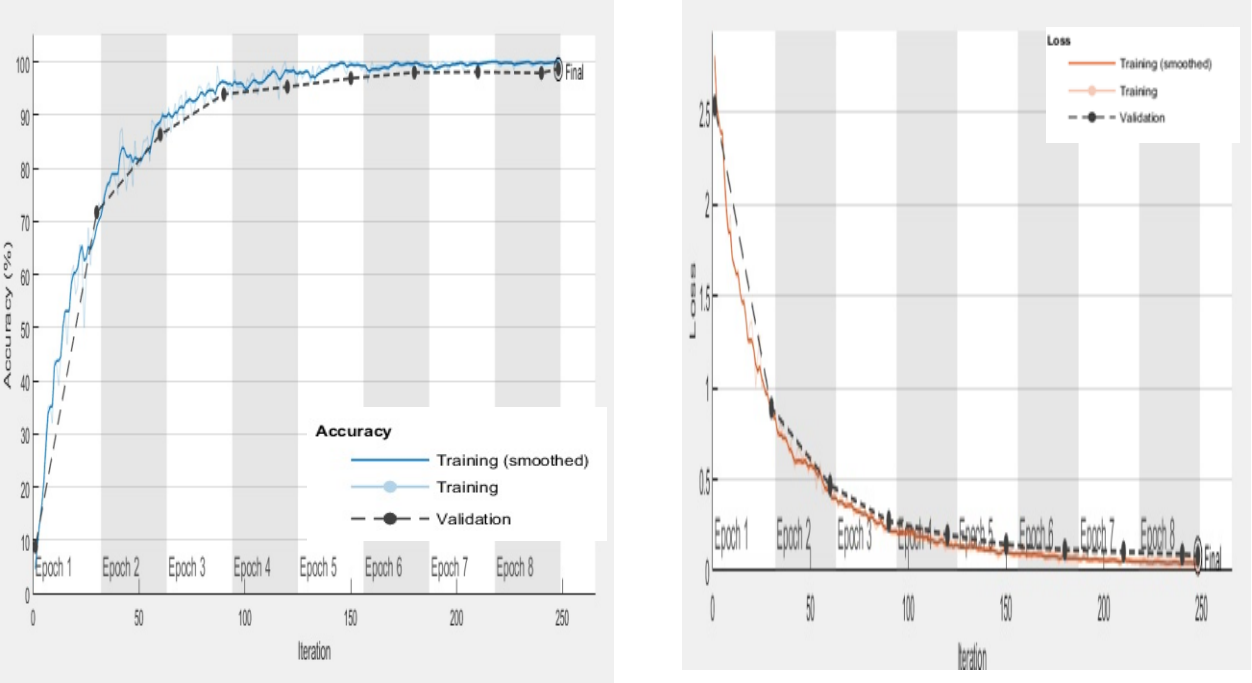

Fig 8: VGG16 Model Validation and Loss

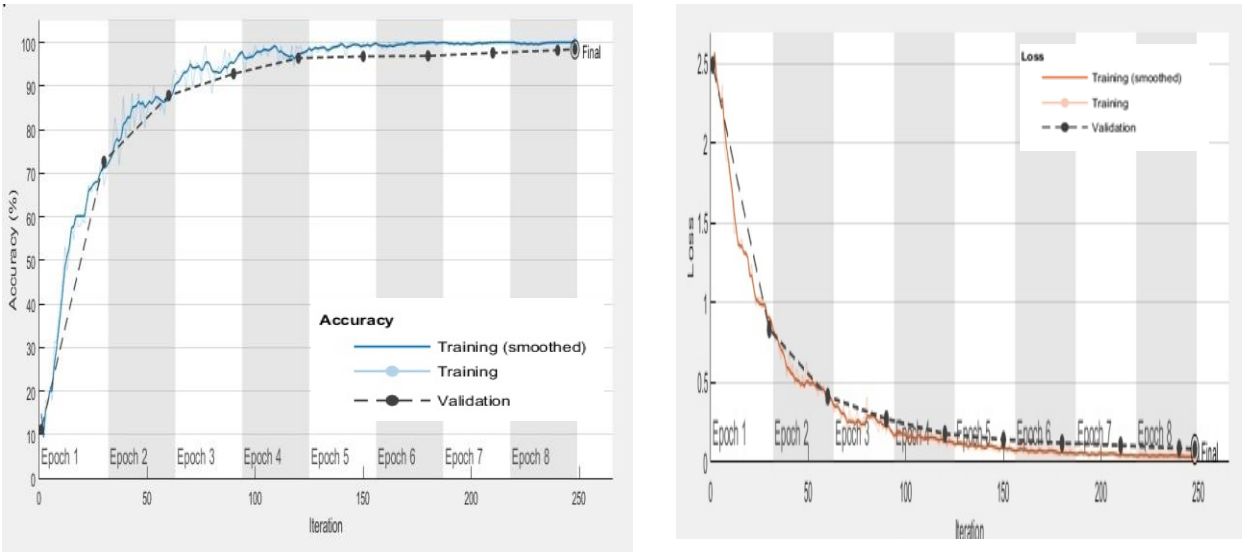

Fig 9: VGG19 Model Validation and Loss

V. CONCLUSION

This work assumes that the proposed methodology a popular pre-trained method ResNet-50 was used as a classifier in the recognition of medicinal plant leaf which provides sensibly high recognition rate when compared with pre-trained models like AlexNet, VGG16 and VGG19. The overall accuracy of proposed model was 98.42. In leaf image processing if we give more images it gives more accuracy because generally the images of the most of the leaves contains similar properties like color, shape etc., in out experiment we used 10 categories of leaf image and the total number of images is 765 so sometimes it is difficult to the network to recognize exact label. If we load hundreds of images in the same model, then there is no doubt we can achieve 100% of accuracy.

REFERENCES

[1] Ishrat Zahan Mukti, Dipayan Biswas (2019). Transfer Learning Based Plant Diseases Detection Using ResNet50, 4th International Conference on Electrical Information and Communication Technology (EICT), 20-22.

[2] Muhammd Hammad Saleem, Johan Potgieter, Khalid Mahmood Arif (2019). Plant Disease Detection and Classification by Deep Learning, Plants, 8(11): 468.

[3] Liwen Gao, Xiaohua Lin (2018). A method for accurately segmenting images of Medicinal Plant leaves with complex backgrounds, Computer and Electronics in Agriculture, 155:426-445.

[4] Soon Jye Kho, Sugumaran Manickam, Sorayya Malek, Mogeeb Mosleh, Sarinder Kaur Dhillon (2018). Automated plant identified using artificial neural network and support vector machine, Frontiers in Life Science, 10(1):98-107.

[5] Xiaolong Zhu, Meng Zhu, HongeRen (2018). Method of plant leaf recognition based on improved deep convolutional neural network, Cognitive Systems Research, 53:223-233. doi: https://doi.org/10.1016/j.cogsys. 2018.06.008

[6] C.H Arun, D. Christopher Durairaj (2017). Identifying Medicinal Plant Leaves Using Textures and Optical Colour Spaces Channel, Journal of Computer Science and Information, 10(1):19-28.

[7] Yu Sun, Yuan Liu, Guan Wang, Haiyan Zhang (2017). Research Article Deep Learning for Plant Identification in Natural Environment, Hindawi Computational Intelligence and Neuroscience, 2017:1-6. https://doi.org/10.1155/2017/7361042.

[8] Wang-Su Jeon, Sang-Yong Rhee (2017). Plant Leaf Recognition using a Convolution Neural Network, International Journal of Fuzzy Logic and Intelligent Systems, 17(1):26-34.

[9] Mostafa Mehdipour Ghazi, Berrin Yanikoglu, Erchan Aptoula (2017). Plant identification using deep neural networks via optimization of transfer learning parameters, Neurocomputing, 235:228-235. http://dx.doi.org/10.1016/j.neucom.2017.01.018.

[10] Adams Begue, Fawzi Mahomodally, Upasana Singh, Samerchand Pudaruth (2017). Automatic Recognition of Medicinal Plants using Machine Learning Techniques, International Journal of Advanced Computer Science and Applications, 8(4):166-176.

[11] Jana Waldchen, Patrick Mader (2018). Plant Species Identification using computer vision techniques: A Systematic Literature Review, Archives of Computational Methods in Engineering, 25:507-543.

[12] Nian Liu, Jiang-ming Kan (2016). Improved deep belief networks and multi-feature fusion for leaf identification. DOI: http://dx.doi.org/10.1016/j.neucom.2016.08.005

[13] Aimen Aakif, Muhammad Faisal Khan (2016). Research Paper Automatic classification of plants based on their leaves, Biosystems Engineering, 139:66-75.

[14] Guillermo L, Grinblat, Lucas C. Uzal (2015). Deep Learning for plant identification using vein morphological patterns, Computer and Electronics in Agriculture, 127:418-424.