



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 6, June 2022

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.118

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)



# Survey on Plant identification for Drug Industry Based on Machine Learning

Namratha Hiremat C<sup>1</sup>, Prof.Md.Irshad Hussain B<sup>2\*</sup>

Student, Department of Master Applications, University B.D.T College of Engineering, Davangere, Karnataka, India<sup>1</sup>

Assistant Professor, Department of Master Applications, University B.D.T College of Engineering, Davangere, Karnataka, India<sup>2</sup>

**ABSTRACT:** Plant species detection is a big help for a wide range of people and groups, such as forestry agencies, ecologists, entomologists, doctors, pharmacological institutions, groups fighting for endangered species, government officials, and the general public. Using traditional keys to figure out what a plant is is hard, takes a lot of time, and can be annoying for people who aren't experts because they use specific botanical terms. This makes it harder for people who are just starting out to learn about species. There may be a growing interest in automating the process of figuring out what species something is today. The idea of computerised species identification is now a reality thanks to technologies like virtual cameras and cell phones, remote access to databases, new ways of processing pictures, and sample recognition. With the help of Deep Learning, you can find out about plants in a completely painless way. This challenge shows an Automatic Plant Recognition system that uses a Convolutional Neural Network to try to build an end-to-end Automatic Plant Recognition system. This device lets people upload pictures of plants, and the device will tell the person if the plant can help keep diabetes in check or not.

## I. INTRODUCTION

Ayurvedic knowledge can help with finding new medicines from plants. The first step in finding a new drug is figuring out what NCEs are. NCEs can be made through chemical synthesis or taken from herbal products through organic hobby-guided fractionation. Herbal products are the source of some of the active ingredients in new medicines and some of their ingredients. For plant-based new drug discovery to begin, the right candidate plants must be found through Ayurvedic knowledge, traditional documented use, tribal non-documented use, and a thorough search of the scientific literature. Frequency analysis of the ingredients in the historically documented formulations and analysis of their Ayurvedic properties can also give an in-depth idea of the prevalence of certain Ayurvedic properties, which can be used to choose the right candidate plants for bioactivity-based totally fractionation.

Combining what we know about Ayurveda with what we know about finding new drugs also means that the way we extract things needs to change from sequential to parallel. Bioassay-guided fractionation of a known plant can also lead to a standardised extract or an isolated bioactive compound that can be used to make a new drug. This method could save money and time while also making it easier to find new drugs. Making a new drug is a complicated, expensive, and time-consuming process. From the time a new drug is found to the time it is used in a hospital, it takes about 12 years and more than a billion dollars in investments. New drug discovery is mostly about finding new chemical entities (NCEs) that can be used to make drugs and have the right medicinal chemistry.

## II. LITERATURE SURVEY

The drug industry, the stability of the environment, as well as the rural areas' ability to be productive and last. Botanists are interested in the different ways leaves have the same traits because it lets them compare plants. Plant Recognition from Photographs is a hard task for computer vision. There are a lot of different types of challenges, and there are a lot of different parts of the plant that need to be identified. There are also a lot of different types of plants in nature, with a lot of variations within each type and few variations between types. Even though research on automated plant taxonomy has led to useful results, these models are still a long way from meeting the needs of a fully automatic ecological surveillance system. Over the past 10 years, many different kinds of studies have been done to find ways to identify plants. Wu et al. [4] have done one of the best studies ever done on how plants identify themselves. From the five most important geometrical parameters, 12 structural features were taken into account. The Principal Component Analysis method was then used to reduce the number of inputs to a probabilistic neural network. In this case, PCA is



used to get rid of the curse of photos having depth. The Flavia information set, which is what they made, had an average accuracy of 90.3%.

(2015), T. Gaber suggested a method for recommending plants that uses 2D photographs of plants. The method of attribute fusion and the process of multilabel classification were used in this programme. The results of the experiments showed that the function fusion method was much more accurate than other single applications. The tests showed how reliable they were at making good suggestions.

A new 426 CNN was made by T. J. Jassmann in 2015. As CNN's non-linearity method, they tried out the new Exponential Linear Unit (ELU) instead of the Rectified Linear Unit (ReLU).

Hulya Yalcin [3] suggested in 2016 that CNN could use picture sequences from smart agro-stations to figure out the shape of a plant. The design is used as a preliminary step to get rid of the picture's properties. Configuration and breadth of the CNN are important points that should be brought up because they affect how well neural network architecture can recognise things. They used 16 different kinds of plants and compared them to other methods. Early results show that the CNN-centered method is better at classifying plants than other methods.

Amala Sabu (2017) shows that identifying all leaves is a hard Computer Vision problem. Botany studies and other parts of society, like medicine, can benefit from a method for getting the leaves off of Ayurvedic plants that works well. See if you can recognise the leaf in the pictures. The study of the different ways to identify leaves and how to group them.

Lee, S.H.,[5] gathered one of the pictures of plant leaves has also been talked about based on the leaf characteristics used as an input and convolution neural network is being used to find patterns for each plant's depth information. CNN was mostly used here to show the characteristics better, and DN (Deconvolutional Network) was used to study leaf organisms more effectively. It makes it easier to recognise plant leaves and how many of them there are.

Ghazi, M.M.,[6] used three models of transfer learning to explain what the different plants are. LIFECLEF 2015 was used to evaluate the Network. These three models got their ideas from GoogleNet, VGGNet, and AlexNet.

Barbedo, J.G. (2018) talked about the analysis of the key factors that affect the architecture and effectiveness of deep neural networks used in plant pathology. The in-depth study of the topic, which shows the pros and cons, will lead to more concrete plant pathology findings.

Shah, Sougatta Singh, et al. [2] have suggested that a leaf classification system based on the dualpath deep CNN could be useful. The main things that the CNN with two channels will do are: i. The shape and texture characteristics are both looked at. ii. The features that were found are optimised for classification.

### III. COMPARISION & DISCUSSION

AUTHOR	TITLE	PURPOSE	ALGORITHM	LIMITATIONS
Angie K. Reyes	Fine-tuning Deep Convolutional Networks for Plant Recognition	Used deep Convolutional Neural Networks (CNNs) to extract features and classify images at the same time. CNN is pre-trained in a large collection of generic web images and to recognize plants with the highest possible accuracy.	Convolutional neural networks	Result is 20% worse due to the absence of plant specific strategies in technique.
Jana Wäldchen	Plant Species	A fuzzy k-nearest	systematic	The identification of

	<b>Identification Using Computer Vision Techniques:</b>	neighbors classifier was proposed. Fuzzy k-NN synthetically considers the congeneric number and the similarity between the k-nearest neighbors and the unknown sample	literature review (SLR)	plants by conventional keys is complex. This creates a hard to overcome hurdle for novices interested in acquiring species knowledge.
Jyotismita Chaki	<b>Classification of Medicinal Plants: An Approach using Modified LBP with Symbolic Representation</b>	Modified Local binary patterns (MLBP) is proposed to extract texture features from plant leaves. The classification is facilitated using a simple nearest neighbor classifier.	Shape Feature Selection Template (SFST)	Storage complexity is high. Computation complexity is high.
Alexis Joly	<b>Multimedia Life Species Identification Challenges</b>	CNN Classifier for video matching. Achieved by grouping the temporally connected video segments classified by the CNN.	FishCLEF	It remains a very timeconsuming process. A content-based approach can then be much more difficult to knowledge-based approaches.
Sue HanLee	<b>How deep learning extracts and learns leaf features for plant classification Working notes of CLEF 2015</b>	Extract useful leaf features directly from the raw representations of input data using Convolutional Neural Networks (CNN). Gain intuition of the chosen features based on a Deconvolutional Network (DN) approach.	Deep learning algorithms for kmeans	Increase the difficulty of the classification problem by constraining the varieties of leaf data to be seen by the CNN during training.
Matthew D	<b>A Review on Plant Leaf Classification and Segmentation Proceedings</b>	They classify the plants based on flowering and associative phenomenon. KNN model take input as the leaf venation morphological	ImageNet benchmark	It was found that this process was time consuming and difficult. Long training time.



		feature and classify them into four different species.		
--	--	--	--	--

Table 1: Comparison of different leaf

#### IV. DATA SET

The known machine learning site called Kaggle The main need for the project is to have an high quality images of the parts of plants. The collection of this datasets manually is a tedious process and sometimes it is an impossible task to collect all these image datasets. The images are collected from.



Figure 1 Image of Alfalfa flower



Figure 2 Image of Alfalfa Leaf

As a result of classifying the plants, the common name, scientific name, medicinal use, place of growth, and features of parts of the plants will be shown. Figure 1 is a picture of an arugula flower. The picture of an Alfalfa leaf is shown in Figure 2. Here are some images that can be used to train and test how well you can classify images. The quality of these images is very good, and none of the information in them has been lost. So, the result will be the best.

#### V. METHODOLOGY

Methodology The pictures of plants in the dataset are pretty different from one another. For example, Amla may have both leaves and fruit, and Neem photographs may show leaves that are different from those of Amla, Lavender,



and Tulsi. Even though Tulsi and lavender may look the same, they are used for different things. All of the pictures are resized so that the version looks the same and doesn't lie to CNN about having different-sized pictures. People know that Deep Learning systems need a lot of training data, so pictures are flipped horizontally to make it look like the model has more training pictures.

Using the Sequential property, the Convolutional Neural Network is set up for the first time. [5] The length of the Convolutional layer is (32, 4, 4), where 32 is the number of characteristic detectors and (4, 4) is the size of each characteristic detector. The input format for pictures is (64, 64, 3), which means that each picture is 64x64 pixels long and has three colour channels (RGB). The "ReLU" activation is given to the Convolutional Layer. The Max Pooling layer of length is the next layer of the CNN (2, 2). Pooling layers are used to make characteristic maps smaller. So, it cuts down on the number of parameters to analyse and the amount of math that has to be done inside the network. In other words, the pooling layer is a summary of the features of the feature map that was made with the help of a convolution layer. The last layer of a CNN is the Flatten layer. This layer is needed to turn the data into a one-dimensional array so that it can be sent to the next layer. We flatten the results of the convolutional layers to make a single long feature vector. And it's connected to the last class version in a way that's called a fully linked layer.

As part of the CNN's hidden layer, 128 neurons with "ReLU" activation are now added to the Dense layers, which are made up of the "Dense layers."

The CNN's output layer is made up of just neurons, since can cure and cant cure are the most useful instructions in our dataset. These are turned on with "Softmax." The "Adam" optimizer is used to connect and put together all the layers of the CNN version. The "categorical crossentropy" is used to measure loss, which lets us put specific facts into groups, and "accuracy" is used to measure the accuracy of the model. The model has been trained for a long enough time to give accurate class results when compared to test facts snapshots. The model is trained for 70 iterations using the whole dataset, which has a success rate of 98.07 percent with almost no errors of less than 0.05 percent.

Once the CNN has been trained for the number of epochs you want, it is put into a model and saved using the joblib library of Python. This saved version can then be loaded into another Python file and used right away to make predictions, without having to do any of the above. For a prediction to be made on test photos, the input photos should be given to the model in the same way that the model found them. In simpler terms, the input photos need to be preprocessed before they are sent to the CNN version as input. This is done with the help of Python's cv2 module. The version gives out 0 if the picture's functions are like those of pictures from the can cure class, and 1 if they are not.

## VI. EXPIREMENTS AND RESULTS

The Indian Gooseberry, also called Amla, is the first plant chosen. It helps keep blood sugar levels steady and gives people a stronger immune system. Holy Basil, also called Tulsi, can help stop glucose from building up in the blood and the problems that come with diabetes. Lavender oil helps diabetics lower their blood sugar and oxidative stress, which can lead to serious problems. [6] Neem, also called Azadirachta indica, is known for its antiviral and antioxidant properties, which help keep blood sugar levels in check.

Author	Technique used	Dataset Size	Accuracy
T. Gaber (2015) [2]	Bagging classifier, 2D based technique	using Flavia a dataset which 1907 colored images.	Accuracy: 95%
T. J. Jassmann (2015) [3]	Rectified Linear Unit (ReLU): is an activation functions and ReLU is the most used activation function in the neural network, moreover in the CNNs	Flavia dataset contains leaves of 32 plants	Accuracy: 60%
Hulya Yalcin (2016) [4]	Convolutional Neural Network (CNN) model.	16 plant species.	Accuracy: 97.47%

Lee, S.H (2017) [6]	CNN	Using 113,205 images	Accuracy: 96.3%
Ghazi, M.M (2017) [7]	Transfer Learning using AlexNet GoogLeNet and VGGN: VGGN is an object-oriented Model and supports 19 layers and VGGN is still the most popular used architecture for image recognition	using 91,758 images.	Accuracy: 80.18%
Abdul Kadir [9]	Commonly, the methods will not capture color information, because color was not identified as an important feature to the identification. In this research, color, and texture, shape and vein, features were combined that are included to classify a leaf.	Using 32 images	Accuracy:93,75%
Habibollah Agh [5]	CNN architecture which is applied for leaf classification. This CNN architecture consists of 5 layers. ReLU/ELU activation function is applied after each convolution layer. MaxPooling function is applied for each pooling layer. The fully connected layers are defined as convolutional layers with the filter size of 1x1	32 and 15 leaf species in Flavia and Swedish leaf	Accuracy:97.24%
Jiachun Liu, S. Yang [6]	This method uses the CNN algorithm for feature extraction and classification. 10 layers are introduced in this CNN architecture. In this method, an augment for leaf was applied to enlarge the database. So that it can improve the classification performance.	Flavia dataset	Accuracy:87.92%

Table 2: Technique to be used

## VII. CONCLUSION

The works that regularly watched CNN to learn about plants. We will also talk about tips and tricks to keep in mind if you want to get the most out of CNNs used in real-world international applications. Many answers that have already been posted are mostly based on CNN's. Herbal medicine is an important part of health care in India, and



medicinal plants are a major source for the whole country. It is important to make sure that they are protected so that they can be used for a long time. This article is about making an easy-to-use Image Recognition tool that can tell the difference between plants that can help keep diabetes away and those that can't. The dataset includes 4 unique medicinal plants like Tulsi, Neem, Lavender, and Amla that have been proven to help control diabetes, as well as other unique plants that have nothing to do with diabetes at all. All of these plants were handpicked from Google images. The CNN version is made to put pictures of plants into two groups: those that can cure and those that can't. This Plant Recognizer has also been put to use at home so that people can use it in a way that is easy for them to understand. For example, a person with diabetes can just upload a picture of a plant and see if it helps control their diabetes because the picture is one that the model has been trained on. After 70 epochs of training, the version is accurate about 89% of the time.

#### **FUTURE WORK**

Transfer learning technique can be used to make the model more accurate.

Plants with very small leaves can be put into groups, and the project can be made to be portable.

#### **REFERENCES**

1. Pešić, M. (2015). Development of natural product drugs in a sustainable manner. Brief for United Nations Global Sustainable Development Report 2015.
2. Bauer, A., Brönstrup, M. (2014). Industrial natural product chemistry for drug discovery and development. *Natural Prod. Rep.* 31 (1), 35–60. DOI: 10.1039/C3NP70058E
3. Gao, W. and Lin, W., 2012. Frontal Parietal Control Network Regulates the AntiCorrelated Default and Dorsal Attention Networks. *Human Brain Mapping*, 33(1), 192–202.
4. Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F. and Xiang, Q.L., 2007. A Leaf Recognition Algorithm for Plant Classification using Probabilistic Neural Network. 7th IEEE International Symposium on Signal Processing and Information Technology, Giza, Egypt, 11- 16.
5. Zhang X., Liu Y., Lin H., Liu Y. (2016) Research on SVM Plant Leaf Identification Method Based on CSA. In: Che W. et al. (eds) *Social Computing, ICYCSEE 2016*. Communications in Computer and Information Science, Vol 624, Springer, Singapore.
6. Hossain, J. and Amin, M.A., 2010. Leaf Shape Identification Based Plant Biometrics. 13th International Conference on Computer and Information Technology, Dhaka, Bangladesh, 458-463.
7. Du, J.X., Wang, X.F. and Zhang, G.J., 2007. Leaf shape-based plant species recognition. *Applied Mathematics and Computation*, 185, 883-893.
8. Du, M., Zhang, S. and Wang, H., 2009. Supervised Isomap for Plant Leaf Image Classification. 5th International Conference on Emerging Intelligent Computing Technology and Applications, Ulsan, South Korea, 627-634.
9. Herdiyeni, Y. and Wahyuni, N.K.S., 2012. Mobile Application for Indonesian Medicinal Plants Identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram. *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, West Java, Indonesia, 301-306.
10. Arai, K., Abdullah, I.N. and Okumura, H., 2013. Identification of Ornamental Plant Functioned as Medicinal Plant-based on Redundant Discrete Wavelet Transformation. *International Journal of Advanced Research in Artificial Intelligence*, 2(3), 60-64.
11. Prasvita, D.S. and Herdiyeni, Y., 2013. MedLeaf: Mobile Application for Medicinal Plant Identification Based on Leaf Image. *International Journal of Advanced Science, Engineering and Information Technology*, 3, 5–8.
12. Le, T.L., Tran, D.T. and Hoang, V.N., 2014. Fully Automatic leaf-based plant identification, application for Vietnamese medicinal plant search. *Fifth Symposium on Information and Communication Technology*, Hanoi, Vietnam, 146-1514.
13. Bhatt K.K.S. 1995. *Medicinal Plant Information Databases. Non-wood Forest Product* FAO – Food and Organisation of the United Nations, Viale dell di Caracalla, 00100 Rome, Italy.
14. .Flavia medicinal plant species dataset, website: <http://flavia.sourceforge.net/>, accessed on 01/01//2018.
15. *A Hand Book of Medicinal Plants – A Complete Source Book*. Agrobios (India), Jodhpur. Purohit S.S. and Vyas S.P. 2004.





**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.118**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details