Journal of Computer Science Engineering and Software Testing
Volume 4 Issue 2

Automated Classification of Commonly used Hybrid and Non-Hybrid Vegetables

[1]Mahantesh C. Elemmi, [2]Shanta Kallur, [3]Supriya Belkar
Department of CSE/ISE, K.L.E. Institute of Technology, Hubli, India
E-mail: *mc_elemmi2004@rediffmail.com, shanta06kallur@gmail.com, supriyabelkar27@gmail.com*

Abstract
The proposed work classifies commonly used hybrid and non-hybrid vegetables. The number of hybrid vegetables is more than the number of non-hybrid ones. Consumption of hybrid food for longer period can lead to various types of cancers due to lack of nutrients. People are increasingly becoming health conscious and thus most of the population started preferring non-hybrid food items. In the proposed work, four types of regularly consumed vegetables in the food items namely, brinjal, tomato, carrot and cucumber. Differentiating hybrid and non-hybrid vegetables in market is tough job for most of the people, especially, for the people living in cities. Thus, we propose to develop an application which will be able to identify and classify the vegetable image into hybrid or non-hybrid. The overall classification rate for classifying hybrid and non-hybrid vegetable images is 92% and 96.75%.

Keywords: *Computer vision, hybrid and non-hybrid vegetable, image processing*

INTRODUCTION

Computer vision is a process of automatic extraction, understanding and analyzing of useful information from single or sequence of images. When compared to human vision, the proposed technique gives more clarity about the information gathered from the captured image. The image data can take different forms, such as video sequences, views from multiple cameras, or multidimensional data from a medical scanner. Computer vision and image processing are widely used in food industry, biological science, material science, medical science, weather forecasting and many other fields.

Computer vision is a process of automatic extraction, understanding and analyzing of useful information from single or sequence of images. When compared to human vision, the proposed technique gives more clarity about the information gathered from the captured image. The image data can take different forms, such as video sequences, views from multiple cameras, or multidimensional data from a medical scanner. Computer vision and image processing are widely used in food industry, biological science, material science, medical science, weather forecasting and many other fields.

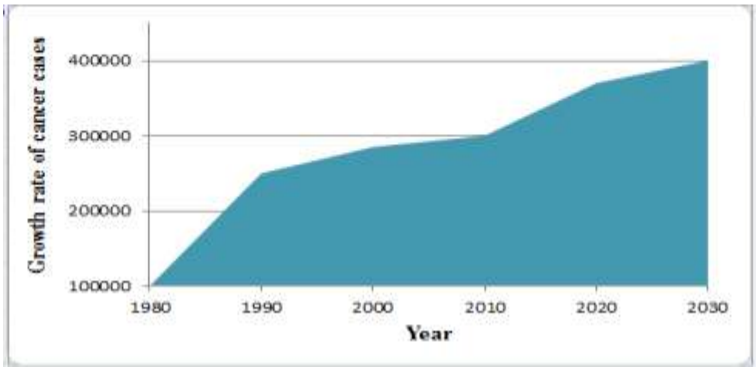


Fig 1: *Growth rate of cancer cases*

Problem Statement

Hybrid vegetables cause health hazards like diabetes and different types of cancer. Therefore there is need to develop a system that identifies hybrid and non-hybrid vegetables. Hence, the work on ***"Automated classification of commonly used hybrid and non-hybrid vegetables"*** is proposed.

LITERATURE SURVEY

[Timothy J Key, et. al,. 1990] presented a work on Diet, nutrition and the prevention of cancer. They have given particular credence to the results of large prospective studies and to the few results from randomized controlled trials. Formulating and Testing hypotheses are discussed in brief.

[B.K. Mohanty and A.M. Prusti, 2002] Carried out Hybrid research especially on the development of specific quality hybrids in tomato, brinjal, chilli and onion which is under progress at twelve centers located in different states.

[Sanjeet Kumar and P. K. Singh, 2004] described genetic (inherited) or non-genetic mechanisms viz., male sterility, self-incompatibility, gynoecium, auxotrophy, use of sex regulators and chemical hybridizing agents, based on their relative importance in hybrid development of vegetable crops.

[Mei chen, et. Al., 2008] have presented a wok on fast-food image data set is used to facilitate research in automated food recognition. PFID contains data of 101 fast-food acquired in both restaurant environments and laboratory settings. It offers still images, video, and stereo images to support different algorithm developments and evaluations.

[TaichiJoutou and KeijiYanai, 2009] Proposed a work on Multiple kernel learning for food image recognition. In this paper, work carried out on image recognition of many types of foods with high accuracy by introducing a MKL-based feature fusion method.

PROPOSED METHODOLOGY

The classification of commonly used vegetables into hybrid and non-hybrid vegetables is shown in Figure 2.

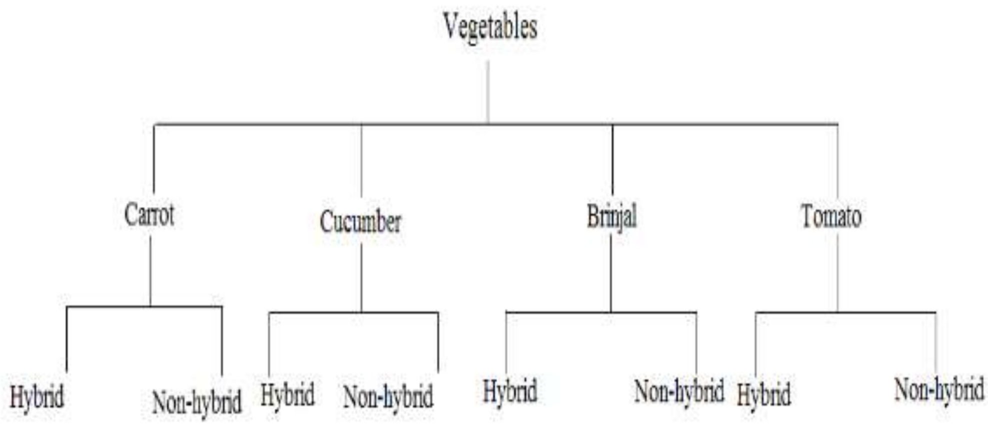


Fig 2: *Tree structure showing the work to be carried out.*

Vegetable image is given as input. Initially the vegetable image is identified into various types as tomato, cucumber, carrot and brinjal. In turn, each of the vegetable images are further classified as hybrid and non-hybrid. The classification is done on the basis of ANN and by using SURF key point detection technique. The vegetable image is identified as tomato, carrot, cucumber and brinjal. In turn each vegetable image is classified into hybrid and non-hybrid vegetables.

In SURF key point detection technique, the original image is converted into gray-scale image then the key points are detected from the given input vegetable image and descriptor values are generated. These generated descriptor values are matched with stored values in the data set and results are produced accordingly.

Identification of commonly used vegetables as carrot, brinjal, tomato, cucumber

The identification of commonly used vegetables such as carrot, brinjal, tomato, cucumber is done using surf key point detection technique. Here when we give sample image as input then it detects and identifies the name of vegetable.

Classification of vegetables into hybrid and non-hybrid

Surf keypoint technique is used to classify the vegetable images into hybrid and non-hybrid. Here we need to provide an image path as an input and then it processes the image and classifies the image into hybrid and non-hybrid.

Phases of initial system design

Different phases used in the proposed work are Data Collection, Pre-processing, Feature Extraction and Classification. The phases are shown in the following Figure 3

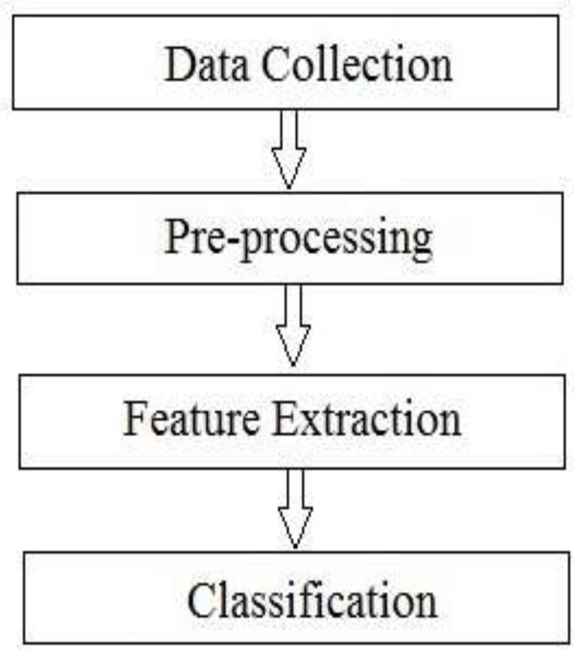


Fig 3: *Phases used in the proposed work.*

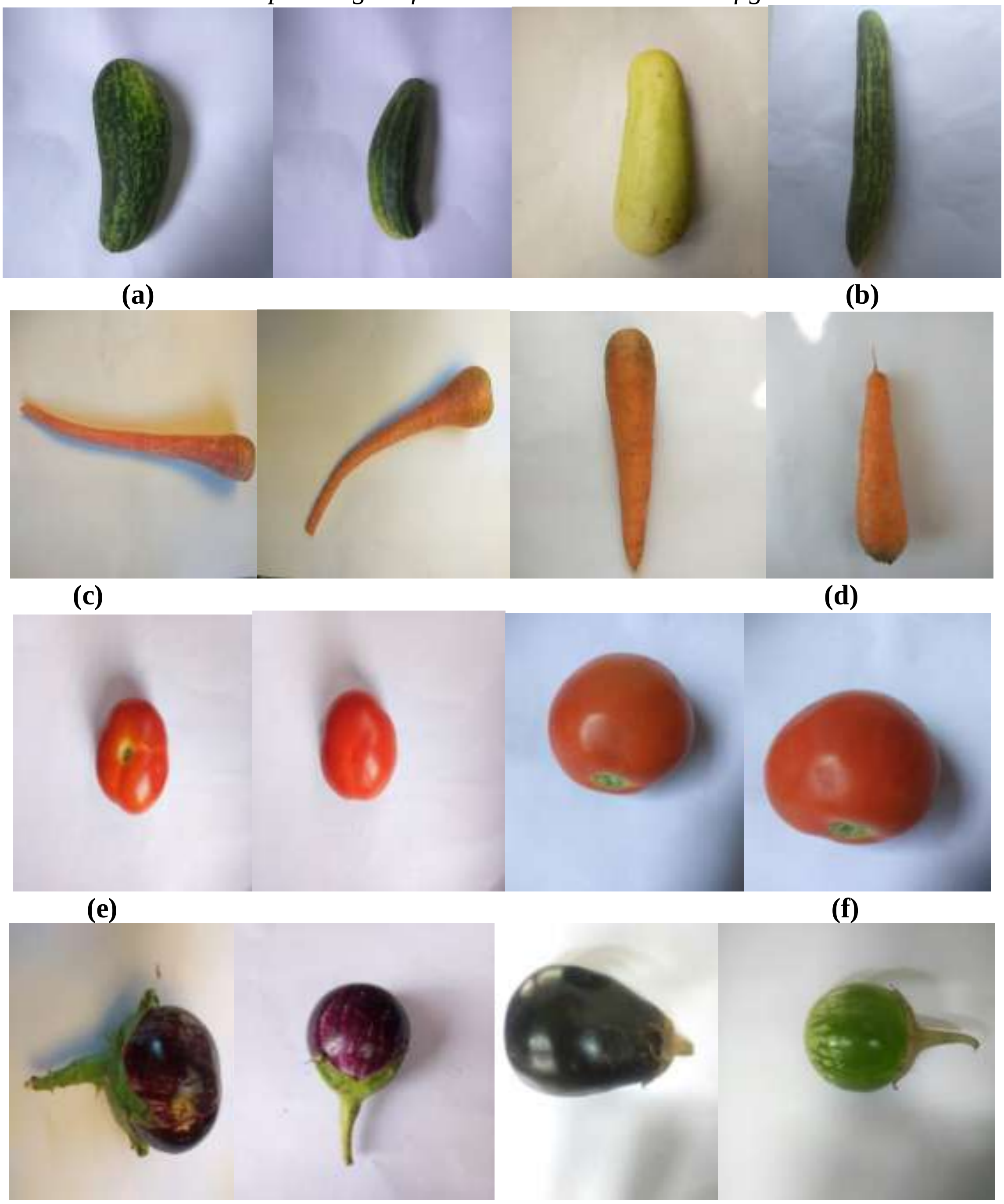
Data collection

Images of different varieties of vegetables are collected by visiting markets, fields and other sources using a high resolution camera. Constant distance and uniform light intensity is maintained while capturing the images. Images collected for hybrid and non-hybrid vegetables are given in table 1.

Table 1: *Data collection of hybrid and non-hybrid vegetables*

Vegetable Variety	Hybrid	Non-Hybrid
Cucumber	100	100
Carrot	100	100
Tomato	100	100
Brinjal	100	100

The sample images of data collection is shown in figure 4.



(a) **(b)**

(c) **(d)**

(e) **(f)**

(g) **(h)**

Fig 4: *Sample images of hybrid and non-hybrid vegetables (a)Non-hybrid cucumber images (b)Hybrid cucumber images (c)Non-hybrid carrot images (d)Hybrid carrot images (e)Non-Hybrid tomato images (f)Hybrid tomato images (g)Non-Hybrid brinjal images (h)Hybrid brinjal images.*

Pre-processing

Here, the images are collected preprocess to improve the quality of image data. The original image is resized to 300*300 pixels. The SURF technique is used for converting the input image into gray scale image.

Algorithm and features

The SURF technique is used .The

algorithm has three main parts such as (1) interest point detection,(2) local neighborhood description and (3)matching.

Detection

The square-shaped filters used in SURF for an approximation of Gaussian smoothing. Filtering the image with a square is much faster if the integral image is used:

$$S(x,y) = \sum_{i=0}^{x} \sum_{j=0}^{y} I(i,j) \quad \text{………...…(1)}$$

The sum of the original image within a rectangle can be evaluated quickly using the integral image, requiring evaluations at the rectangle's four corners as given in equation (1).

SURF uses a blob detector based on the Hessian matrix to find points of interest. The determinant of the Hessian matrix is used as a measure of local change around the point and points are chosen where this determinant is maximal. In contrast to the Hessian-Laplacian detector by Mikolajczyk and Schmid, SURF also uses the determinant of the Hessian for selecting the scale, as is also done by Lindeberg. Given a point p=(x, y) in an image I, the Hessian matrix H(p, σ) at point p and scale σ, is given as equation(2):

$$H(p,\sigma) = \begin{pmatrix} L_{xx}(p,\sigma) & L_{xy}(p,\sigma) \\ L_{yx}(p,\sigma) & L_{yy}(p,\sigma) \end{pmatrix} \quad (2)$$

where $L_{xx}(p,\sigma)$ etc. is the convolution of the second-order derivative of Gaussian with the image $I(x,y)$ at the point.

The box filter of size 9×9 is an approximation of a Gaussian with σ=1.2 and represents the lowest level (highest spatial resolution) for blob-response maps

Scale-space representation and location of points of interest

Interest points can be found at different scales, partly because the search for correspondences often requires comparison images where they are seen at different scales. In other feature detection algorithms, the scale space is usually realized as an image pyramid. Repeatedly smoothed with a Gaussian filter, then they are sub sampled to get the next higher level of the pyramid. Therefore, several floors or stairs with various measures of the masks are calculated in equation (3):

$$\sigma_{\text{approx}} = \text{current filter size} \times \left(\frac{\text{base filter scale}}{\text{base filter size}} \right) \quad \text{...…(3)}$$

A number of octaves are formed from scale space, where an octave refers to a series of response maps of covering a doubling of scale is refereed as an octave .The lowest level of the scale space is obtained from the output of the 9×9 filters in SURF technique.

By applying box filters of different sizes, the scale spaces in SURF are implemented. The analysis of space scale is done by up-scaling the filter size rather than iteratively reducing the image size. The initial scale layer at scale *s* =1.2 is considered as output of 9x9 filter. The following layers are obtained by filtering the image with gradually bigger masks, taking into account the discrete nature of integral images and the specific filter structure. The filters of size 9×9, 15×15, 21×21, 27×27 will be given as results.

Descriptor

The unique and robust description of an image feature is given by the Descriptor. It describes the intensity distribution of the pixels within the neighborhood of the point of interest. Hence a description is obtained for every point of interest identified previously. The Computational complexity and point-matching robustness/accuracy have direct impact from dimensionality of the descriptor.

Descriptor based on the sum of Haar wavelet responses

A square region is extracted, centered on the interest point and oriented along the orientation to describe the region around the point. The interest region is split into smaller 4x4 square sub-regions, and for each one, the Haar wavelet responses are extracted at 5x5 regularly spaced sample points.

Feature extraction

In feature extraction phase, unique features of hybrid and non-hybrid vegetables are extracted using SURF technique and then later used to detect the images. It is as shown in the following figure 5.

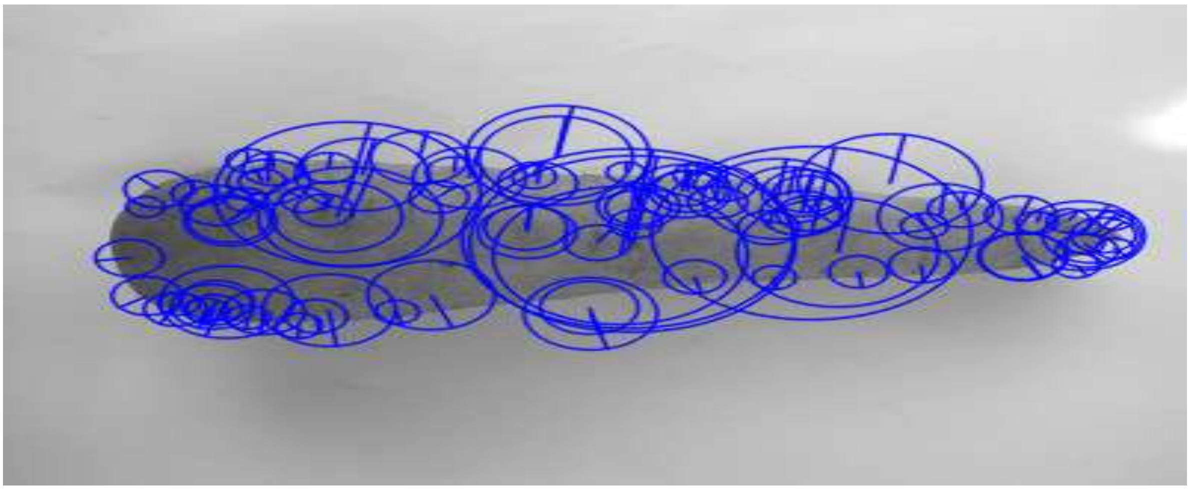


Fig 5: *SURF key point detection of vegetables*

The vegetable images are classified into two categories as Hybrid and Non-hybrid vegetables. Non-hybrid vegetables are small in size and dim in color whereas hybrid vegetables are big in size and bright in color when compared with non-hybrid vegetables. Non-hybrid vegetables are small in size and dim in color whereas hybrid vegetables are big in size and bright in color when compared with non-hybrid vegetables. Hybrid tomato is big in size, oval in shape and bright in color where as non-hybrid tomato is small in size, round in shape and dim in color. Hybrid cucumber big in size, long in shape and light in color where as non-hybrid cucumber is small in size and dark in color. Hybrid carrot is big in size, long in shape and bright in color where as non-hybrid carrot is small in size and dim in color. Hybrid brinjal is big in size and dark in color where as non-hybrid brinjal is small in size and dim in color.

Results and Discussion for identification of hybrid and non-hybrid vegetable images.

The classification rate for variety of hybrid and non-hybrid vegetable images is as shown in figure 6. For cucumber vegetable the classification rate is 81% and for carrot vegetable the classification rate is 85% and for brinjal vegetable the classification rate is 75% and for tomato vegetable the classification rate is 100%. From classification rate we can observe that maximum classification rate is obtained for tomato vegetable images.

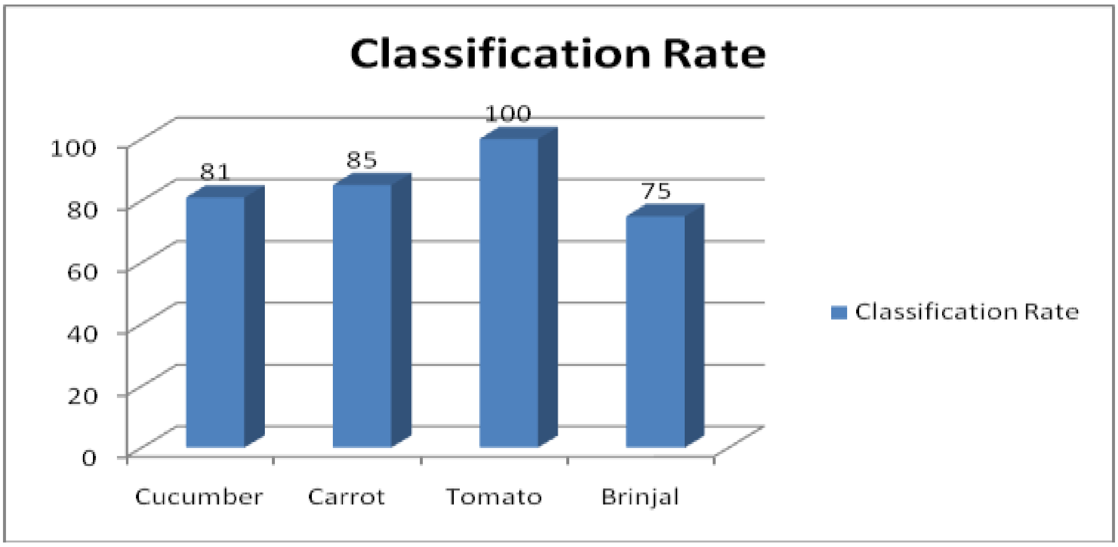


***Fig 6:** Classification rate of varieties of vegetable images.*

The overall classification rate for all these vegetables is 85.25%. The system has been trained to distinguish between hybrid and non-hybrid vegetables a confusion matrix will summarize the results of testing the algorithm for further inspection. By considering texture, shape and color of vegetable images, we are considering the RGB component values of the images. Based on these obtained values we are identifying the variety as cucumber, carrot, brinjal and tomato.

Result and discussion for classifying hybrid images

Classification rate for variety of hybrid vegetable images as shown in the figure. For cucumber vegetable images classification rate is 95% and for carrot, brinjal and tomato vegetable images classification rates are 86%, 92% and 95% respectively. From this recognition rate we can observe that the classification rate for hybrid vegetables is high.

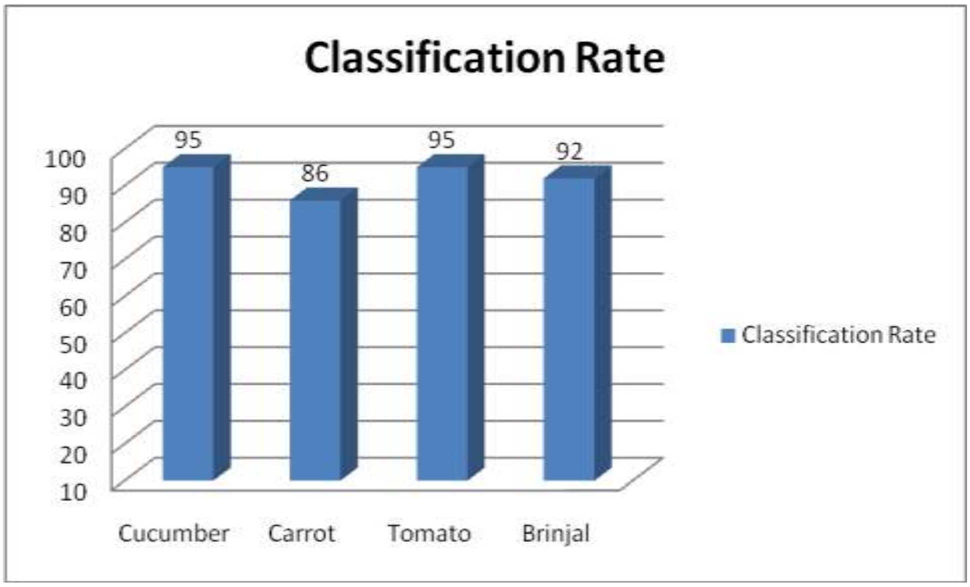


***Fig7:** Classification rate of varieties of hybrid vegetable images.*

The overall classification rate obtained for classifying hybrid vegetable images is 92%. The system has been trained to distinguish between different hybrid and non-hybrid vegetable images, confusion matrix will summarize the results of testing the algorithm for further inspection.

Result and discussion for classifying non-hybrid images

Classification rate for variety of non-hybrid vegetable images as shown in the figure. For cucumber vegetable images classification rate is 98% and for carrot, brinjal and tomato vegetable images classification rates are 94%, 96% and 99% respectively.

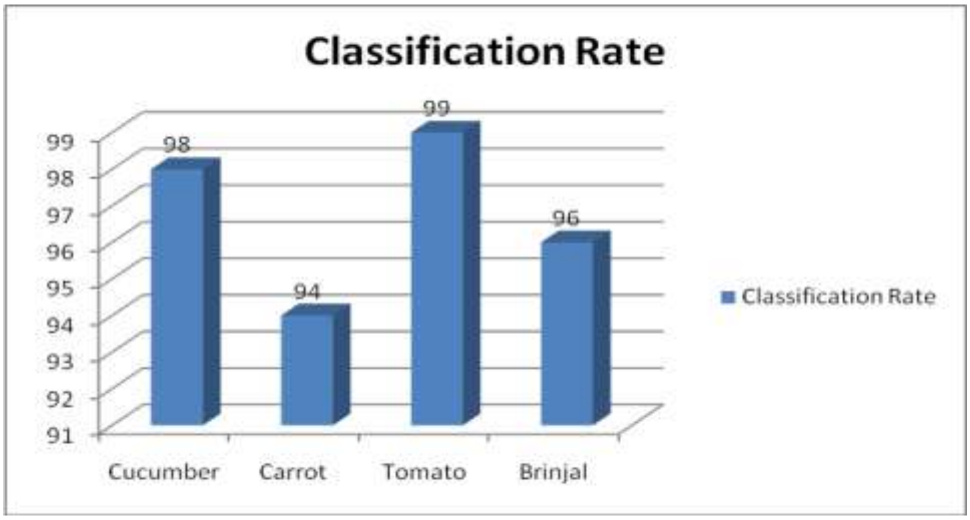


Fig 8: *Classification rate of varieties of non-hybrid vegetable images.*

The overall classification rate obtained for classifying non-hybrid vegetable images is 96.75%. The system has been trained to distinguish between hybrid and non-hybrid vegetable images,

Result and discussion for category of vegetable images

The classification rate for category of vegetable images is as shown in figure 9. For hybrid vegetable images the classification rate is 90% and for non-hybrid vegetable images the classification rate is 97%. From the classification rate we can observe that classification rate is high for non-hybrid vegetable images.

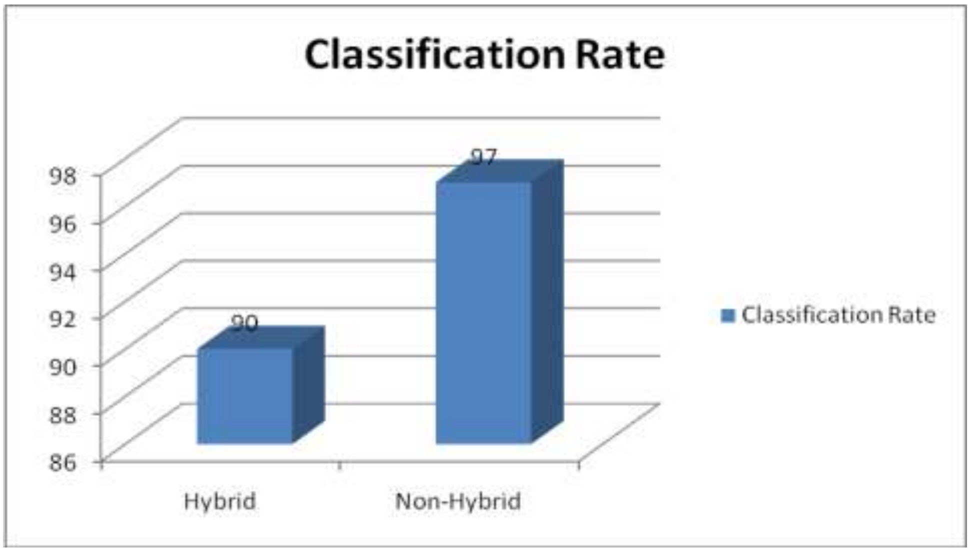


Fig 9: *Classification rate of vegetable category*

CONCLUSION

This paper will help the people to differentiate between hybrid and non-hybrid vegetables and they can avoid taking more hybrid vegetables and thus they can maintain their health. The SURF technique is used for feature extraction from images. We can go for quality analysis of non-hybrid vegetables in future enhancement.

REFERENCES

1. Aparajita, Rudransh Sharma, Anushikha Singh, Malay Kishore Dutta, KamilRiha, PetrKriz, (2017) - Image processing based automated identification of late blight disease from leaf images of potato crops.http://ieeexplore.ieee.org/document/8076090/
2. Hung-Lin Chi, Jian Chai, Changzhi Wu, Junxiang Zhu, Xiangyu Wang, Chongyi Liu (2017)-Scaffolding progress monitoring of LNG plant maintenance project using BIM and image processingtechnologies.http://ieeexplore.ieee.org/document/8002505/
3. Mahantesh C. Elemmi, ShantaKallur, NitinRamesh, Nikhil Aivalli, PrabhatNayak, AdarshNaik (2017) - Vision Based Classification of Different Diseases of Grape Leaves and their Severity.http://matjournals.in/index.php/IJISPR/article/download/1655/1061
4. Dayanand G Savakar / Basavaraj S anami (2009)-Recognition and classification of food grains, fruits and flowers using machine vision. https://www.researchgate.net/publication/250147479_Recognition_and_Classification_ofFood_Grains_Fruits_and_Flowers_Using_Machine_Vision
5. Sanjeet Kumar and P.K. Singh(2004)-Mechanisms for Hybrid Development in vegetables.https://www.researchgate.net/publication/254370246_Mechanisms_for_Hybrid_Development_in_Vegetables
6. B.K. Mohanty and A.M. Prusti(2002)-hybrid vegetable technology-a review. https://www.google.co.in/url?sa=t&source=web&cd=1&ved=0ahUKEwiU77HX7I3XAhXFq48KHa9DA4cQFggnMAA&url=http%3A%2F%2Fwww.arccjournals.com%2Fuplods%2Farticles%2Far233001.pdf&usg=AOvVaw0wY3a2KA5UNFpBAc1EdRp
7. Timothy J Key, Arthur Schatzkin, Walter C Willett , Naomi E Allen , Elizabeth A Spencer and Ruth C Travis-Diet, nutrition and the prevention of cancer.
8. http://www.google.co.in/url?sa=t&source=web&cd=4&ved=0ahUKEwi06ODl7o3XAhVGvo8KHaHLAnoQFghGMAM&url=http%3A%2F%2Fcdrwww.who.int%2Fnutrition%2Fpublications%2Fpublic_health_nut6.pdf&usg=AOvVaw0hKKl6699XvpYfFYY_V09A