Leaves classification using Recurrent Neural network based on the edge features

Shubham Vijay Bargal, computer dept, PVPIT college, Pune
Sanket Vinayak Chavan, computer dept, PVPIT college, Pune
Amey Sanjay Ghongade, computer dept, PVPIT college, Pune
Gorakshnath B. Jagdale, computer dept, PVPIT college, Pune
Prof.(Mrs.) R. V. Gawande,computer dept, PVPIT college, Pune

Abstract

Having information about the trees and plants is as important as of having information about the peoples. This may help us to maintain a database of the plants and trees, which can be helpful in many different processes like medicinal search, plant census and many more. Most of the times identifying a tree or a plant by its name may be difficult due to their complex and lengthier botanical terminology. So it is always an advantage to know about trees and plants by the shape of their leaves. As leaves are the most intricate features of trees that are like fingerprints of the trees, So considering leaves for the uniqueness of the tree census and other processes. By taking the advantage of the uniqueness of the plant leaves shape proposed system put forwards an idea of classifying plants leaves based on the morphology and Recurrent neural network. This technique is catalyzed by the using of Edge feature estimation and fuzzy classification model.

Keywords: *Recurrent Neural network, Edge Detection, Fuzzy Classification , Correlation.*

I. INTRODUCTION

Recurrent Neural networks are powerful and robust applications of neural networks that are predominantly used in applications ranging from speech synthesis and image processing. They are a class of networks belonging to the family of Neural networks. They are a subset of a large subsection of Neural network and was first envisioned in the late 1980s, but was not utilised until very recently as the computational demands could not be met by the processors in that era.

Neural Networks are computational networks that are inspired by the human brain. The human brain is the seat of knowledge and consciousness in humans and has enabled us to reach new heights and has got us from cave dwellers to the masters of our planet. The brain has immense learning potential and can solve extremely complex problems. Therefore, the neural networks are modelled after the human brain which has the neuron as its basic cellular unit.

Similarly, the neuron is replicated in the neural networks. The neuron is capable of being activated when the stimuli it receives, exceeds a certain threshold, this concept too has been utilised in the neural networks and various parameters have been isolated and used as the stimulus. When activated, the neuron fires an electrical signal through the neuron into the adjacent neurons in the brain. This allows for a pattern in the firing of the neurons that can be quantified and utilised as output.

There are about typically a billion neurons in a healthy brain. Each of this neuron is capable to responding to stimulus and collectively their firing patterns are recognised as a memory. This is very useful in the neural networks as it helps identify certain patterns and occurrences that could only be noticed by the human eye or more precisely comprehended like a human brain. Neural networks are really useful as they allow us to get a human level understanding from a network which is very valuable and can reduce the load on humans so that their time can be utilised for something much more productive.

Recurrent Neural networks are a type of neural network that works sequentially. That would mean that the recurrent neural network is capable of learning the problems and providing a deeper understand of the problem. And this is due to the fact that, to enable a sequential algorithm, some memory is required to remember the input that has been fed to the machine earlier. This enables a far greater degree of understanding the system that can be utilised for extensive deep learning applications.

As the progresses in the silicon industry have been exponential in the past couple of year and has followed the Moore's law diligently. This has led to an unprecedented increase in the computing power of computers as a whole. This has enabled a far more ludicrous usage of the increased computing power to power the Recurrent Neural Networks as they are quite computationally extensive. This has led to the development of LSTM (Long Short Term Memory), these special breed of RNNs are capable of proving memory and the computational power for the RNNs to generate an output.

This research paper dedicates section 2 for analysis of past work as literature survey, section 3 deeply elaborates the proposed technique and whereas section 4 evaluates the performance of the system and finally section 5 concludes the paper with traces of future enhancement.

II. LITERATURE SURVEY

This section of the literature survey eventually reveals some facts based on thoughtful analysis of many authors work as follows.

Alagukumar [1] Associative classification is one of the most effective methods of the data mining technique.Associative algorithm is depended on Classification based Association (CBA), Classification based on the Predicted Association Rule (CPAR), Multi-class Classification using Association Rule (MCAR). In this paper they have used four phases of Statistical Gene filtering, Discretization, Class Association Rules and prediction or class assignment.The result of this compares the breast cancer data which are available on the NCBI biological database. Thus this helps to detect the disease.

Linking Zhang [2] Generally the leaf detection is one of the challenging part of agriculture automation systems due to variation in the shape of leaves and the leaving plants changes there pose. In this paper the leaves detection is done by examining the vein of the leaves. There is SKEDET method which extracts the skeleton of the leaves is used. In this leaf which is last for longer time is taken as main leaf and then it is compared with leaf to check the thickness of the leaf. This experiment is tested on sweet potato and it is successfully implemented

Y. Cao [3] Botanical studies place an important role in leaf detection studies. It is very hard to detect the leaved veins from the edges of the leaf. Vein extraction and the angle measurement is proposed in this paper. There are basically three approaches first to extract the foreground and the background to get the proper images of the leaves and the second is to change the color RGB to HSI and last is to get results by using the Otsunalgorithm.Thus, this experiment can be done even after the image is very blurry.

M. Sardogan [4] In recent years there has been tremendous growth in agriculture field for the betterment of the framers. This experiment is done on the tomato leaves by taking 500 images of the tomato with the 4 symptoms of diseases. Convolutional Neural Network (CNN) model and Learning Vector Quantization (LVQ) algorithm to detect the disease. They have used CNN for the feature extraction and for the classification. The result of this validation and the proposed method effectively recognizes four different types of diseases

K. Fukuta [5] The early detection of diseases is can help you to predict or to cure it from the disease. In this paper Discriminant Power Score (DPS) approach is used to build the informative genes which evaluate the efficiency of class discriminant. There have used the dataset for leaf from (ALL/AML, ALL/MLL and MLL/AML), in a public database. Then the guns with the DPS are compared with each other and disease is detected using this methodology. The accuracy rate of this method 100 % accurate.

A. Codizar [6] As there are n number of plants and the different type shape of the leaves some of these leaves are unknown some of these known leaves. There has been research on different plant species in which system can identify the plant quick and accurate system due to this people those who don't have knowledge of plant species it helps them to study them on their own. There is an application called LeaVes which based on the leaf shape and identification through their veins by help of machine learning and Image processing system

M. Mat [7] There are many applications developed to detect the disease of different plants this is possible only when we extract the needed content from the images given. There is an approach called as Photogrammetry this approach is used to extract the spatial part from the image. With this approach, there is one more method is used called as Ringed Automated Detection (RAD) is using this. RAD automatic detect the matching points from the multiple photographs. This method is better than the traditional one. This approach, called RAD is one of the best approach, execute with less error and the complex time.

Manojkumar. P [8] Plants play an important role in the life of humans begins. Some of the medicinal plants are used for the making medicine in an Ayurvedic medicinal plant. There are some important features to identify the plant by the leaf shape and the color. In this paper their identification of the plants from both the ends that is from the front side and also from the back side. There is already a database created with images of the plants which are used in medical form later, these images is compared with the new images result of these decide the accuracy rate of the images and it is achieved above94%.

X. You [9] In this paper multiscale crossing representation of leaf identification is used to identify the plant trough their leave by color and venation in this there she used three database set that is the soybean cultivar leaf of the dataset, Flavia leaf dataset and the last one is Swedish leaf set dataset. The methodology used in this paper is one of the successful resulted method, then the other method previously implemented thus this indicated the solution for a leaf identification problem.

Karthik.G [10] The banana is one of the most valuable plants in India. The banana plant can easily get viral disease this disease will come to notice in the fifth month of banana lifecycle. After that it will spoil the banana plant and also retaining of the banana of the banana fingers.There is camera called Embedded Linux development board which can identify the viral disease of the banana leaves. There threshold value which holds the value of healthy banana leave when the image is captured the value of the captured image is compared with a healthy value if the value exceeds there is a sign of disease presence

Arya M S [11] Agriculture is one of the important part of India where seventy percent of our population depend on agriculture, When more than half of our country is depended on the agriculture, then identification of disease in plant becomes important it plays an important role in the farmers life. Identification of the disease can be done only by using a concept called Image processing toolbox of Matlab is used determine the color differences of the plant. This can be extended till the symptoms of any other plant disease. Thus reduce the monitoring plants of farms at a very early stage of disease

M.Merchat [12] National fruit of India is mango, the leaves of the mangos are affected by shortage of nutrient such as iron , potassium and copper Thus the shortage of nutrients can cause the change of color in the mango leaves . By using the Image processing the method there is dataset created of leaves mango. These datasets later divided in clusters by using unsupervised machine learning then this cluster helps to detect various deficiencies which will help or detection of disease.

Z. Husin [13] This paper discusses how to protect the chilly plant from the different attacks and the different type of disease of the leaves. Image is captured of leaf to check their health status. There has been image captured and the dataset is created in the database. There are different type chemicals are pesto the plants so that they should damage from the different bacteria or from insects this can cause the shortage nutrients due to pest. Then this image is compared with the dataset where hundred of images are stored by using the image processing technique. By this the monitoring of plantis done and detection of disease is done.

III PROPOSED METHODOLOGY

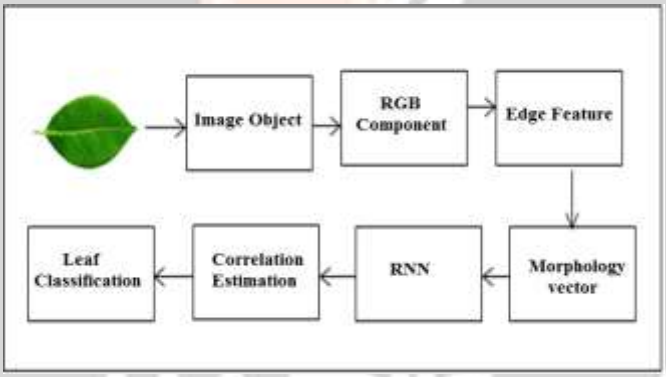

Figure 1: Overview of the Proposed Methodology

The proposed methodology for leaf classification is pictorially depicted in figure 1. The stages that involved in the proposed model for leaf classification is narrated in depth with the below mentioned steps.

The complete proposed approach contains mainly two major tasks. 1) Leaf image indexing 2) searching.

1. Leaf Image Indexing - This is the primitive steps of all, here many images of a leaf are collected in a folder by the name of leaf to feed the system for learning purposes through indexing. This indexing includes some steps as mentioned below.

1.a. Resizing through Graphics- As the image folder is fed, then each and every image are selected as a file and then read in the form of the image objects. Once these image objects are properly buffered, then they are subjected to resize based on the graphical object in the size of 170 X 170.

1.b. Edge Formation - This resized image is now subjected to form the edges. Before this step begins, every image need to convert in the average grayscale channel. By doing this, the edges of the images will be more enhanced. Gray scale conversion is done based on averaging the color model of the pixel into one entity as shown in the below mentioned algorithm 1.

ALGORITHM 1: Average Grayscale Conversion

// Input: Input Image IN_{IMG}
//Output: Grayscale image GRY_{IMG}
// avgGrayScale_BinaryConversion(IN_{IMG})
1: Start

2:avg[766]= Ø
3: ***for*** i=0 TO 765
4:avg[i]=i/3
5:***end for***
6: GRY_{IMG}= Ø
7: ***for*** i = 0 to size of Width of IN_{IMG}
8: ***for*** j=0 to size of Height of IN_{IMG}
9: P_{SIGN} = $IN_{IMG\ (ij)}$ RGB
10: R= P_{SIGN} >> 16 & H_D
11: G= P_{SIGN} >> 8 & H_D
12: B= P_{SIGN} >> 0 & H_D
13: V=(R+G+B)
14: K=avg[V]
15: SET $GRY_{IMG\ (i,j)}$ RGB → (K,K,K)
16: ***End for***
17: ***End for***
18: **return** GRY_{IMG}

Edge Formation process contain some steps as mentioned below

- Smoothing- Here the image is blurred by adding the kernel values of all other pixels towards the pixel nearer to the center. By doing this the edges of the objects in the image are widening.
- Gradient estimation- Here the kernel value for the bigger range is evaluated based on some threshold.
- Non- maximum Suppression - Edges with the maximum width is labeled.
- Double Thresholding - Labeled threshold edge pixels are increased twice by their size.
- Edge Tracking- All the edge pixels are traversed and label them in white pixels and remaining all into black pixels.

1.c . Recurrent Neural Network - Morphology vector formation - First Layer Evaluation- This is the core part of the system where the morphology of the each of the leaf images will be estimated and stored in a file to classify the same during the searching process. So this step takes the input of the edge images, then the pixel position with respect to the vertical top alignment is estimated to store in an array. This array eventually represents the shape of the leaf. This array is written in a text file with the image name. This correlation array or morphology array formation can be shown in the below shown algorithm 2.

ALGORITHM 2: Morphology Vector

// Input: Input Edge Image E_{IMG}
//Output: Morphology Array M [ ]
// morphologyVector(E_{IMG})
1: Start
2: ***for*** i = 0 to size of Width of E_{IMG}
3: index=0
4: ***for*** j=0 to size of Height of E_{IMG}
5: P_{SIGN} = $IN_{IMG\ (ij)}$ RGB
6: R= P_{SIGN} >> 16 & H_D
7: G= P_{SIGN} >> 8 & H_D
8: B= P_{SIGN} >> 0 & H_D
9: ***IF*** (R=255 AND G=255 AND B=255)
10: index=j
11 **BREAK**
12: ***End for***
13: Ratio=index/ Height of E_{IMG}
14: M[i]=Ratio
15: ***End for***
16: **return** M[]

All the input images that are fed to the system in a folder are estimated for their morphology vector estimation to write in a text file with the name of the input leaf.

2.a - Searching- RNN Deep Layer- Here in this step a single leaf image is fed to the system to search with all other leave images to classify properly according to its name. Again, this step follows all the step till morphology vector formation. Then formed morphology vector is correlated with the stored one for each of the leaves to estimate the biggest correlation values using the Pearson Correlation equation as mentioned below in the equation 1. Then finally the image morphology file which is having the highest correlation value with the given image is identified as the searched image name. And it is classified with its name to display its image along with the Medicinal name, common name and medicinal property description to help the user to get more information about the unknown input leaf.

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{(x^2 - \frac{\sum x^2}{n})}\sqrt{(y^2 - \frac{\sum y^2}{n})}} \quad \text{_____(3)}$$

Where
x is the Morphology array of input image
y is the Morphology array of stored image
n is the array Size
r= correlation value in between -1 to +1.

IV RESULT AND DISCUSSIONS

The proposed methodology for the detection and classification of leaves has been deployed on a machine running a Windows Operating System. The implementation machine is powered by a Core i5 Processor paired with a 6GB of Random-Access Memory. The presented technique has been coded in the Java Programming language on a NetBeans Integrated Development Environment. For the storage purposes, MySQL Database server has been utilized. To ascertain the efficiency and performance of the proposed technique needs to extensively experiment on. Mean Absolute Error is one of the most commonly used techniques that can ascertain the accuracy of a system. Mean Absolute Error has been utilized primarily for the evaluation of the errors in the system and the extent of these particular errors.

The Mean Absolute Error evaluates the performance in terms of percentage, which is highly useful and easy to understand the result. The Parameters used in this method are continuous in nature which is useful for the system. Therefore, the MAE is one of the most robust and effective techniques for the evaluation of the performance. The error percentage of the proposed methodology is evaluated through the MAE for the number of accurate Leaf classifications performed. The absolute difference between the number of trials and the accurate classifications is utilized for the calculation of the Mean Absolute Error from the Equation given below.

$$\text{MAE} = \frac{(\sum_{i=1}^{n} | xi - yi |)}{n}$$

Where,
xi - Number of Actual Trials Conducted.
yi-Number of Accurate Leaf Classifications performed.
n- Number of Experiments Conducted.

Experiment Number	Number of Trials Conducted (xi)	Number of Accurate Leaf Classifications performed (yi)	Difference (xi-yi)
1	10	9	1
2	10	8	2
3	10	10	0
4	10	10	0
5	10	9	1
6	10	10	0
7	10	7	3
8	10	10	0
9	10	9	1
10	10	10	0
		MAE	0.08

Table 1: Experimentation and Calculation of MAE.

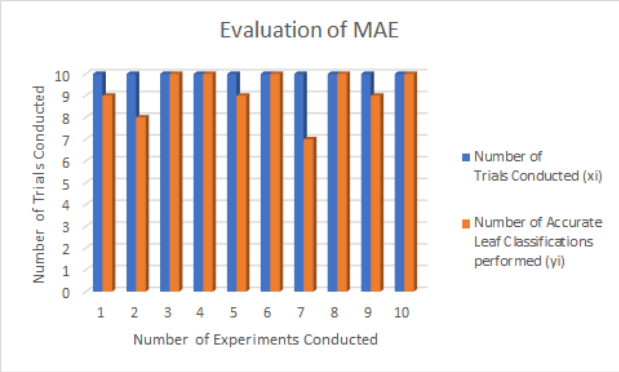

Figure 2: Evaluation of MAE.

Table 1 Given above has the values generated through the Mean Absolute Error evaluation, which reveals certain important aspects of the process and the calculations performed. The Mean Absolute Error has been calculated through executing 10 Experiments, where, each of the experiment was formed by performing 10 trials each. The values obtained through the numerous trials and experiments have been plotted onto a graph that is given in Figure 2 above. This indicates that the Mean absolute error for this system has been calculated to be around 0.08. This is a highly impressive score which has been evident from the results.

V CONCLUSION AND FUTUREWORK

The proposed system of leaves classification system is deployed in real time environment using the leaf images of more than 1000 belongs to 50 different species. Every input leaf images are indexed properly by converting the image into edge images and then a morphology vector is formed with respect to the edges to store in a file. This morphology array eventually represents the shape of the image with respect to the single axis. While searching also for the input image same kind of morphology array will be formed, then both the stored vector and input leaf vector are evaluated for the Pearson correlation value to measure the correlation between the two input arrays. Based on this the highest correlated array for the stored image is matched with the input image. The proposed system is measured for the Mean absolute error to measure the effectiveness of the proposed model. On applying this the experimental results show that the proposed system yields the MAE of 0.08 which is indeed the best for leaf classification models.

In the feature this system can be enhanced to work as mobile application where on clicking the image of a leaf can provide the complete leaf information.

REFERENCES

[1] Alagukumar, S and Lawrance, R, “Classification of Microarray Gene Expression Datausing Associative Classification”, International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), 2016.

[2] Liankuan Zhang, Chunlei Xia, Deqin Xiao et al, “A leaf vein detection scheme for locating individualplant leaves”, International Conference on Information and Communication Technology Robotics, 2018.

[3] Ye Cao, Chenjun Yan, Jingyue Li,and Hong Zhou, “Leaf vein extraction and angle measurement using hue information and line detection”, 9th International Conference on Intelligent Human-Machine Systems and Cybernetics, 2017.

[4] MelikeSardogan, AdemTuncer and YunusOzen, “Plant Leaf Disease Detection and Classification basedonCNN withLVQAlgorithm”, 3rd International Conference on Computer Science and Engineering, 2018.

[5] KentaroFukuta, Tomomasa Nagashima and Yoshifumi Okada, “LEAF: leave-one-out forward selection methodfor cancer classification using gene expression data”, IEEE/ACIS International Conference on Computer and Information Science, 2010.

[6] AzeilLouisseCodizar and Geoffrey Solano, “Plant Leaf Recognition byVenation and ShapeUsing Artificial Neural Networks”, 7th International Conference on Information, Intelligence, Systems & Applications (IISA), 2016.

[7] MuhdSafarudinChek Mat, Jezan Md Diah, Mokhtar Azizi Mohd Din & Abd. Manan Samad, "Data Acquisition and Representation ofLeaves using Digital Close-RangePhotogrammetry for Species Identification", IEEE 5th Control and System Graduate Research Colloquium, Aug. 11 - 12, UiTM, Shah Alam, Malaysia, 2014.

[8] Manojkumar P., Surya C. M. and Varun P. Gopi, "Identification of Ayurvedic Medicinal Plants byImage Processing of Leaf Samples", Third International Conference on Research in Computational Intelligence and Communication Networks, 2017.

[9] Xiaohan Yu, ShengwuXiong, Yongsheng Gao, Yang Zhao,andXiaohui Yuan, "Multiscale Crossing Representation Using CombinedFeature of Contour and Venation for Leaf ImageIdentification", International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2016.

[10] Karthik. G and Praburam. N, "Detection and Prevention of Banana LeafDiseases From Banana Plant Using EmbeddedLinux Board" Online International Conference on Green Engineering and Technologies, 2016.

[11] Arya M S, Anjali K,andMrs.DivyaUnni, "Detection of Unhealthy Plant Leaves UsingImage Processing and Genetic Algorithmwith Arduino", International Conference on Power, Signals, Control and Computation (EPSCICON), 2018.

[12] M. Merchant, V. Paradkar, M. Khanna and S. Gokhale, "Mango Leaf Deficiency Detection Using DigitalImage Processing and Machine Learning", 3rd International Conference for Convergence in Technology (I2CT)The Gateway Hotel, XION Complex, Wakad Road, Pune, India. Apr 06-08, 2018.

[13] Zulkifli Bin Husin, Abdul Hallis Bin Abdul Aziz, Ali Yeon Bin Md Shakaff and Rohani Binti S Mohamed Farook, "Feasibility Study on Plant Chili Disease Detection Using Image ProcessingTechniques", Third International Conference on Intelligent Systems Modelling and Simulation, 2012.

*****