Available online at http://www.mecs-press.net/ijem

Analysis of Late Blight Disease in Tomato Leaf Using Image Processing Techniques

Megha P Arakeri[1], Malavika Arun[2], Padmini R K[3]

[1,2,3]*M.S Ramaiah Institute of Technology, M.S.R Nagar, Bangalore 560-054, India*

Abstract

Tomato (*Lycopersicon esculentum* L.) is one of the most widely grown crops in the world. This crop is easily prone to various diseases. One such disease is late blight, caused by the fungus *Phytophthora infestans*. The first symptoms of late blight on tomato leaves are irregularly shaped, water soaked lesions, which are typically found on the younger leaves of the plant canopy. During humid conditions, white cottony growth may be visible on the underside of affected leaves. As the disease progresses, lesions enlarge causing leaves to brown, shrivel and perish. Hence in the present paper, a novel computer vision system has been proposed for detection and analysis of late blight disease. The proposed system implements thresholding algorithm to classify the leaf as diseased or healthy. Later it uses K -means clustering algorithm for analyzing late blight disease. The experiment was carried out on leaves of tomato collected from various plantations. The accuracy, sensitivity and specificity of the developed system in analyzing the late blight disease are 84%, 85% and 80% respectively.

Index Terms: Agriculture, tomato, late blight, segmentation, K- means clustering.

1. Introduction

Agriculture plays a decisive role in India's economy. It is the major source of raw materials, food, fodder and also foreign exchange. Besides, it is the most common occupation which supports 58% of Indian population (Pinaki Mondal and Manisha Basu, 2009). It is the premier source of our national income. Vegetables are important constituents of Indian agriculture due to their short duration, high yield, nutritional richness and ability to generate on-farm and off-farm employment. Farmers have a wide range of choices regarding the crops they can cultivate based on climatic conditions. However, in order to produce quality yield, scientific farming methods need to be followed. Engineering intervention in the form of technology and tools facilitates in maximizing agricultural productivity and profitability on a sustainable basis. In India, tomato

* Corresponding author. Tel.:
E-mail address:

covers an area of 4.97 lakh hectares with an annual production of 86 lakh tons which is 8.5% of the total crop production (Md. Rokunuzzaman and H.P.W. Jayasuriya, 2013) indicating its popularity. It is an important protective food, because of the presence of the several nutritive values. In addition to this, it also plays a major role in Indian traditional cooking. It also tops the list of canned vegetables. However, this crop is threatened by many foliar and fruit diseases.

One such disease is late blight (Late blight on tomato, 2014), caused by a fungus called *Phytopthora infestans*. It is a specialized pathogen for tomato. It can completely defoliate and destroy a crop within 2-3 weeks of infestation. This fungus is dispersed by air born sporangia, produced by branched hyphae that emerge from the stomata of infected leaves. It is favored by clouds protecting the spores from exposure to UV radiation and wet conditions, allowing the spores to infect when they land on leaves. As a result, lesions may appear on leaves within 3-5 days. The characteristic symptoms of this disease are water-soaked lesions which become black with white sporulation on the margin, frequently occurring on lower surface and occasionally on the upper surface of the leaf and appearing beige in color, once lesions dry out in hot weather. One unsatisfactory solution for this disease was "disease forecasting" that enabled farmers to predict when the environmental conditions were highly favorable to the spread of pathogen. Further to this, the naked eye observation of experts is the main approach adopted in practice for the detection and identification of such diseases.

Thus, farmers require the aid of experts, a deprival of which results in loss of crop. Hence, to bridge the gap between experts and farmers, this paper proposes a software for tomato leaf disease detection and analysis of late blight using image processing techniques. Here, the application can take the role of the expert and help the farmer identify the disease instantly at an early stage and take control measures to protect the crop. The application developed performs the following functions:

- It accepts images of type JPEG of specific resolution taken at a specific distance from the source.
- It is capable of detecting whether the given tomato leaf is healthy or diseased by applying the thresholding technique based on the color differences observed.
- If it is diseased, it also identifies and analyses whether it is diseased with late blight, using K-means clustering and thresholding techniques.
- Further, to calculate the total area of the leaf affected by disease, binary thresholding is applied to calculate the number of pixels corresponding to infected area. Accordingly, the remedies to the disease at different stages are provided to the farmer.

2. Related Work

Efforts have been made by various researchers to adopt image processing techniques for detection of diseases that produce symptoms like spots, patches, lesions, and variation in the roughness of the surface. Generally mango, grapes and pomegranate are affected by anthracnose disease. Jagadish et al., 2013 segmented the lesion areas and graded them based on percentage of affected area. They used neural network classifier to classify normal and anthracnose affected fruits. However, the classification accuracies for normal and affected anthracnose fruit types were 84.65% and 76.6% respectively. A similar study was carried out by Minghua et al., 2005 on late blight disease detection in tomato plantation using remote sensed spectral imagery. Here, disease could be detected only in the third stage and not in early stages. The author used 5-index feature vector method to distinguish between healthy and diseased plants.

Arivazghan et al., 2013 segmented unhealthy region of plant leaves and extracted their texture features to classify the disease. Leaves from banana, beans, guava, jackfruit, lemon, mango, potato and tomato were classified for various types of disease using support vector machines leading to overall accuracy of 87.66%. Niket Amoda et al., 2014 proposed a software based solution for automatic detection and classification of plant based diseases. Their work is based on color transformation structure for the input leaf image, without pre-processing techniques to enhance the image. The image was further subjected to color space transformation and segmented using the K-means clustering. Texture features were extracted from segmented infected objects. The

extracted features were fed into a neural network to classify the leaves.

A method was proposed by Xu et al., 2011 to detect nitrogen and potassium deficiencies in tomato plants. The algorithm extracted a number of features from the color image. Also, texture features were extracted using three different methods namely difference operators, Fourier transform and Wavelet packet decomposition. The selection of the features was carried out by means of a genetic algorithm. The accuracy of this diagnostic system was above 82.5% and it could diagnose disease about 6–10 days before experts could determine. Sannakki et al., 2011 presented a method to quantify disease symptoms based on Fuzzy logic. The test images were pomegranate leaves. The algorithm begins converting the images to the L*a*b* color space. The pixels are grouped into a number of classes through K- means clustering. One of the groups will correspond to the diseased areas, however the paper does not provide any information on how the correct group is identified. In the following, the percentage of the leaf that is infected is calculated. Finally, a Fuzzy Inference System is employed for the final estimation of the disease rating.

The method proposed by Al Bashish et al., 2010 tries to identify five different plant diseases. The authors did not specify the species of plants used in the tests. After a preprocessing stage to clean up and filter the image, a K- means clustering algorithm is applied to divide the image into four clusters. According to the authors, at least one of the clusters must correspond to one of the diseases. After that, for each cluster a number of color and texture features are extracted by color co-occurrence Method, which operates with images in the HSI format. Those features are fed to a MLP Neural Network with ten hidden layers, which performs the final classification. Pugoy and Mariano, 2011 developed a system to identify two different types of diseases that attack rice leaves. The algorithm first converts the image from RGB to HSI color space. The K- means technique is applied to cluster the pixels into a number of groups. Those groups are then compared to a library that relates colors to the respective diseases. This comparison results in values that indicate the likelihood of each region being affected by each of the diseases.

Wang et al. 2012 implemented a method to discriminate between pairs of diseases in wheat and grapevines. The images are segmented by a K- means algorithm, and then 50 color, shape and texture features are extracted. For the purpose of classification, the authors tested four different kinds of neural networks: Multilayer Perceptron, Radial Basis Function, Generalized Regression, and Probabilistic. The authors reported good results for all kinds of neural networks. Zhang Yongqin et al., 2011 comparatively studied the performance of color image segmentation using the optimized Chan-Vese model with initialization mask in multiple color spaces. This model was robust to noise and had good adaptability in RGB, CIE XYZ, and YCbCr color spaces. But, it produced poor segmentation results on a set of test images.

3. Proposed Methodology

The automated late blight disease detection system is designed in two phases:

- Detection
- Analysis as shown in Fig.1, where the leaf image is subjected to image processing algorithms.

3.1. Phase-I: Detection

In order to detect the late blight disease in tomato leaf, the system must be capable of differentiating between diseased leaves and healthy leaves as shown in Fig. 2 and Fig.3 respectively. The main color differentiation that exists on the diseased leaves helps in detecting the diseases or any abnormality existing on the leaf.

Three major steps are involved in achieving this:

- Pre-processing
- Thresholding
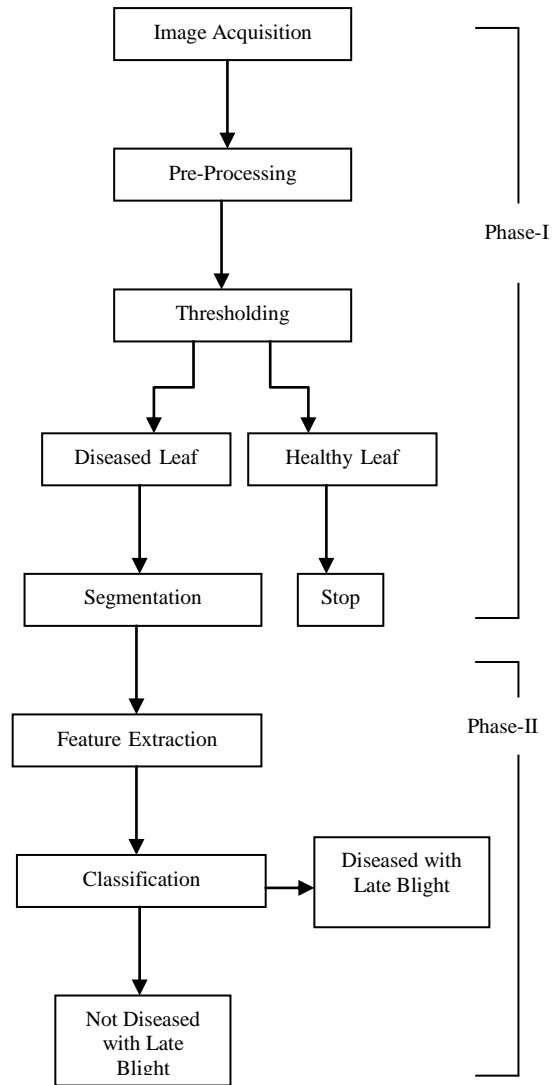- Identifying the leaf as healthy or diseased.

```
                    ┌─────────────────────┐
                    │  Image Acquisition  │
                    └─────────────────────┘
                               │
                               ▼
                    ┌─────────────────────┐
                    │   Pre-Processing    │
                    └─────────────────────┘           Phase-I
                               │
                               ▼
                    ┌─────────────────────┐
                    │    Thresholding     │
                    └─────────────────────┘
                         │           │
                         ▼           ▼
              ┌──────────────┐  ┌──────────────┐
              │ Diseased Leaf│  │ Healthy Leaf │
              └──────────────┘  └──────────────┘
                     │                 │
                     ▼                 ▼
              ┌──────────────┐     ┌────────┐
              │ Segmentation │     │  Stop  │
              └──────────────┘     └────────┘
                     │
                     ▼
              ┌──────────────────┐              Phase-II
              │ Feature Extraction│
              └──────────────────┘
                     │
                     ▼
              ┌──────────────┐     ┌─────────────────┐
              │Classification│────▶│  Diseased with  │
              └──────────────┘     │   Late Blight   │
                     │             └─────────────────┘
                     ▼
              ┌──────────────┐
              │ Not Diseased │
              │  with Late   │
              │    Blight    │
              └──────────────┘
```

Fig.1. Proposed methodology for late blight analysis

(a)   (b)

Fig.2. (a) Tomato leaf diseased with late blight; (b) Healthy tomato leaf

The acquired tomato leaf image is stored in the database and then it is pre-processed to reduce the noise. In the pre-processing step, R-, G- and B- channels are extracted separately and median filter is applied to each channel. The purpose here is to reduce the noise as much as possible in each R-, G- and B- channels, whose average values play major role in identifying correct threshold value. The filtered image is then restored using Lucy-Richardson algorithm (Swati Sharma et al., 2013). Further, mean of R-, G- and B- channels are calculated and an appropriate value is identified as threshold value. Based on this threshold value the leaf is identified as healthy or diseased.

3.2. Phase-II: Analysis

Once the leaf image is identified as diseased, the analysis proceeds with segmentation. Since color features show dominance, color based segmentation using K-means clustering (Vijay Jumb et al., 2014) is followed. The images are converted into L*a*b* space. It consists of luminosity layer 'L*', chromaticity-layer 'a*' indicating where color falls along the red-green axis, and chromaticity-layer 'b*' indicating where the color falls along the blue-yellow axis. The colors in a*b* color space are classified using K-means clustering algorithm as given in Fig. 3. K-means finds partitions such that objects within each cluster are as close to each other as possible and as far from objects in other clusters. In the present work, K-means clustering algorithm is applied on the leaf image to obtain the following three clusters:

- Background
- Healthy part of leaf
- Diseased part of leaf

Input: Image I with $X = \{x_1, x_2, x_3 \dots x_N\}$ pixels,

      $K = \{k_1, k_2, \dots k_K\}$ number of clusters.

Output : Segmented image with K clusters.

Steps:

  1. Randomly assign each pixel to any of the clusters

  2. Compute the distance between the pixel and all

     centroids $C = \{c_1 \dots c_K\}$ as:

$$\sum_{i=1}^{k} \sum x_j |x_j - c_k|^2; \ x_j \in k_i \qquad (1)$$

     where |.| denotes the Euclidean norm,

     c_k = mean of all pixels in k^{th} cluster

   3. Reassign the pixel x_j to the cluster c_i If and only if,

$$|x_j - c_i|^2 \leq |x_j - c_k|^2; 1 \leq k \leq K \qquad (2)$$

   4. Repeat above steps until there is no further re- assignment

---

Fig.3. K- means clustering algorithm

K-means algorithm returns an index corresponding to a cluster. The indices returned by K-means are used to label every pixel in the image. As a result of pixel labeling, objects in the image are separated by color. This leads to a formation of segmented images based on the number of colors present in the whole image. Features like R-, G-, B- components are extracted from the segmented diseased part. Next, mean of R-, G-, B- components is calculated and an experimental threshold value is set. This helps in differentiating between late blight disease and other diseases. If the system identifies the leaf as diseased with late blight, the area of diseased part is calculated to find the severity of disease. Based on the severity, the proposed system provides remedy which helps the farmer to control the disease.

4. Experimental Results

The experiment was carried out using MATLAB version R2008a. The developed system was tested with JPEG images having a resolution of 384x256 pixels. The dataset consists of 100 tomato leaf images pertaining to:

- Healthy leaves
- Late blight affected leaves
- Leaves affected by other diseases.

These images were captured using a camera in tomato plantations belonging to University of Agricultural Sciences, Bangalore. The leaf images were acquired in natural lighting conditions. The acquired leaf image of tomato plant is loaded onto the application. On clicking "Detect" button, the system pre-processes and thresholds the image to detect whether the leaf is diseased or not as shown in Fig. 4 and Fig. 5.
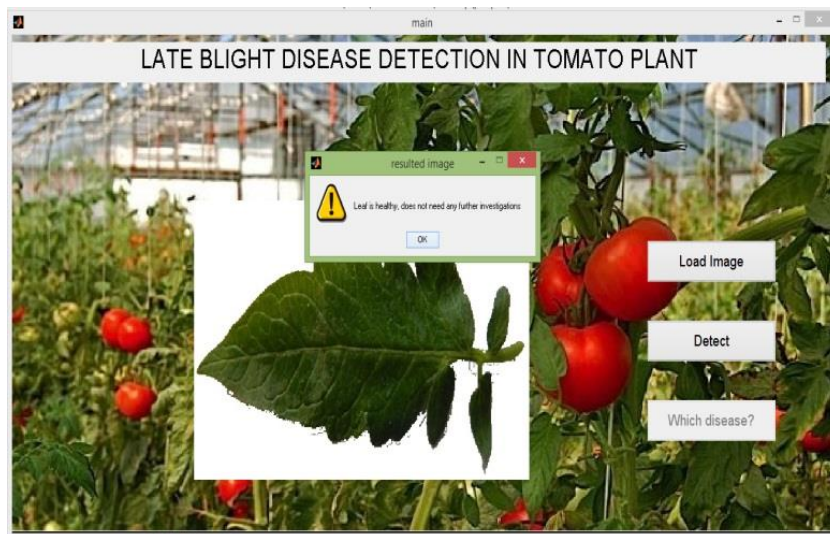

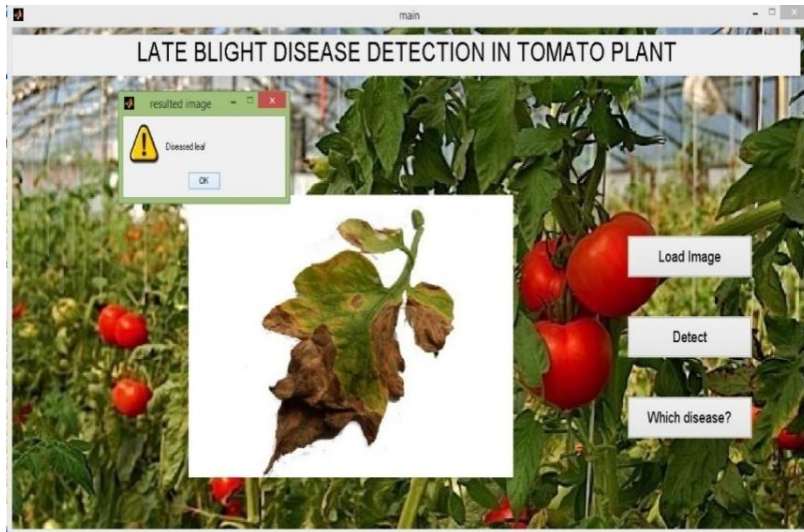Fig.4. Healthy tomato leaf detected by the proposed method

Fig.5. Diseased tomato leaf detected by the proposed method

If the application identifies it as a diseased leaf, segmentation is carried out on clicking "Which disease?" button. Here, the diseased part is extracted from the leaf image using K-means clustering. Then based on the color feature, the disease is classified as late blight or other disease. Fig. 6 shows the application classifying the leaf as diseased with late blight. Further, area (number of pixels) of the diseased part is calculated from the segmented image. Based on the extent of disease spread, remedial measures are provided, as shown in Fig. 7. This helps the farmer to identify the disease and apply control measures based on the severity of the disease.

Fig.6. Diseased part segmented from the tomato leaf

Fig.7. Remedy for late blight disease

From Fig. 7, the area of the diseased part of leaf is found by converting the diseased segmented part into binary image by applying thresholding. The number of black pixels is counted to give the area of diseased part. The effectiveness of the segmentation method was calculated using relative error as given in Eq. (3). Table. 1 gives the relative error of 15 image samples of late blight diseased leaves. Column 2 shows the actual area of the disease (in pixels) and column 3 shows the area of the diseased part segmented using K- means clustering described in our methodology. The relative error was calculated for all images in the data set and it was found that on average the relative error was 8.69.

$$Relative\ Error = \frac{A - P}{A} \times 100 \tag{3}$$

Where,
A: Actual diseased area
P: Diseased area identified using the proposed algorithm.

Table 1. Computation of relative error for proposed method

Sample	Actual diseased area (*pixels*)	Diseased area extracted from algorithm (*pixels*)	Relative Error(%)
1	18567	19636	5.7575
2	12168	10324	15.1545
3	5267	6738	27.9286
4	4596	4825	4.9825
5	3916	3692	5.7201
6	6258	5305	15.2285
7	68976	71242	3.2852
8	72097	74101	2.7795
9	56101	58917	5.0195
10	67178	70776	5.3559
11	3295	4049	22.8831
12	69875	74430	6.5187
Average relative error = 8.6994			

The disease classification performance of the developed method was evaluated using various metrics such as, sensitivity, specificity and accuracy as given below.

$$Sensitivity = (TP)/(TP + FN) \tag{4}$$

$$Specificity = (TN)/(FP + TN) \tag{5}$$

$$Accuracy = (TN + TP)/(TN + TP + FN + FP) \tag{6}$$

Where,

TP: True Positive, which is the proportion of positive cases that were correctly identified.
TN: True Negative, which is the proportion of negatives cases that were classified correctly.
FP: False Positive, is the proportion of negatives cases that were incorrectly classified as positive.
FN: False Negative, is the proportion of positives cases that were incorrectly classified as negative.

The calculation of the above metrics is shown in the Table 2. The developed method achieved sensitivity, specificity and an accuracy of 85%, 80% and 84% respectively.

Table 2. Performance evaluation of the proposed method

Test samples	
Condition positive	Condition negative
True positive (TP) = 16	False positive (FP) = 12
False negative (FN) = 4	True negative (TN) = 68
Sensitivity 16 / (16 + 4) =85%	Specificity 68 / (12 + 68) = 80%
Accuracy = (16+68)/100= 84%	

5. Conclusions

The early detection of the disease can be difficult because of less prominence of disease features. The developed system detects the diseased leaves and identifies late blight disease by processing the given input images. The thresholding and K-means color based segmentation helped to differentiate diseased leaves from healthy ones. Thus the developed system helps the farmer in monitoring and managing disease in tomato crop fields. Further to improve accuracy, texture features and classifiers can be included in the present system as a part of future work.

Acknowledgement

References

[1] Pinaki Mondal, Manisha Basu. Adoption of precision agriculture technologies in India and in some developing countries. Scope, present status and strategies, progress in Natural Science 2009;19:659-666.

[2] Md. Rokunuzzaman, H.P.W. Jayasuriya. Development of a low cost machine vision system for sorting of tomatoes. Agricultural Engineering International: CIGR Journal 2013;15(1):173-180.

[3] Late blight on tomato.[online], 2014, http://livegpath.cals.cornell.edu/research/late-blight-of-tomato/

[4] Jagadeesh Devdas Pujari, Rajesh Yakkundimath, Abdulmunaf Syedhusain Byadgi. Grading and Classification of Anthracnose Fungal Disease of Fruits based on Statistical Texture Features. International Journal of Advanced Science & Technology 2013;52:121-132.

[5] Minghua Zhang, Zhihao Qin, Xue Liu. Remote sensed spectral imagery to detect late blight in field tomatoes. Precision Agriculture, Springer 2005;6:489–508.

[6] S. Arivazghan, R. Newlin Shebiah, S. Ananthi, S. Vishnu Varthini. Detection of Unhealthy Region of Plant leaves and Classification of Plant Diseases using Texture Features. Agricultural Engineering International: CIGR Journal 2013;15(1):211-217.

[7] Niket Amoda, Bharat Jadhav, Smeeta Naikwadi. Detection and Classification of Plant Diseases using Image Processing. International Journal of Innovative Science, Engineering and Technology 2014;1(2): 1-7.

[8] Guili Xu, Fenling Zhang, Syed Ghafoor Shah, Yongqiang Ye, Hanping Mao. Use of leaf color images to identify nitrogen and potassium deficient tomatoes. Pattern Recognition Letters 2011;32(11):1584-1590.

[9] Sannakki SS, Rajpurohit VS, Nargund VB, Kumar A. Leaf Disease Grading by Machine Vision and Fuzzy Logic. International Journal. 2011;2(5):1709-1716.

[10] Al Bashish D, Braik M, BaniAhmad S. A framework for detection and classification of plant leaf and stem diseases. International conference on signal and image processing Chennai: IEEE 2010;113–118.

[11] Pugoy RADL, Mariano VY. Automated rice leaf disease detection using color image analysis. 3rd International conference on digital image processing 2011;8009:F1–F7.

[12] Wang H, Li G, Ma Z, Li X. Proceedings of the 2012. International Conference on Systems and Informatics (ICSAI) Yantai: IEEE; 2012. Application of neural networks to image recognition of plant diseases; 2159–2164.

[13] Zhang Yongqin, Chen Hui, Wang Ling, Xiao Yongjun, Huang Haibo. Color Image Segmentation Using Level Set Method with Initialization Mask in Multiple Color Spaces. International Journal of Engineering and Manufacturing 2011;4:70-76.

[14] Swati Sharma, Shipra Sharma and Rajesh Mehra. Image Restoration using Modified Lucy Richardson Algorithm in the Presence of Gaussian and Motion Blur. Advance in Electronic and Electric Engineering 2013;3(8):1063-1070.

[15] Vijay Jumb, Mandar Sohani, Avinash Shrivas. Color Image Segmentation Using K-means Clustering and Ostu's Adaptive Thresholding. International Journal of Innovative Technology and Exploring Engineering 2014;3(9):72-76.

Authors' Profiles

Dr. Megha. P. Arakeri, (born November 10, 1979) obtained her Master's degree in computer science from Jawaharlal Nehru National College of Engineering, Shimoga, India. She obtained Ph.D in Information Technology from National Institute of Technology Karnataka (NITK), Surathkal, India. She is currently working as Associate Professor in the department of Information Science and Engineering in M. S. Ramaiah Institute of Technology. Her research interests include Medical Image Processing, Computer Vision, Information Retrieval, and Pattern Recognition.

Malavika Arun (born December 24, 1994) is currently pursuing her B.E degree in Department of Information Science and Engineering, M. S. Ramaiah Institute of Technology, Karnataka, India. Her area of interest is Image Processing.

Padmini R. K (born March 30, 1994) is currently pursuing her B.E degree in Department of Information Science and Engineering, M. S Ramaiah Institute of Technology, Karnataka, India. Her area of interest is Image Processing.