PLANT RECOGNITION FOR DRUG INDUSTRY USING CNN

Pratiksha Bongale[1], Bhumika K M[2], Chaitanya K B[3], Teju G M[4], Ayesha Sahareen Shaik[5], Waseem Khan[6], Dr. Roopa G M[7], Dr. Nirmala C R[8]
Department of Computer Science and Engineering
Bapuji Institute of Engineering and Technology, Davanagere, India

Abstract.

Plant species detection offers significant advantages for a diverse range of sectors, including forestry agencies, ecologists, entomologists, doctors, pharmacological institutions, groups campaigning for critically endangered, governmental authorities, and the general public.The identification of plants by conventional keys is complex, time-consuming, and due to the use of specific botanical terms frustrating for non-experts. This makes it harder for novices interested in acquiring species knowledge. Today, there is an increasing interest in automating the process of species identification. The availability and ubiquity of relevant technologies, such as digital cameras and mobile devices, remote access to databases, new techniques in image processing, and pattern recognition let the idea of automated species identification become reality. With the help of Deep Learning, one can identify plants in a very hassle-free manner. This paper showcases an Automatic Plant Recognition system that aims at developing an end-to-end Automatic Plant Recognition System using a Convolutional Neural Network. This system allows users to upload images of plants to the system and the system will tell the user whether the plant can help maintain diabetes or not. This automatic recognizer can be a very useful tool for institutions that are involved in making medicines that can help maintain diabetes. For instance, if a user uploads an image of Holy Basil, popularly known as Tulsi in India, the system will classify it as a plant that can help maintain diabetes and if the user uploads an image of silver maple leaves then the system will classify it as a plant that cannot help maintain diabetes.

Keywords: Computer Vision, Image Recognition, Convolutional Neural Network.

1 Introduction

AI (ML) is arranged under the counterfeit insight of (AI) which encourages the PC with proficiency to perform and learn even after not being especially customized. ML is a methodology for data assessment that robotizes consistent model structure. Machine learning has made all the difference in our lives. It has become the source of our everyday interactions and deals with issues like a small phone call to a substantial industry! For countless generations, herbal medicines have been a foundation of treatment in indigenous people throughout the world. It is still important today as a primary care strategy for about 85 percent of the planet's population [1] and as a reservoir of pharmaceutical research, with 80 percent of all synthetic pharmaceuticals coming from it [2]. Similarly, the inception, evolution, and refinement of natural component assessment have seen a rapid surge in the last few centuries.

Countless species of plants exist on the planet, many of which have therapeutic value, while some are endangered while others are dangerous to humans. Plants are not only necessary for humans, but they also serve as the foundation for all nutrient cycling. It is critical to understand and categorize species appropriately to utilize and conserve them. A knowledgeable botanist's innate knowledge is essential for detecting different species. A man-made strategy morphology-based traits have proven to be the most effective way of accurately and quickly identifying plants.As a result, several of the procedures associated with identifying these medicinal plants are "reliant on the human being's capability of understanding the medicinal properties of plants" [3]. Human acknowledgment, on the other hand, is a cumbersome and arduous operation. As a result, a large number of researches have been published to substantiate the automatic detection of various plants based on their physical properties [4][5]. Although the methods are very identical, the methods produced up to this point have used a different sequence of iterations to computerize automatic plant categorization.These procedures are fundamentally the preparation of gathered leaves, pre-processing of the leaves to determine their specific characteristics, leaves categorization and database compilation, identification learning, and performance evaluation. Any part of the plant can be used for image recognition and not the leaves always. The flowers, trunk, seeds, and other parts of a plant can be considered for the image recognition process. Non-botanical professionals can use the proposed automatic image identification model to easily recognize species of plants.

2 Literature Review

During the past decade, multiple kinds of research were done to come up with methodologies for plant recognition. Wu et al. [4] have carried out one of the most remarkable research in the area of plant identification. Twelve structural characteristics have been considered out of five fundamental geometrical parameters, and the Principal Component Analysis method is then utilized for dimensionality suppression to minimize the volume of input to a probabilistic neural network. PCA here is used to get rid of the curse of dimensionality of images. The Flavia data set, which is their production, attained an average accuracy of 90.3 percent.

Researchers in [6] obtained a comparable degree of accuracy with identical characteristics leveraging a distinct data set while keeping the same predictor. Du et al. (2007) achieved 93 percent using the k-nearest neighbor classification method [7], which employed comparable functions but unique data with only 20 species. Du et al, (2009) have achieved a precision of 92.3 percent on 2000 photos encompassing 20 different leaf kinds using a new distance metric termed 'isomap' [8]. The authors in [9] went to utilize a combination of fuzzy local binary series and fuzzy color histogram and a probabilistic neural network (PNN) forecaster on a database of close to 2.5k images of leaves extracted from a variety of plants having medicinal values from the dense forests of Indonesia to attain a categorization precision of around 75 percentage. In 2013, Arai et al. [10] utilized the disjunct wavelet transformation to take up the transfer of the invariant characteristics across eight distinct decorative plant species in Indonesia. They came up to an accuracy level of 95.8 percentage by leveraging Support Vector Machines. The proposed technique has already been evaluated on 55 Vietnamic medicinal plants, with a classification support vector machining (SVM) of very high accuracy of 98.3 percent. Thecomparable smartphone application, based on earlier research, was built by Prasvita and Herdiyeni [11].

An entirely computer-controlled plant classification model has been designed using the Kernel Decriptor (KDES) as a novel characteristics mining technology by Le et al. (2014) [12]. With the k-nearest neighbor classification of 20, a higher classification rate of 87.1 percent can be achieved. The authors in [13] were able to control conventional methodologies based on Gabor filters and Fourier analysis with the use of volumetric fractal aspect techniques to build a textural imprint for a leaf and the Linear Discriminant Analysis Algorithms (LDA).

3 Dataset

The dataset consists of handpicked images of two categories of plants from Google images. One of the above two contains images of plants like Tulsi, Amla, Neem, and Lavender. These are a few plants that help maintain diabetes in people. The other class of images contains images of Silver Maple Leaves and Eastern Cottonwood that do not help in maintaining diabetes in diabetic people at all.

Fig. 1. The above images are from the can_cure class beginning with Amla, followed by Neem, Lavender, and Tulsi at the end from left to right order

.

Fig. 2. The above images are from the cant_cure class with Eastern Cottonwood on the left and Silver Maple on the right-hand side.

The can_cure category containing Tulsi, Amla, Neem, and Lavender has a total of 414 images in it. These 414 images are then divided into the training and testing datasets respectively. The cant_cure class containing Silver Maple Leaves and Eastern Cottonwood contains 153 images. These 153 images are then divided into the training and testing datasets respectively.

4 Methodology

The dataset consists of plant images that are quite different from each other. For instance, Amla would have leaves and fruit together and Neem images would have leaves that are different from those of Amla or Lavender and Tulsi. Tulsi and Lavender might look similar but have different features. To not mislead the CNN with different image sizes, all the images are resized to help the model produce a consistent result. As it is known that Deep Learning methods involve a lot of training data, images are flipped horizontally to create an illusion of more training images to the model.

The Convolutional Neural Network is initialized using the Sequential( ) function. The Convolutional layer of size (32, 4, 4) is built where 32 is the number of feature detectors, (4, 4) is the size of the feature detector. The input shape of images is set to (64, 64, 3) where each image is of size 64x64 pixels and they have 3 color channels (RGB). The Convolutional Layer is given “ReLU” activation function. The next layer of the CNN is the MaxPooling layer of size (2, 2). Pooling layers are used to reduce the dimensions of the feature maps. Thus, it reduces the number of parameters to learn and the amount of computation performed in the network. In other words, the pooling layer summarizes the features present in a region of the feature map generated by a convolution layer. The final layer of the CNN is the Flatten layer that is required in converting the data into a 1-dimensional array for inputting it to the next layer. We flatten the output of the convolutional layers to create a single long feature vector. And it is connected to the final classification model, which is called a fully connected layer. The Dense layers are added now consisting of 128 neurons with “ReLU” activation functions as part of the hidden layer of the CNN. The output layer of the CNN consists of just two neurons as we have only two classes in our dataset, can_cure, and cant_cure. These are given “Softmax” activation.

All the layers of the CNN model are now attached and compiled using the “Adam” optimizer. The loss function used is the “categorical_crossentropy” that allows us to classify categorical data and the metric used for the measurement of the accuracy of the model is “accuracy”. The model is trained for enough epochs to provide good classification results when tested with test data images. The model is trained for 70 epochs through the whole dataset. At the end of the 70th epoch, the model has achieved an accuracy of 98.07% with nearly zero loss of 0.05%.

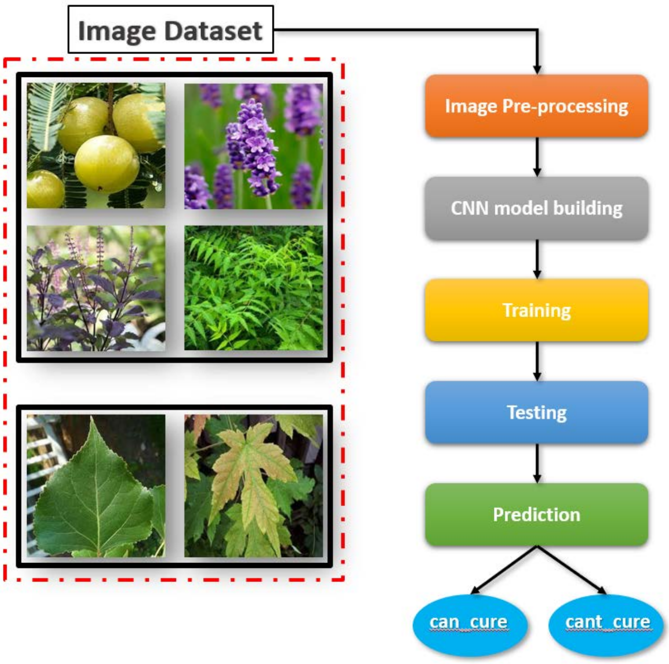

Fig. 3. The flow diagram.

Once the CNN is trained for the required number of epochs, it is fit into a model and saved using the joblib library of python. This saved model can then be loaded into any other python file and can be used for predictions directly without the need for any of the above steps mentioned. For a prediction to be done on test images, the input images are supposed to be passed to the model in the same way that the model has learned them. In simpler words, the input images are to be preprocessed before they are sent as input to the CNN model. This is done using the cv2 module in python. The model outputs 0 if the image has features that resemble features of images from the can_cure class and returns 1 otherwise.

5 Results

Fig. 4. Home page of Plant Recognition for Drug Industry.

In figure 4, the first plant chosen is the Indian Gooseberry or Amla. It helps maintain blood sugar levels and boosts immunity in humans. Holy Basil or Tulsi helps prevent glucose in blood and complications caused due to diabetes. Lavender oil helps reduce blood sugar levels and oxidative stress in diabetic patients that might lead to major complications. Azadirachta indica or Neem is known for its antioxidant and anti-viral properties that help in keeping blood sugar levels under control.

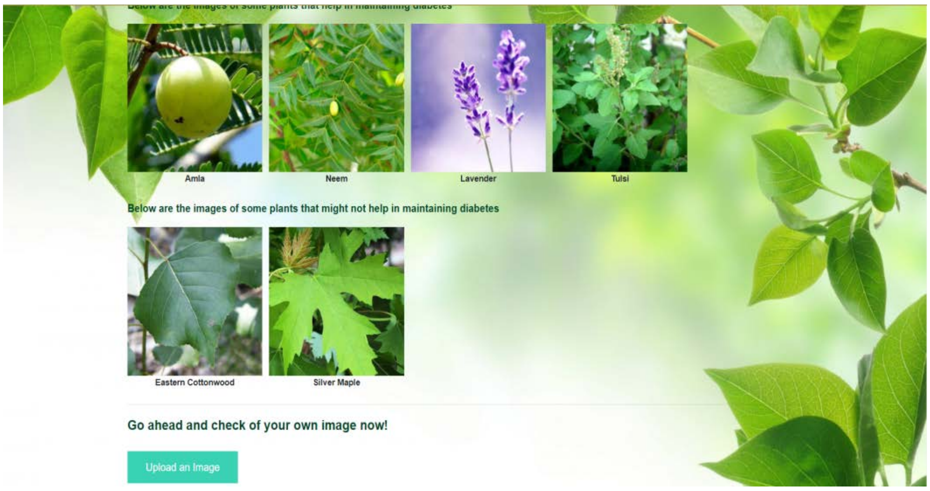

Fig. 5. Eastern cottonwood and silver maple have no connection with diabetes management.

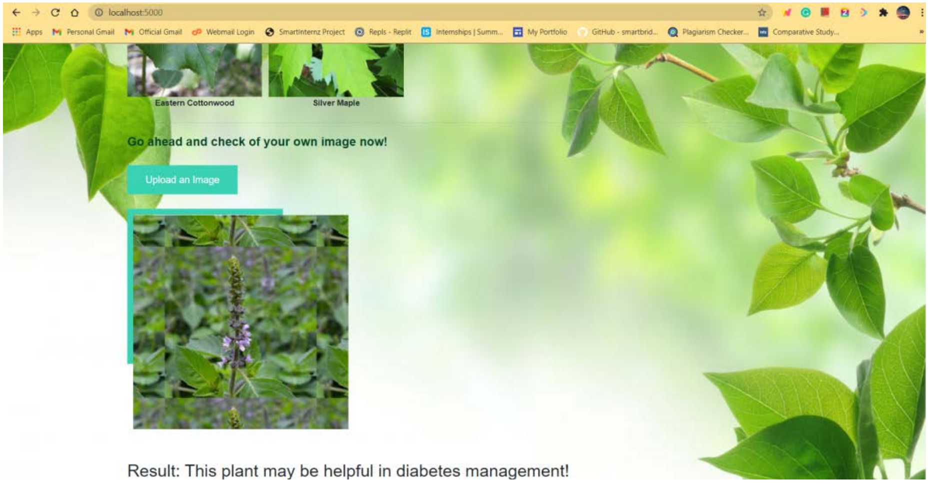

Fig. 6. User-uploaded image of Tulsi plant that helps in maintaining diabetes.

The system processes the uploaded image and renders the output back to the screen displayed to the user. This way the user can easily know which plant is useful for diabetes and which are not

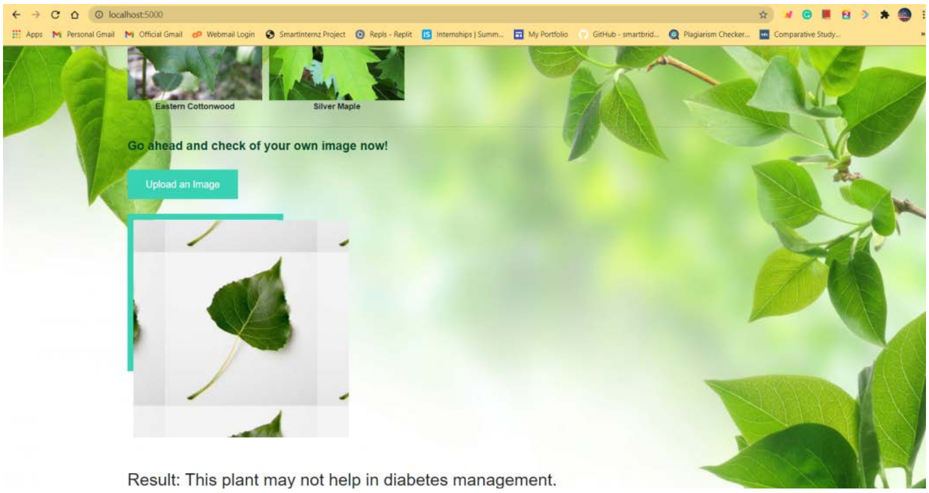

Fig. 7. User-uploaded image of Eastern cottonwood leaf that does not helps in maintaining diabetes.

6 Conclusion

This article aims at building a convenient Image Recognition system that can efficiently categorize different varieties of plants that can help maintain diabetes and those that cannot. The dataset consists of four different medicinal plants including Tulsi, Neem, Lavender, and Amla that are proven solutions to diabetes management and two different plants that have no connection at all with diabetes, all of which are handpicked from Google images. The CNN model is built to classify plant images into can_cure and cant_cure classes. This Plant Recognizer has also been deployed locally for users to use it in a user-friendly manner; wherein the user just has to upload a plant image and see if it helps manage diabetes, given that the uploaded image is one of those that the model has been trained with. The model gives around 89% accuracy after training it for 70 epochs. The number of epochs can be significantly reduced by collecting more images of the plants.

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 Anti-Correlated 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 Technolgy 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 Journalof 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, 146-154.