IDENTIFICATION OF ARTIFICIALLY RIPENED MANGOES USING DIGITAL IMAGE PROCESSING TECHNIQUE

BY

MAHAMUDUL HASAN KHAN SUVRO

ID: 161-15-7009

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Nusrat Jahan

Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

Saiful Islam

Senior Lecturer

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 201

APPROVAL

This Project/internship titled **"Identification of Artificially Ripened Mangoes Using Digital Image Procession Technique"**, submitted by Mahamudul Hasan Khan Suvro, ID No: 161-15-7009 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on December 6, 2019

BOARD OF EXAMINERS

Dr. Syed Akhter Hossain **Professor and Head** Department of Computer Science and Engineering Faculty of Science & Information Technology Daffodil International University	**Chairman**
Nazmun Nessa Moon **Assistant Professor** Department of Computer Science and Engineering Faculty of Science & Information Technology Daffodil International University	**Internal Examiner**
Dr. Fizar Ahmed **Assistant Professor** Department of Computer Science and Engineering Faculty of Science & Information Technology Daffodil International University	**Internal Examiner**
Dr. Mohammad Shorif Uddin **Professor** Department of Computer Science and Engineering Jahangirnagar University	**External Examiner**

DECLARATION

I hereby declare that, this project has been done by me under the supervision of **Nusrat Jahan, Lecturer, and Department of CSE**Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Nusrat Jahan
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Saiful Islam
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Mahamudul Hasan Khan Suvro
ID: 161-15-7009
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete the final year project/internship successfully.

I am really grateful and wish my profound, my indebtedness to **Nusrat Jahan**, **Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of "*Machine Learning*" to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

I would like to express my heartiest gratitude to Prof. Dr. Syed Akhter Hossain, Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

I would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

In this Report, a proficient technique of image processing has been used to detect artificially ripened mangoes. Mango is one of the significant fruit crops across the world and if we mention Bangladesh, it is the most popular one during summer. Even the word mango has been originated from manga in Indian subcontinent. But now a days, this fruit is being ripened artificially with chemicals by some mischievous sellers unknowingly or sometimes knowingly the hazardous effects of this chemicals on human health. Among those chemicals that has been used for ripening mango artificially, calcium carbide is the most used one. Lately, the national agency for food and drug administration and control (NAFDAC) warned the mass people against devouring fruits ripened with calcium carbide by saying that those fruit cause heart, kidney and liver failures when consumed. Hence, consumers must have to be careful while buying fruits. It is kind of difficult for human eye to find out the fruits that have been chemically ripened. Here in this research, a method of digital image processing has been used to aid the detection. This research considers sample images of more than 150 mangoes. Total 24 features have been extracted from this images and then categorization is done using j48 decision tree, multilayer perceptron and naïve bayes. Also a comparison of accuracy between j48 decision tree, multilayer perceptron and naïve bayes has been mentioned.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Ripening is the procedure which causes the fruit to get more toothsome. In natural, when a fruit gets ripened, it becomes sweeter, redder or one can say less green and softer. It has been pointed out that even the natural ripening procedure increases the acidity level in the fruit, it doesn't get tarter with the rise of acidity level [1].

But ripening fruit is a slow natural process and in many cases ripened fruits don't ship well. So, the traders tend to fasten the process of fruit ripening by using chemicals which they seem like been getting more benefited than the naturally ripened ones. Various artificial procedures of fruit ripening have been audited for meeting consumers demand. But it is not mandatory that the artificial fastening procedure will bring good to everyone. While it seems beneficial to the traders, it is leaving hazardous effect on the health of the consumers. Calcium carbide is one of those chemicals that is used for ripening fruits artificially. Lately in Bangladesh, people are spotted consuming fruits that are ripened with calcium carbide. Due to its availability in local market at low cost, calcium carbide (CaC_2) is a famous chemical. It is making healthy fruit poisonous by carrying toxic materials like arsenic and phosphorous to consumers and for this, usage of it in the fruit industry is being discouraged across the world [2]. For explaining calcium carbide, once an expert of food safety told that if calcium carbide is put in the water, it will release acetylene that is used for welding where the natural ripening agent is also acetylene but the one in the air [3].

Pure calcium carbide is transparent or sometimes grayish to black and it produces a garlic odor. The following reaction, $CaC_2 + 2H_2O = C_2H_2 + Ca(OH)_2$ shows how calcium carbide produce ethylene gas when it reacts with water which is a chemical agent for ripening fruits artificially. The fruits which were ripened with calcium carbide are likely to be soft, and they often have an attractive surface though the inside tissues may or may not be ripened properly and is lower in flavor and taste. Acetylene is not intensely poisonous if it remains under the permissible limit. But if it crosses the permissible limit it may induce prolonged hypoxia I.e. deficiency of oxygen and thus affect the neurological system. Findings related to carbide poisoning have shown that it may cause dizziness, headache, mental confusion, memory loss, cerebral edema, mental confusion, mood disturbances and seizures [4]. Even by consuming carbide ripened fruit, it can lead the pregnant women to miscarriage. It also contains carcinogenic properties one of the popular seasonal fruits which is being ripened from the beginning of the artificial ripening process is mango. It is the juiciest fruit all over the world and researches showed that it is one of the most consumed fruits. Thus the traders get mostly beneficial by selling mangoes to the consumers. It may take several days for the transportation and distribution of mangoes from the farmer's orchards to the consumer's basket. The naturally ripened mangoes sometimes become over ripen and unwholesome during the whole time.

During the poor conditions and transportations, some part of the mangoes ripened 'naturally' or without adding chemical agents can also get damaged which is a great economic loss for the sellers. For this reason they prefer artificial ripening with calcium carbide as it is inexpensive and easily accessible in local markets. But a great amount of calcium carbide on a juicy fruit like mango leave a bad effect on the consumers' health. The consumers can play a vital role by not buying the mangoes ripened artificially. There are many difference in texture, aroma, firmness etc. between naturally and artificially ripened mangoes.

But it is not always possible for common people to identify the difference between natural and artificially ripened mangoes. For this reason, in this research, I have used three methods of digital image processing to identify naturally ripened mangoes and artificially ripened mangoes and proposed the most accurate one.

Table 1.1.1: Quality parameters of artificially ripened and naturally ripened mangoes

Quality Parameters	Types of mangoes	
	Artificially Ripened	Naturally ripened
Texture	Not very attractive but uniformly colored	Attractive but uniformly Colored
Aroma	Mildly good	Good
Firmness	Fair	Fair
Taste	Mildly in-core sour, not very pleasant	Sweet, pleasant

But it is not always possible for common people to identify the difference between natural and artificially ripened mangoes. For this reason, in this research, I have used three methods of digital image processing to identify naturally ripened mangoes and artificially ripened mangoes and proposed the most accurate one

1.2 Motivation for This Study

Calcium Carbide (CaC_2) is banned under PFA rules 1955 and also under Food Safety and Standards (Prohibition and Restrictions on Sales) Regulations, 2011 made thereunder. When Calcium Carbide reacts with water, it makes acetylene gas which is also knows as carbide gas. Carbide gas is in analogue to ethylene and boosts the ripening process. It is said to have the same effect as ethylene the natural ripening hormone.

However acetylene is not nearly as effective for ripening as is ethylene and acetylene is not a natural hormone as ethylene. Calcium Carbide has traces of arsenic and phosphorous hydride. A strong reactive chemical, Calcium Carbide has carcinogenic properties and is used in gas welding. Being cheap and easily available in the local markets, CaC_2 is indiscriminately being used in ripening fruits.

Consumption of fruits like mangoes with Calcium Carbide causes stomach upset because the alkaline substance is an irritant that erodes the mucosal tissues in the stomach and disrupts intestinal functions. As CaC_2 imitates acetylene gas, it may affect the neurological system by inducing prolonged hypoxia (low oxygen reaching the blood and tissues). The fast ripened fruits contain harmful properties because CaC_2 contains traces of arsenic and phosphorous and the production of acetylene gas has a hazardous effect on human. It may affect the neurological system by causing headache, dizziness, mood disturbances, sleepiness, mental confusion, memory loss, cerebral edema and seizure [4].

Most of the people consume mangoes which look more yellow thinking that they are naturally ripened where they have no idea that how harmful these mangoes are. By applying my method people can identify the natural mangoes. They don't have to fall in the trap of those people who sell mangoes naming them naturally ripened

1.2 Rationale of the Study

I want to identify the natural ripened mangoes and artificially ripened mangoes based on digital image processing technique. The research has focused on the following specific objectives:

>To extract and analyze the RGB and HSI features of mango images.

>To distinguish the uses of calcium carbide (CaC_2) in artificially ripened mangoes through image processing as automated identification.

>To develop a model of mango shape detection and extraction of mango image based on proper analysis of RGB values of sample images.

1.3 Research Questions

There were several dispersed questions floating around my mind at the very first time I decided to do this work. Thinking specificity, complexity and relevancy to a social or scholarly issue, the questions pinpointed me what exactly I want to find out and gave my focus a purpose. Eventually, I was chasing the answers of why people were poisoning the most popular summer fruit of Indian subcontinent? What laws are constituted against these acts and how well are they preventing people from ripening fruits artificially? How can anyone compute featured data which those distinguishes naturally ripen fruits from poisoned fruits? What computing environment is best for extracting features from simple mango images and how to mine these features? What methods or algorithms will be more accurate and precise to find artificially ripened fruits that can strip the established procedure?

1.4 Expected Outcome

I expect that this study will be able to distinguish between chemically and naturally ripened mangoes via different classifiers. The most effective classifier can be used for implementing a machine vision technique which will be able to identify our objectives from a large data set. Eventually, this study will help other researchers to have a head start on digital image processing as well as it will help general people to avoid eating chemically ripened fruits.

1.6 Report Layout

The work is divided into five segments. The general notations and concepts are given in chapter 1 which are required to understand the problems. Background of the study is also presented in chapter 2. My methodologies are appended in chapter 3. Results, discussions, my proposed model and its working procedures are included in the chapter 4 and finally we conclude our thesis with a brief overview in chapter 5.

CHAPTER 2

BACKGROUND

2.1 Introduction

Natural fruit ripening process can be stimulated using different artificially ripening processes. As it has quite a bad impact on health, the artificially ripening processes has become questionable. There are direct and indirect health hazards associated with artificial ripening agents and their impurities, which require qualitative and quantitative analysis of chemical toxicity and their impact on fruit quality. Identification between artificially ripened fruits and naturally ripened fruits is important to prevent these health hazards.

2.2 Related Works

It is a matter of surprise that is spite of all this health hazards related to those chemicals which are used to ripen fruits, there is no marketization of any established technology. General people can't differentiate between artificially and normally ripened fruits. They need a technology for it which they can carry with them anywhere.

There are some methods which are applied on fruits to detect the artificial ones. But they don't have that much accuracy. Mansor announced that "by using RGB color sensor and fuzzy logic as classification algorithm, the accuracy of mango grading is more than 85%". In ripeness mango classification, Fuzzy logic is successfully categorized as a decision support technique. Human expert result and fuzzy logics grading results shows quite related agreement. By integrating with other features, performance result of the system can be improved. Furthermore, this research is a first attempt of fruit ripeness determination by using RGB color sensor. The result shows that RGB color sensor can be used accurately as data acquisition and can be applied to other classification of fruits [8].

A machine vision technique is proposed to cluster mangoes into four stages which are determined on the factor of market distance and market value. They have used Support Vector Regression (SVR) for maturity prediction in terms of actual-days to-rot and

Multi Attribute Decision Making (MADM) system for determination of quality from the quality attributes. Finally for grading based on maturity and quality, fuzzy incremental learning algorithm has been used. The performance accuracy of their proposed system is nearly 87% [9].

Color, size, shape, texture and different defects are external properties of fruits which are very important attributes for classification and grading. Now a days due to availability and advancement of machine vision, hardware, software etc. manual classification of fruits has been replaced with automated machine vision systems. There are more reasons like producing accurate, rapid, objective and efficient results over manual work. This paper reviews the basic process flow of fruit classification and grading. Feature extraction methods for size, color, shape and texture are explained with SURF, HOG and LBP features. Finally some machine learning approaches like K-NN, SVM, and ANN and CNN are briefly discussed. Though some challenges are still need to overcome, but machine vision will prove to be the future for non-destructive fruit classification and grading [10].

'A Computer Vision System' is proposed to distinguish varieties of French olives has been developed. Pit images (frontal and profile) were used, characteristics such as the histograms of the RGB model and form descriptors (area, perimeter, length, width, etc.) were computed. The classification procedure is performed by Partial Least Squares Discriminate Analysis [11].

A machine vision technique had been developed for artificially ripened and naturally ripened bananas using image processing. Statistical features had been extracted from GLCM matrix in wavelet domain using Haar filter. The wavelet decomposition had been used and a discriminatory variation between the naturally ripened and artificially ripened bananas were spotted. SVM classifier was used for the classification between artificially ripened and naturally ripened bananas [12].

A research had also developed an android application for identifying artificially ripened mangoes. In that research, the classification was done with the histogram threshold values using MATLAB. Then the procedure was carried out by developing an android application in android developer toolkit. The method needed an operating system where they can install and run it. The accuracy of the technique was 91% [13].

Spatial details are not focused in histogram analysis, and therefore it cannot guarantee the segmented regions to be contiguous. This problems can be solved by using decision tree as a classifier. It has been a matter of fact that even though the advantages of the decision tree as a classifier, researchers don’t think of it as an option most often. The decision tree gave the accuracy of approximate 96% [14].

Hall and Smith (1998) developed a feature selection algorithm that uses a correlation based heuristic to assess the merit of features. The heuristic by which the correlation based feature selection (CFS) measures the “goodness” of feature subsets takes into account the usefulness of individual features for predicting the class label along with the level of inter correlation among them. The hypothesis is that good feature subsets contain features highly correlated with (predictive of) the class, yet uncorrelated with (not predictive of) each other [15].

Another technique for image classification which is also not popular but effective and with higher accuracy is multilayer perceptron (MLP). Not many researches have been found regarding image classification using MLP. In a research paper, it has been observed that by using MLP there they got a slight amount of decrease in accuracy [16].

2.3 Research Summary

This chapter has elaborately discussed different kinds of chemical agents used for artificial mango ripening. Calcium Carbide is vital chemical in ripening of mangoes artificially has been discussed briefly. The effects of this procedure leave a vital threat to consumer’s health and this may cause many diseases. There are possible technologies which can be used for artificially fruit ripening has also been discussed in this chapter.

2.4 Scope of the Problem

Mango is the most popular summer fruit in Indian continent and other countries as well. But why people are poisoning this beautiful healthy fruit? There are many reasons for ripening fruits artificially. The problem arose when people were consuming chemically poison fruits which they have no idea how are they endangering their lives. In the study, I'm not going to discuss about how to avoid people from poisoning fruits but I'll talk more about how to detect it. It people can distinguish between chemical fruits and natural fruits, people will eat natural fruits and chemically ripen fruits will have no reason to stay on. Detection of artificially ripen fruits can be done by many ways rather than computation. There are many chemicals which can detect calcium carbide within seconds. One just have to take a sample amount of a mango and have to react it with another chemical to get the result that whether it is artificially ripened or not. One can just simply cut the mangoes and see its inside and easily identify. But these processes has some boundaries or hangovers. Like if anyone cuts a mango to test its purity, the mango is wasted. Same goes for chemical testing. There are other discomforts like when we have a huge amount of mangoes which will be exported to another country and we have to test its purity, things will get messy. It will be very time consuming and requires a lot of workers to detect artificiality from all these huge amount. The scope of the problem is expanding to businessmen as an opportunity to run their system at maximum profit. For example, naturally ripened mangoes are very delicate. In order to export them into different cities, it takes time and which can cause a naturally ripened mango to waste. The distribution can also be a problem. Most of the cases, mangoes are transported via truck or these types of vehicles which produce chances that the mangoes can be wasted. But artificially ripened mangoes won't face these types of problems because they are forcefully ripened before time and calcium carbide make the surface strong. These are opportunities for the dealers to use chemicals more for more profit but in these proposal, detection can be machine based and comforts with large amount of data. There is no need of manual workers, just an analyst can detect even thousands of mangoes just looking at a screen.

2.5 Challenges

Initially, collecting data was a huge challenge. Mangoes are summer fruits so they are not always available in markets. When I first purchased one kilo mangoes from Farmgate, Dhaka for data input, most of the mangoes were chemically ripened and how hard I was searching for natural mangoes I couldn't find it anywhere. Then mango season passed away and there were no data for my study. I tried to find images from Google but there are a lots of variations of mango species. Finally my friend from Satkhira, Khulna helped me out by sending images of mangoes from their own garden. Next challenge was to learn MATLAB for feature extraction and background subtraction. In WEKA, after selecting the features, some unknown bug was appearing at first and it paused my study for a long time.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The goal of this study is detection of artificially ripened mangoes from their images. For starting the procedure, pictures of mangoes have been taken as inputs and then the backgrounds have been subtracted in order to get noise free images for analyzing the features of the mangoes. Different features have been identified and extracted. HSI and RGB color values of the mangoes have been used. The most informatory attributes of features have been specified by evaluating them in two different ways. I have used MATLAB for feature extraction and WEKA (Waikato Environment for Knowledge Analysis) – a data mining tool in order to process the rest of these steps which includes feature selection as well as classification.

3.2 Research Subject and Instrumentation

In digital image processing, different types of color models are used such as RGB (Red, Green, and Blue), HSI (Hue, Saturation, and Intensity), CMY (Cyan, Magenta, and Yellow), and HSV (Hue, Saturation, and Value).

The RGB color model is Cartesian coordinate system based and in this model, each color appears in its primary spectral components of red, green and blue. We get a composite color image in combination of these three component images on the screen. In terms of human exposition, the RGB and CMY color models are not very well suited. Except human eyes perception for primary colors, RGB is kind of worthless. In terms of actual perception, RGB is unnatural. One don't look at a mango and think about the proportions of red, green and blue inside the somewhat dull-green, yellowish color. Rather the think about the color more in terms of hue, saturation and intensity.

The HSI model was designed for improvement of the RGB color model. Hue is the color itself. By looking at something when one try to assign a word to the color that he/she sees, the hue is being identified. Saturation refers to the "density" of the hue within the light that is reaching eyes. Intensity is essentially brightness. In a grayscale photograph, brighter areas appear less gray (i.e., closer to white) and darker areas appear more gray [11]. It can minimize effects from variations in illumination intensity. For this it is often used to develop color image processing algorithms.

In order to differentiate between naturally ripened mangoes and artificially ripened mangoes, the input must be the images of mangoes. Then the background of the images were dismissed so that the features can be properly analyzed from those images .In accordance with the RGB and HSI color model, overall 24 features were extracted. The most significant features were selected using classifier subset evaluator and information gain. The detection of artificially ripened mangoes was performed using decision tree, multilayer perceptron and naïve-Bayes method of those features. Finally the accuracy of these three methods had been compared and the most accurate one had been suggested to get approximately perfect result (Figure 4.3.2).

3.3 Data Collection Procedure

A total of 120 mangoes were selected from an orchard at Tala upazila, Satkhira. Half of them were naturally ripened and the other half were detached unripen from the tree. The raw mangoes were preserved in an isolated container. Packets of calcium carbide were kept in that container where acetylene gas that acts as a ripening agent was produced in contact with moisture. The ripening process had been observed and monitored carefully. After 24 hours the forepart of the container had been opened and the rest of the process took place with the container opened. And after 48 hours, the surface of the mangoes got colored yellowish like perfectly ripe mangoes where it took almost 55 to 60 days for a mango to ripe naturally. Calcium carbide hastened the procedure of ripening mangoes. Images of naturally ripened mangoes were captured before and after the end of the artificially ripened procedure the images of the artificially ripened mangoes were captured.

I took sample digital image of 120 “Amrapali” mangoes (Figure 3.3.1). After that, those images were preprocessed. Images had been taken with a resolution of 480 * 960 pixels using Casio Exilic Ex-FC100. To do the segmentation properly, we used a white background. A tripod had been used in capturing those images to avoid the fuzziness of the pictures.

Figure 3.3.1: Collection of both artificially and naturally ripened mangoes

There is a slight difference between artificially ripened and naturally ripened mangoes which can't always be spotted with human eye observation. The naturally ripened mangoes take time to get ripened for this reason they never get total yellow in their ripening process where the artificially ripened mangoes get ripened within a shortened period and they got some extra yellowish texture on them comparing with the artificially ripened ones (Figure 3.3.2).

Artificial Norma Artificial Normal

Figure 3.3.2: Artificially and naturally ripened mangoes

3.4 Statistical Analysis

3.4.1 Background Subtraction Using MATLAB:

Image segmentation has always been defined as "a process of partitioning a digital image into multiple segments" in digital image processing. Image segmentation is mainly implemented to facilitate or change image representation into something which is cabalistic and easier to analyze. In image segmentation, background elimination is important to avoid any other noise. For this reason, in my research a preprocessing stage was implemented to subtract the background. As mentioned earlier I took images of the mangoes over a white background so it made the process of subtraction even easier. At first I separated the three channels i.e. red, green and blue from the images. Then I used global threshold value to proselyte the original image into a binary one. To minimize the intra class variance of black and white pixels, global threshold value is used. (Figure 3.4.1.1).

Now among all other techniques of segmentation I used this one because it can be easily observed that in our images the intensity values are quite dissimilar in different regions but much similar in each regions. For further analysis, the image was then subdivided into significant non-overlapping regions. After the stage of image segmentation, the part of extracting the features from segmented images took place.

Input Image Thresholding Background subtraction

Figure 3.4.1.1: Background Subtraction Process

3.4.2 Feature Extraction Using MATLAB:

Total twenty four (24) mango features had been extracted. These features were – mean value of red channel (R-Mean), median value of red channel (R_Median), highest value of red channel (R_H),lowest value of red channel (R_L), mean value of green channel (G-Mean), median value of green channel (G_Median), highest value of green channel (G_H),lowest value of green channel (G_L), mean value of blue channel (B-Mean), median value of blue channel (B_Median), highest value of blue channel (B_H),lowest value of blue channel (B_L), mean value of hue (H-Mean), median value of hue (H_Median), highest value of hue (H_H),lowest value of hue (H_L), mean value of saturation (S-Mean), median value of saturation (S_Median), highest value of saturation (S_H),lowest value of saturation (S_L), mean value of intensity (I-Mean), median value of intensity (I_Median), highest value of intensity (I_H),lowest value of intensity (I_L).Some of this features were successfully used to detect artificially ripened and naturally ripened mangoes. I have got 24 features of 120 mangoes. Then extracted features are placed at excel and the file saved as .csv format because a machine learning tool weak support .csv format.

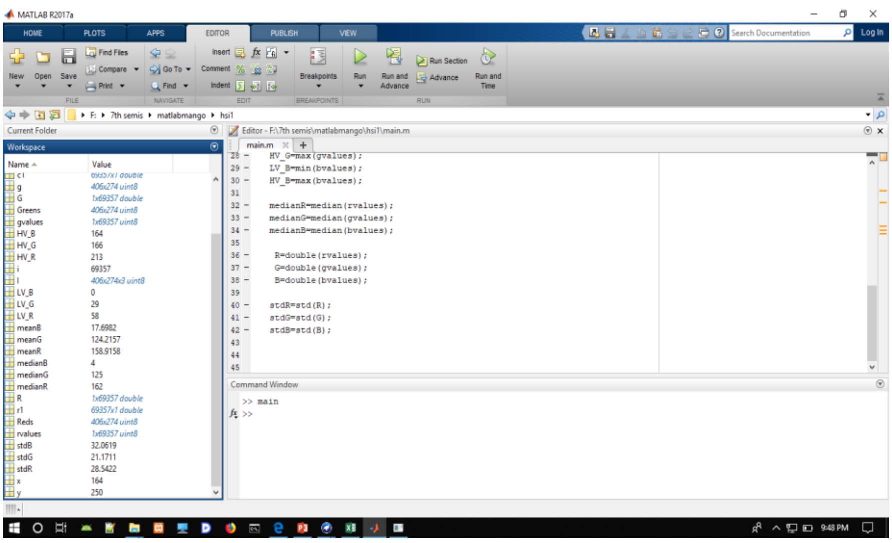


Figure 3.4.2.1: MATLAB code for extracting features from sample images

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	R_Mean	G_Mean	B_Mean	R_Median	G_Median	B_Median	R_H	R_L	G_H	G_L	B_H	B_L	decission
2	159.7414	126.0358	16.6608	162	127	7	216	55	167	41	168	0	artificial
3	131.0936	95.0964	8.415	136	99	2	180	30	137	2	135	0	artificial
4	89.7563	81.9624	11.3813	69	82	3	177	15	141	26	129	0	normal
5	105.2933	135.1148	15.8694	97	133	4	209	19	210	41	174	0	normal
6	136.5295	106.3017	12.3843	139	108	3	185	39	143	37	139	0	artificial
7	121.723	126.2113	11.6299	123	126	2	198	39	195	36	168	0	normal
8	150.191	118.7947	21.7007	155	124	6	213	42	169	20	163	0	artificial
9	176.1612	130.8033	31.2735	182	135	15	231	29	174	25	185	0	artificial
10	95.1921	127.2116	15.401	87	126	4	215	17	202	47	165	0	normal
11	129.5462	149.4472	21.7128	131	158	4	236	38	220	51	197	0	normal
12	166.3792	119.0936	17.373	176	127	9	222	23	156	20	162	0	artificial
13	91.1437	104.9192	18.0191	91	107	9	172	17	169	33	152	0	normal
14	158.1012	169.8113	28.2574	161	162	7	235	52	221	52	201	0	normal
15	166.6741	123.1581	28.9381	170	126	12	223	44	162	33	171	0	artificial
16	115.57	113.7544	16.0788	118	115	3	193	29	181	29	172	0	normal
17	168.3782	127.8023	13.9903	171	133	3	224	72	162	57	169	0	artificial
18	134.0394	117.7199	37.5153	138	112	22	232	38	187	45	200	0	artificial
19	66.2788	79.0176	11.6258	66	80	5	134	11	136	24	118	0	normal
20	128.0227	108.4304	32.3836	131	109	2	221	44	177	37	178	0	artificial
21	131.2448	112.9044	32.1473	130	112	10	223	51	182	34	190	0	artificial
22	75.1355	92.5411	19.7077	69	90	14	153	15	156	28	139	0	normal
23	76.1771	93.3925	20.716	69	91	16	138	15	155	27	159	0	normal
24	122.8834	91.0099	28.8541	123	93	7	194	22	139	11	154	0	artificial
25	136.1354	124.3869	32.354	139	127	10	212	38	171	33	198	0	artificial
26	106.724	107.7983	17.6117	101	105	6	186	36	171	39	162	0	normal
27	132.9748	98.2824	21.1776	133	99	5	222	40	152	32	165	0	artificial
28	85.9651	91.5067	14.4647	82	90	4	173	20	158	19	136	0	normal

Figure 3.4.2.2: Extracted features placed at excel file and saved as .csv file

3.4.3 Feature Selection Using WEKA (Waikato Environment for Knowledge Analysis):

24 mango features regarding RGB and HSI values were extracted for detection process. Feature selection is the process of selecting a subset of pertinent features for using in model construction to avoid the course of dimensionality and enhance generalization by reducing over fitting.

I had used a dataset of 24 features of each mango for selecting the most important features. In this selection filter I used information gain based attribute evaluator. Information gain specifies the attributes with the most information. For this attribute evaluator, ranker search method was used to rank the attributes in accordance with their information gain

The attribute with more information gain is ranked higher. I found some high ranked features by using this attribute evaluator and then top five was selected and they were – R_Mean (mean value of red channel), R_Median (median value of red channel), R_L (lower value of red channel), H_Median (median value of hue), I_Median (median value of intensity).

Then again I had used another technique i.e. classifier subset evaluator to specify some features. Here, I used Best First search. This attribute selector only keeps four attributes. They are – H_Median (median value of hue), R_Median (median value of red channel), I_L (lowest value of intensity) and G_L (lowest value of green channel).After all these feature selection, I applied decision tree, MLP and Naïve Bayes for the classification.

3.4.4 Information gain based feature selection:

The measure of a reduction of uncertainty is called information gain. It represents expected amount of information that would be needed to place a new instance in the branch. Information gain increases with the average purity of the subsets. The strategy is choosing attributes which gives greatest information gain.

How to compute Information Gain:

>> Calculate entropy of the target.

>> Split the dataset on different attributes. Calculate entropy of each dataset. Add it proportionally to get total entropy for the dataset. The result is the Information Gain, or decrease in entropy.

>> Choose attribute with the largest information gain as the decision node, divide the dataset by its branches and repeat the same process on every branch.

I found some features a few of those are R_Mean (mean value of red channel), R_Median (median value of red channel), R_L (lower value of red channel), H_Median (median value of hue), I_Median (median value of intensity). After that I have used these four features for classification measurement using J48 decision tree algorithm.

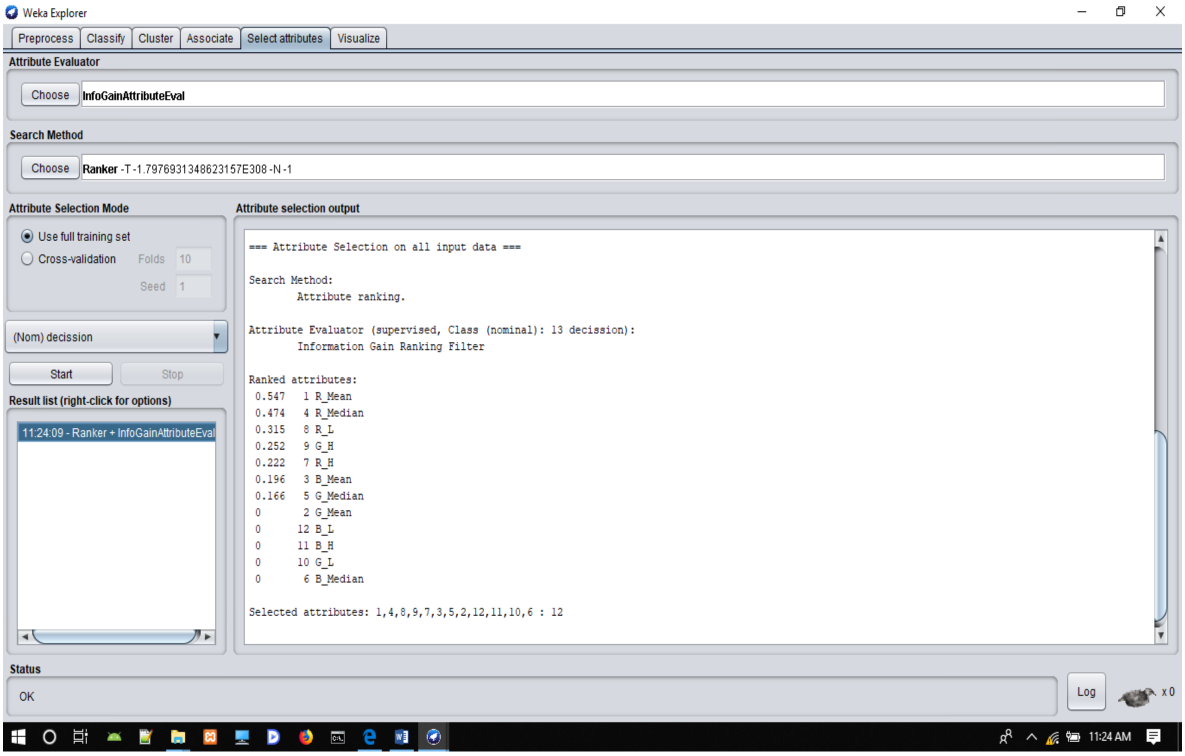


Figure 3.4.4.1: Top ranked features using Information Gain attribute selector and Ranker search method

3.4.5 Classifier subset evaluator based feature selection:

It is the most famous technique for the selection of the most relevant attributes in my dataset which are correlated. A feature is useful if it is correlated with or predictive of the class; otherwise it is irrelevant. We should reduce irrelevant features because they will have low correlation with the class.

So I have observed that Information and Classifier Subset Evaluator selection method has given a slight different attributes.

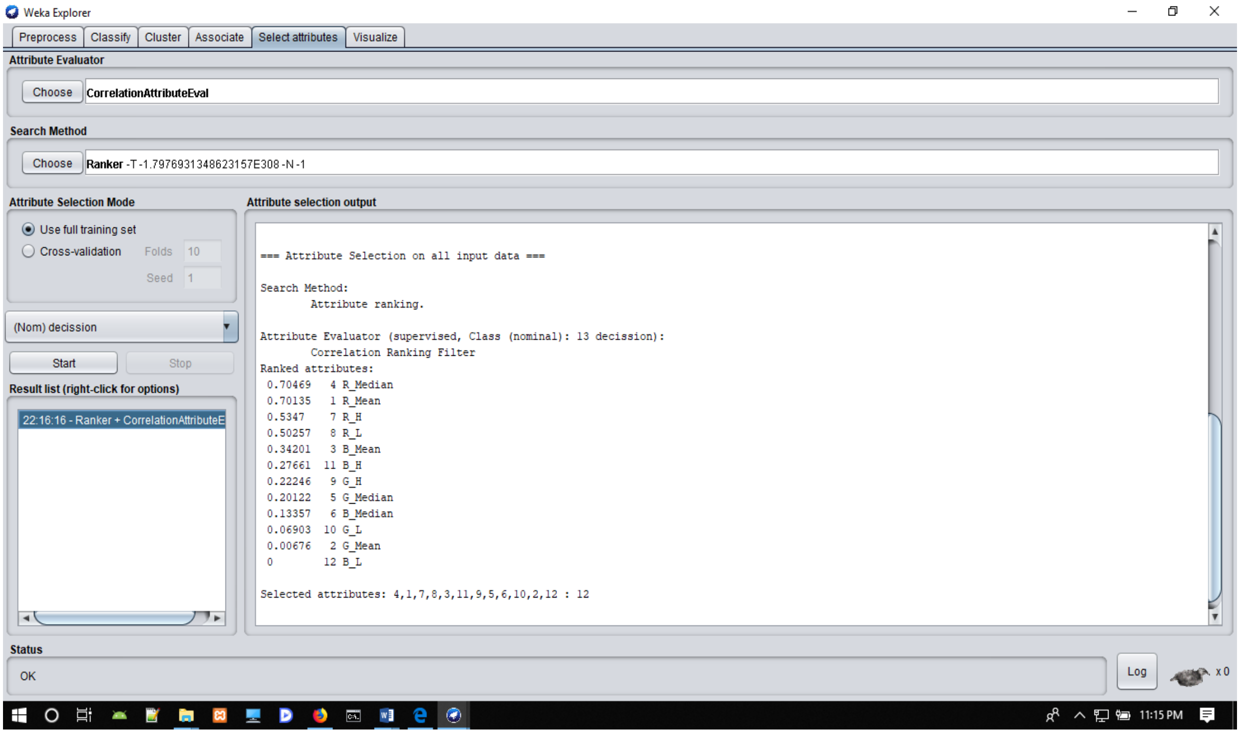


Figure 3.4.5.1: Top ranked features using Correlation based attribute selector and Ranker search method

3.4.6 Classification Using J48 Decision Tree:

In order to classify the artificially ripened mangoes we have used WEKA tools for accurate results. WEKA is an open source Java application produced by the University of Waikato in New Zealand. It is free software licensed under the GNU General Public License. Weka is a collection of machine learning algorithms that contains tools for data preprocessing classification, regression, clustering, association rules and visualization. It provides 49 data preprocessing tools, 76 classification algorithms, 8 clustering algorithms, 15 attribute evaluators, 10 search algorithms for feature selection. WEKA's software platform is java based.

The classifier tool in WEKA is modeled for predicting normal or numeric quantities. Another significant tool of WEKA is attribute or feature selection.

For the identification of artificially ripened mangoes from their images I have used J48 decision tree in WEKA which is a WEKA version of C4.5 algorithm. This algorithm automatically performs "pruning" while building the decision tree.

Decision treebuilds classification or regression models by taking form of a tree structure. It breaks down a data set into smaller and smaller subsets and finally give a result in the form of a tree consisting of decision nodes and leaf nodes. The topmost decision node is often called the root node where the leaf odes often represent the decisions or classifications. Pruning is important as it reduces the size of the tree by turning some branch nodes into leaf nodes. There are several approaches to decision trees like ID3, C4.5,CART etc. but among them ID3 is perfect for splitting nominal values dataset.

ID3 is the acronym for Iterative Dichotomizer 3. Dichotomization means diving a data set into two completely contrary things. With respect to that name the algorithm iteratively divides attributes into two groups based on the most dominant attribute. The most dominant attribute can be found by calculating the entropy and information gain of each attribute. This procedure continues until reaching at a decision for the branch.

C4.5 is a software extension and improved version of the basic ID3 algorithm. As it is the improved version of ID3, there is a slight difference between these two regarding the working procedure. Where the ID3 splits the dataset set based on entropy and information gain, the C4.5 splits the dataset based on the gain ratio. One of some constraints of Id3 is that it can't handle the incomplete data points but as a new version the C4.5 actually can deal with the missing or incomplete data. C4.5 solves over-fitting problems through pruning which is helps the decision tree by increasing its classification accuracy. It automatically performs "pruning" while building the tree.

The J48 decision tree is the WEKA implementation of the standard C4.5 algorithm which is the successor of ID3. WEKA allows the generation of visual vision of the decision tree for the J48 algorithm. For the application of training and testing data, we have used 10 folds cross-validation technique. It breaks data into 10 sets. Then it trains on 9 datasets and tests on 1 dataset. Then it repeats 10 times and takes the mean accuracy.

3.4.7 Features of the Algorithm:

Both the discrete and continuous attributes are handled by this algorithm. A threshold value is decided by C4.5 for handling continuous attributes. This value divides the data list into those who have their attribute value below the threshold and those having more than or equal to it. This algorithm also handles the missing values in the training data.

After the tree is fully constructed, this algorithm performs the pruning of the tree. C4.5 after its construction drives back through the tree and challenges to remove branches that are not helping in reaching the leaf nodes. By applying the above methods, I have got some results which can identify the artificially ripened mangoes. These estimation and evaluation of the results are discussed in the next chapter.

According to Information Gain based attribute selection we see that about 41 mangoes are correctly classified out of 49 mangoes and accuracy is 83.673%.

According to Classifier subset based attribute selection we see that about 42 mangoes are correctly classified out of 50 mangoes and accuracy is 85.7134%.

3.4.8 Classification Using Multilayer Perceptron

Multilayer perceptron’s are networks of perceptron’s, networks of linear classifier. In fact, they can implement arbitrary decision boundaries using “hidden layers .Weka has a graphical interface that lets you create your own network structure with as many perceptron’s and connections as you like.

Artificial Neural Network (ANN) is a computing system that is inarticulately inspired by the animal’s biological brain’s neural network [12]. ANN consists of a activation function which results in output. The most basic form of an activation function is binary function that gives output either as 0 or 1.It is mostly known as perceptron. Two types of perceptron’s are generally known, one is single layered perceptron and the other is multilayer perceptron. Perceptron with one layer i.e. single layer perceptron’s are limited with learning only linearly separable patterns where the perceptron with multiple layers i.e. multilayer perceptron has far greater processing power than the single layer perceptron. Multilayer perceptron is efficient in finding nonlinear associations between input vectors and output vectors, based on training sets.

MLP generally consists of three layer i.e. input layer, hidden layer (s) and output layer. Each connection in the network has a weight and each node which are often called as neurons, in the network performs a weighted sum of its inputs and thresholds the result

Usually with a sigmoid function. It has an input layer where there is only one for each attribute and an output layer where there is one for each class and in case of numeric class the number of class is one. Before directing from input to output layer, the hidden layer does intermediate computation. There can be different hidden layers such as zero hidden layer, one hidden layer, and two hidden layers and so on. The number of hidden layer mostly depends on the data set or the algorithm. Having zero hidden layers is a standard perceptron algorithm and it is suitable if data is linearly separable. One hidden layer is suitable for single convex region of the decision space where two hidden layer can generate arbitrary decision boundaries. The more the number of hidden layer, the more the accuracy is:

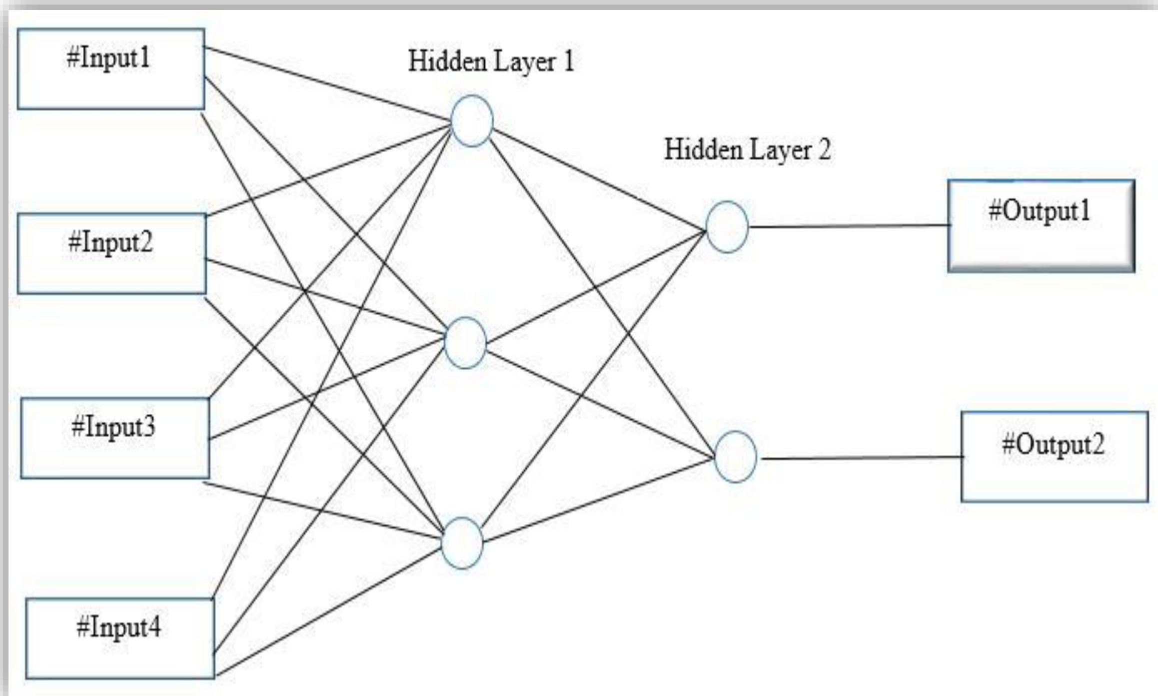


Figure 3.4.8.1: Neural Network based on Multilayer Perceptron

There is a term often used in MLP is weight. Weights are learned from the training set and it iteratively minimize the error using steepest descent. The "backpropagation" algorithm is used to determine the gradient. Change in weight computed by multiplying the gradient by the "learning rate" and adding the previous change in weight multiplied by the "momentum":

Wnext = W + ΔW

$\Delta W = -\text{learning rate} \times \text{gradient} + \text{momentum} \times \Delta W_{previous}$

Excellent results can be obtained by experimenting the numbers and sizes of hidden layers and the values of learning rates and momentums.

Regarding the processing power of MLP, I splatted our dataset into two sets ,one of which was training data and the rest one was testing data by using 10 folds cross validation. By applying MLP I got a visual neural network and some results which showed the identification of artificially ripened and naturally ripened mangoes though a confusion matrix and gave the percentage of accuracy of this method. It was observed that the accuracy was more than decision tree technique.

3.4.9 Classification using Naïve Bayes Classifier

Naive Bayes classifiers is a probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful in the field of medical science for diagnosing heart patients.

Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods. Bayes theorem provides a way of calculating the posterior probability, P (c|x), from P(c), P(x), and P (x|c). Naive Bayes classifier assumes that the effect of the value of a predictor. On a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \ldots.. \times P(x_n \mid c) \times P(c)$

Where,
P (c|x) is the posterior probability of class (target) given predictor (attribute).

P(c) is the prior probability of class.
P (x|c) is the likelihood which is the probability of predictor given class.
P(x) is the prior probability of predictor
Where C and X are two events (e.g. the probability that the train will arrive on time given that the weather is rainy). Such Naïve Bayes classifiers use the probability theory

to find the most likely classification of an unseen (unclassified) instance. The algorithm performs positively with categorical data but poorly if we have numerical data in the training set.

3.5 Implementation Requirements

In the proposed model a large number of mangoes will be categorized according to their presence of calcium carbide in fruit body automatically using several equipment's like camera, conveyor belt, and computer and by taking the help of a controller and switch selector.

At first the collected sample mangoes will be placed on the conveyor belt. The conveyor motor will work and the camera will wait for an object to take the picture. That camera should be placed on the top of the conveyor belt. It will continuously check if there is an object (mango). When it will detect an object the conveyor motor will stop for seconds and the image of the mango will be captured by the camera. After that the conveyor motor will start again and the color features of the mango image will be extracted. By evaluating those color features (RGB and HSI values). The presence of calcium carbide in the fruit body will be identified. Then the mangoes will be packed according to the consequence by taking the assistance of a controller and switch selector (Figure 3.5.1).

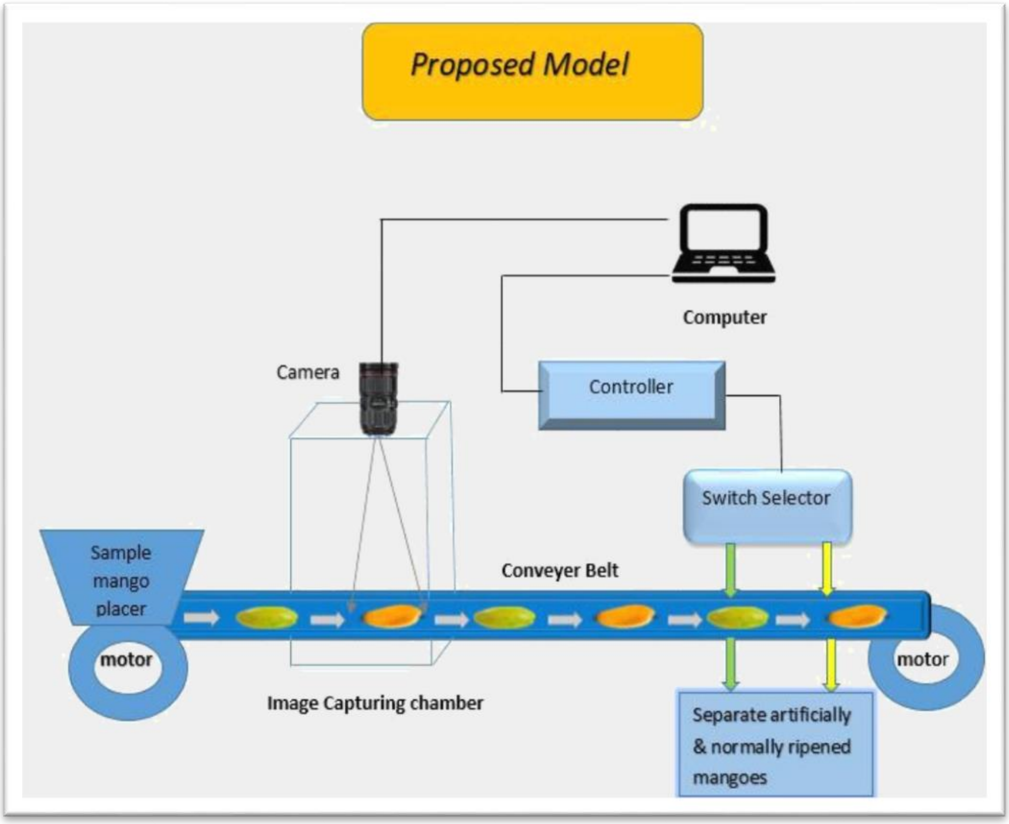


Figure 3.5.1: Proposed automated system for mango categorization

The proposed system is a completely automated system where thousands of mangoes can be classified according to the presence of chemicals in fruit body. By implementing such an automated system, an industry can sort mangoes with very short time and can ease their working procedure. The overall cost of the industries will be reduced and effectiveness will be increased.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1 Introduction

The objective of this research is to categorize artificially ripened mangoes and naturally ripened mangoes based on their images so that an automated system of mango classification can be built. To meet that expectation, I have implemented several techniques. At first the mango images were preprocessed to eliminate the background, after that 24 RGB and HSI features were extracted for each mango image. For better result, the RGB values were converted into HSI values so that the human perception is reflected on the classification measurement.

To increase performance, attributes have been analyzed using a decision tree, multilayer perceptron and naïve Bayes in Weka. For the selection of the most significant attributes two types of evaluation methods of supervised attribute selection filter were used. One of them is Info Gain Attribute with ranker search and the other one is CfsSubsetEval attribute evaluator with Best First search. I have been able to separate the important features from those 24 attributes by applying the two attribute evaluators and those key features of classification are then used for J48 algorithm, Multilayer Perception and Naïve Bayes algorithm in Weka. 10 fold cross-validation has been used to specify the training and testing data.

I have discussed the method and techniques in details in chapter 3. In this chapter I have evaluated the results that we obtained by using those approaches and by the analysis of the accomplished features. The estimated accuracy of different methods has also been shown here and finally comparing the accuracy of the three we have suggested the best method for our dataset to identify the artificially ripened mangoes.

4.2 Experimental Results

4.2.1 Results of Classification Using Decision Tree in WEKA

In order to classify mangoes in WEKA, 24 features regarding RGB and HSI values of mango images have been taken reserved. For the reduction of data dimension two types of attribute evaluator of supervised attribute selection filter have been used. After the selection of the most informative attributes. J48 decision tree of WEKA has been used for classification measurement. To specify the training and testing data, 10 folds cross validation technique is employed. The results show that mangoes are effectively classified using decision tree. A sample of 120 mango images have been used in this case.

4.2.1.1 Decision tree with attribute evaluator CfsSubsetEval

For the selection of most significant features CfsSubsetEval attribute evaluator has been used with Best first search. It identified four important features H_Median (median value of hue), Median (median value of red), I_L (lower value of hue) and G_L (lower value of green) upon which the decision tree is built using J48 decision tree algorithm of WEKA. The following is the decision tree which is the graphical representation of the result of the classification measurement based on those selected features.

The decision tree (Fig.4.1) consists of H_Median (median value of green channel), R_Median (median value of red channel), I_L (lower value of intensity), G_L (lower value of green channel) as root and internal node where the leaf nodes represent categorization of artificially and normally ripened fruits. Among 120 mangoes 103 mangoes were accurately determined either artificial or normal. As a result, the accuracy was 85.7143%.After that we supplied the model a test set of 49 mangoes to verify the accuracy. The succeeding confusion matrix was obtained from the test set (Table 4.2.1.1.1).

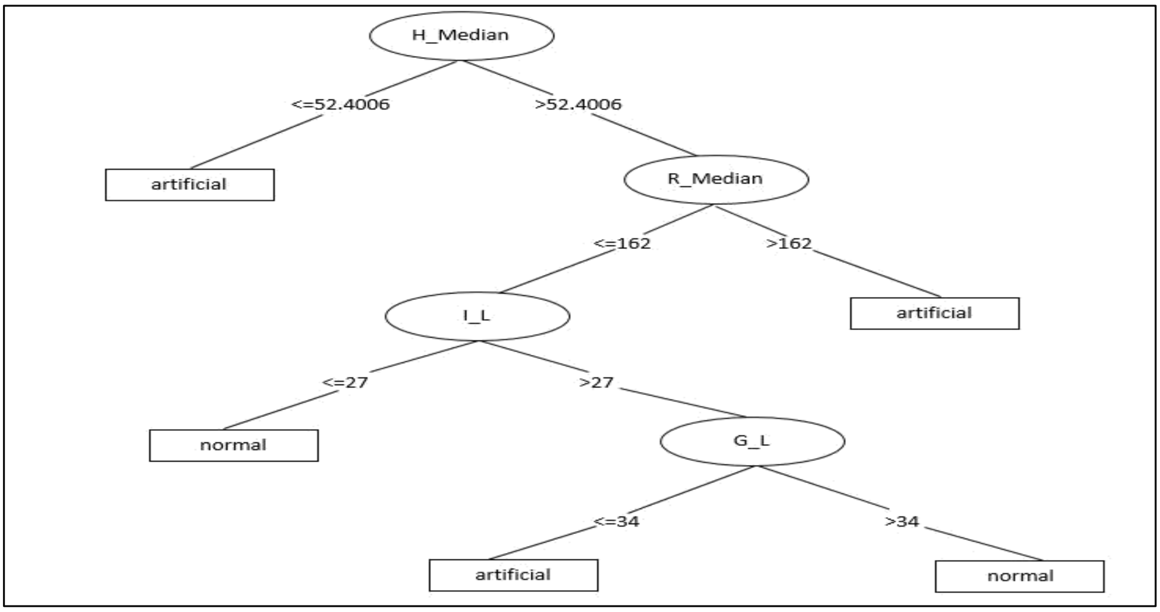


Figure 4.2.1.1.1: Decision tree based upon four selective attributes H_Median, R_Median, I_L, and G_L

The confusion matrix shows that among 26 artificially ripened mangoes, 22 mangoes were predicted correctly and the rest 4 mangoes were misclassified. And among 23 normally ripened mangoes, 20 mangoes were predicted correctly where the rest 3 were misclassified.

Table 4.2.1.1.1: Confusion matrix of test set after being trained by 10 folds cross-validation technique (classifier subset evaluator)

Predicted Class / Actual Class	Artificial	Normal
Artificial	22	4
Normal	3	20

4.2.1.2 Decision tree with attribute evaluator InfoGain Attribute Eval

For the selection of most significant features another attribute evaluator InfoGainAttributeEval has been used with Ranker search. It identified four important features R_Mean (mean value of red channel), R_Median (median value of red channel), R_L (lower value of red channel), H_Median (median value of hue), I_Median (median value of intensity) upon which the decision tree is built using J48 decision tree algorithm of WEKA. The following is the decision tree which is the graphical representation of the result of the classification measurement based on those selected features.

The decision tree (Fig.4.2) consists of H_Median (median value of hue) and R_Median (median value of r channel) as root and internal node where the leaf nodes represent categorization of artificially and normally ripened fruits. Among 120 mangoes 97 mangoes were accurately determined either as artificial or as normal. As a result, the accuracy was 83.6735%.After that we supplied the model a test set of 49 mangoes to verify the accuracy. The succeeding confusion matrix was obtained from the test set (Table 4.2.1.2.1).

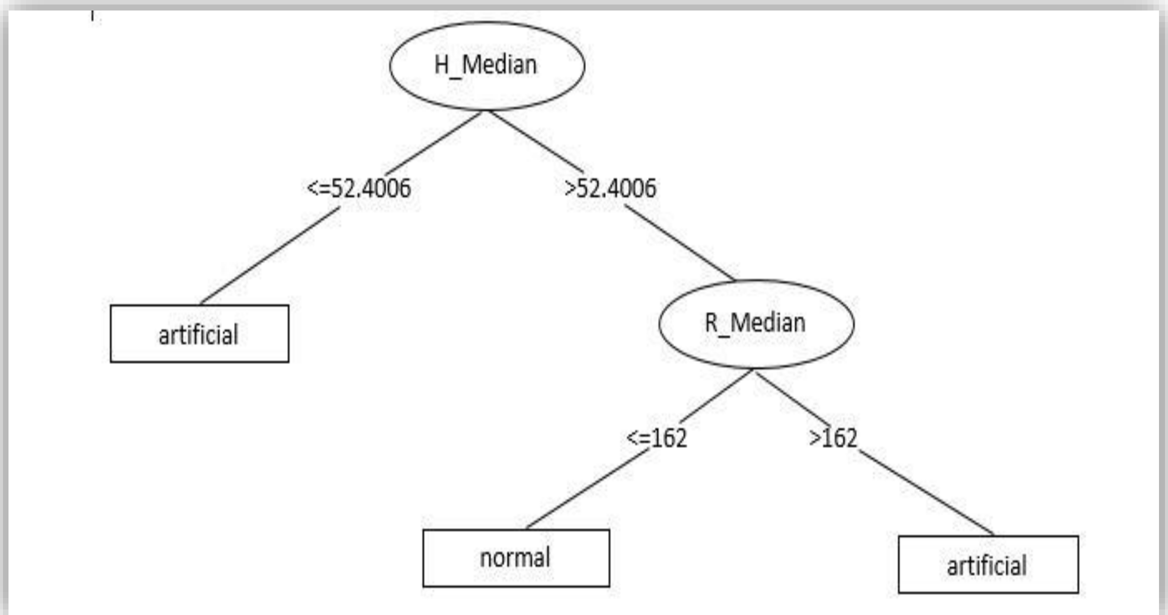


Figure 4.2.1.2.1: Decision tree based upon four selective attributes H_Median, R_Median.

The confusion matrix exhibits that among 26 artificially ripened mangoes, 22 mangoes were predicted correctly and the rest 4 mangoes were misclassified. And among 23 normally ripened mangoes, 19 mangoes were predicted correctly where the rest 4 were misclassified. Now the accuracy is 91percent.

Table 4.2.1.2.1: Confusion matrix of test set after being trained by 10 folds cross-validation technique (Information gain)

Predicted Class / Actual Class	Artificial	Normal
Artificial	22	4
Normal	4	19

4.2.2 Results of Classification Using MLP in WEKA

In order to classify mangoes in WEKA, 24 features regarding RGB and HSI values of mango images have been taken into account. For the reduction of data dimension two types of attribute evaluator of supervised attribute selection filter have been used. After the selection of the most informative attributes. Then another technique which is Multilayer Perceptron of WEKA has been used for classification measurement. To specify the training and testing data, 10 folds cross validation technique is employed. The results show that mangoes are effectively classified using decision tree. A sample of 120 mango images have been used in this case.

4.2.2.1 Multilayer Perceptron with attribute evaluator CfsSubsetEval

Now, applying another technique i.e. multilayer perceptron detected artificially ripened mangoes (Fig.4.3). By using this technique 112 among 120 mangoes were accurately determined either as artificial or as normal. Here, the accuracy was 93.8367%. After that we supplied the model a test set of 49 mangoes to verify the accuracy. The succeeding confusion matrix was obtained from the test set.

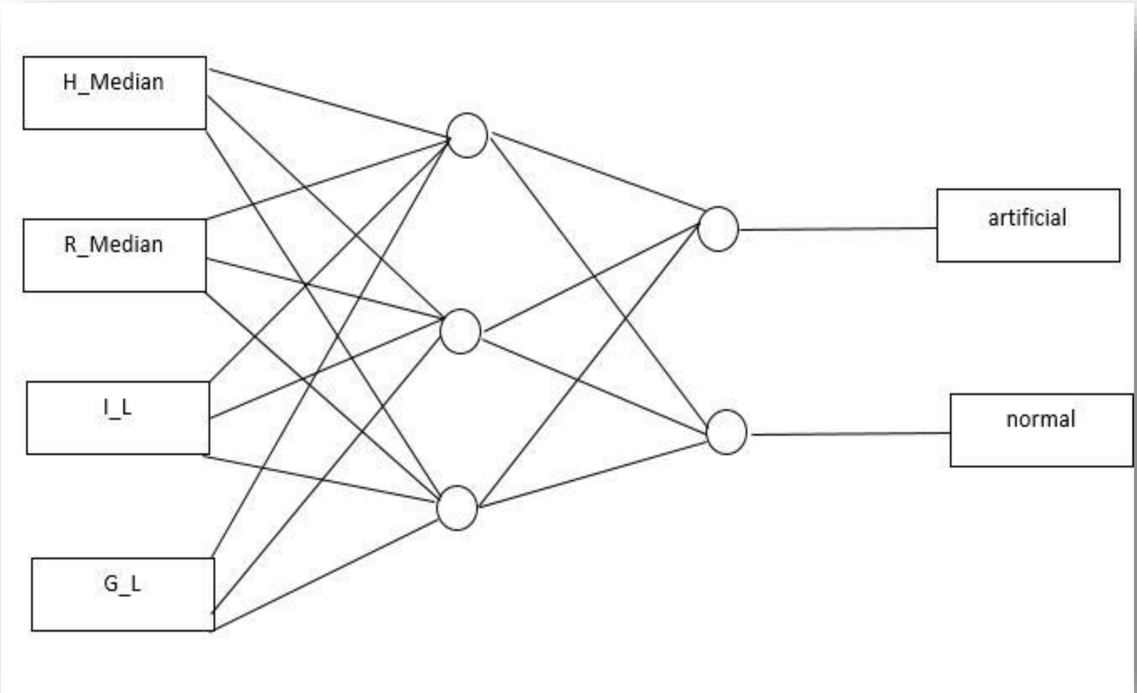


Figure 4.2.2.1.1: Neural Network based on Multilayer Perceptron with classifier subset evaluator

The confusion matrix shown in (Table 4.2.2.1.1) demonstrates that among 26 artificially ripened mangoes, 24 mangoes were predicted correctly and the rest 2 mangoes were misclassified. And among 23 normally ripened mangoes, 21 mangoes were predicted correctly where the rest 2 were misclassified. The accuracy was 95 percent.

Table 4.2.2.1.1: Confusion matrix of test set after being trained by 10 folds cross validation technique (MLP)

Predicted Class / Actual Class	Artificial	Normal
Artificial	24	2
Normal	2	21

4.2.2.2 Multilayer Perceptron with attribute evaluator Infogain

Now, applying another technique multilayer perceptron for detecting artificially ripened mangoes (Fig.4.4). By using this technique 105 among 120 mangoes were accurately determined either as artificial or as normal. Here, the accuracy was 89.667%. After that we supplied the model a test set of 49 mangoes to verify the accuracy. The succeeding confusion matrix was obtained from the test set

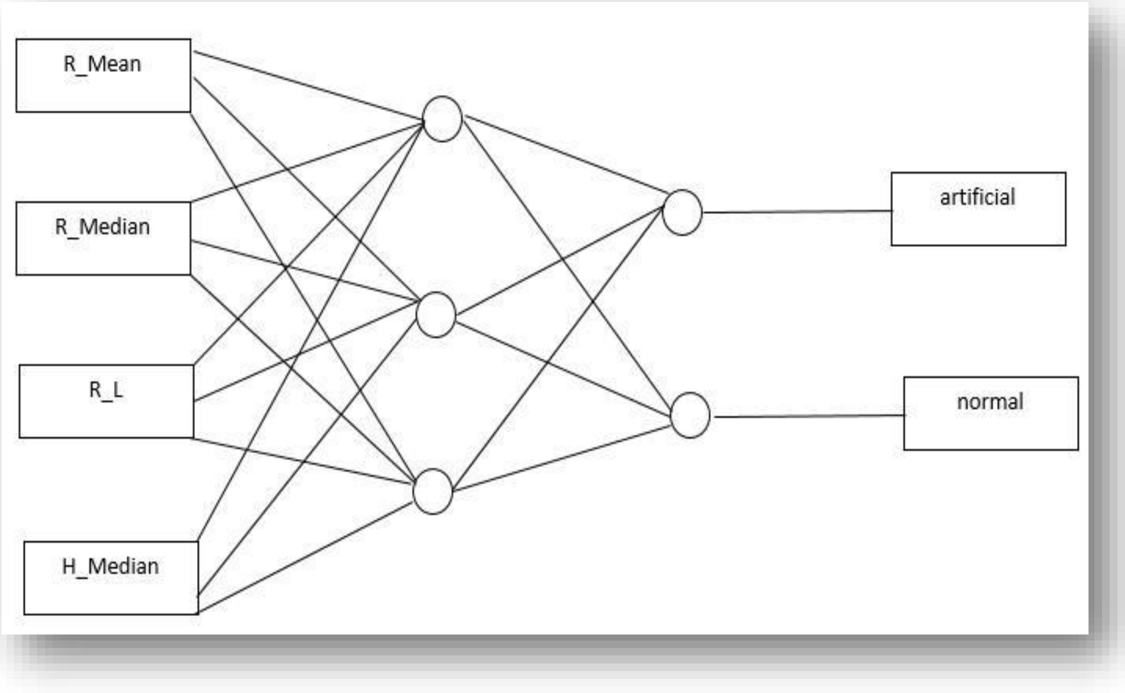


Figure 4.2.2.2.1: Neural Network based on Multilayer Perceptron with Information gain attribute evaluation

The confusion matrix shown in (Table 4.2.2.2.1) demonstrates that among 26 artificially ripened mangoes, 24 mangoes were predicted correctly and the rest 2 mangoes were misclassified. And among 23 normally ripened mangoes, 22 mangoes were predicted correctly where the rest 1 was misclassified. The accuracy was then 93 percent.

Table 4.2.2.2.1: Confusion matrix of test set after being trained by 10 folds cross-validation technique (MLP)

Predicted Class / Actual Class	Artificial	Normal
Artificial	25	1
Normal	1	22

4.2.3 Results of Classification Using Naïve Bayes in WEKA

In order to classify mangoes in WEKA, 24 features regarding RGB and HSI values of mango images have been taken into account. For the reduction of data dimension two types of attribute evaluator of supervised attribute selection filter have been used. After the selection of the most informative attributes. Then another technique which is Naïve Bayes technique of WEKA has been used for classification measurement. To specify the training and testing data, 10 folds cross validation technique is employed. The results show that mangoes are effectively classified using decision tree. A sample of 120 mango images have been used in this case.

4.2.3.1 Naïve Bayes with attribute evaluator CfsSubsetEval

Now, applying another technique i.e. Naïve Bayes detected artificially ripened mangoes. By using this technique 112 among 120 mangoes were accurately determined either as artificial or as normal. Here, the accuracy was 91.837%. After that we supplied the model a test set of 49 mangoes to verify the accuracy. The succeeding confusion matrix was obtained from the test set.

Table 4.2.3.1.1: Confusion matrix of test set after being trained by 10 folds cross-validation technique (Naïve Bayes)

Predicted Class / Actual Class	Artificial	Normal
Artificial	24	2
Normal	2	21

The confusion matrix shown in (Table 4.2.3.1.1) demonstrates that among 26 artificially ripened mangoes, 24 mangoes were predicted correctly and the rest 1 mangoes were misclassified. And among 23 normally ripened mangoes, 21 mangoes were predicted correctly where the rest 2 were misclassified. The accuracy was 91.837 percent.

4.2.3.2 Naïve Bayes with attribute evaluator Info Gain Attribute Eval

Now, applying another technique i.e. Naïve Bayes detected artificially ripened mangoes. By using this technique 112 among 120 mangoes were accurately determined either as artificial or as normal. Here, the accuracy was 93.8367%. After that we supplied the model a test set of 49 mangoes to verify the accuracy. The succeeding confusion matrix was obtained from the test set.

Table 4.2.3.2.1: Confusion matrix of test set after being trained by 10 folds cross-validation technique (Naïve Bayes)

Predicted Class / Actual Class	Artificial	Normal
Artificial	25	1
Normal	2	21

The confusion matrix shown in (Table 4.2.3.2.1) demonstrates that among 26 artificially ripened mangoes, 25 mangoes were predicted correctly and the rest 1 mangoes were misclassified. And among 23 normally ripened mangoes, 21 mangoes were predicted correctly where the rest 2 were misclassified. The accuracy was 93.87 percent.

4.3 Descriptive Analysis

I have observed that the accuracy increases in case of J48 decision tree if I use the pruned tree while by using unpruned tree the accuracy get decreased. I see variations in result because of applying different techniques. If I use information gain based attribute evaluator, I will get an accuracy of 91% for J48 decision tree and 93% for multilayer perceptron. The accuracy of J48 decision tree will increase to 93% (same as MLP of information gain) and accuracy of multilayer perceptron also increase to 95% if we use classification subset evaluator (Fig.4.5). But in case of MLP the time taken to build the model is more than that of decision tree. If I consider Naïve Bayes then it can be observed that it takes time less than decision tree and gives accuracy more than decision tree. So, ultimately I can come to the conclusion that for our dataset Naïve Bayes is the best classifier as it gives more accuracy and takes less time compare to decision tree and multilayer perceptron (Table 4.3.1).

The time comparison has been shown in Table 4.3.1 is given below

Table 4.3.1: Time taken to build the model

Attribute Selector	J48 Decision Tree	Multilayer Perceptron	Naïve Bayes
Information Gain	0.02sec	0.08sec	0 sec
Classifier Subset Svaluator	0.02sec	0.19sec	0 sec

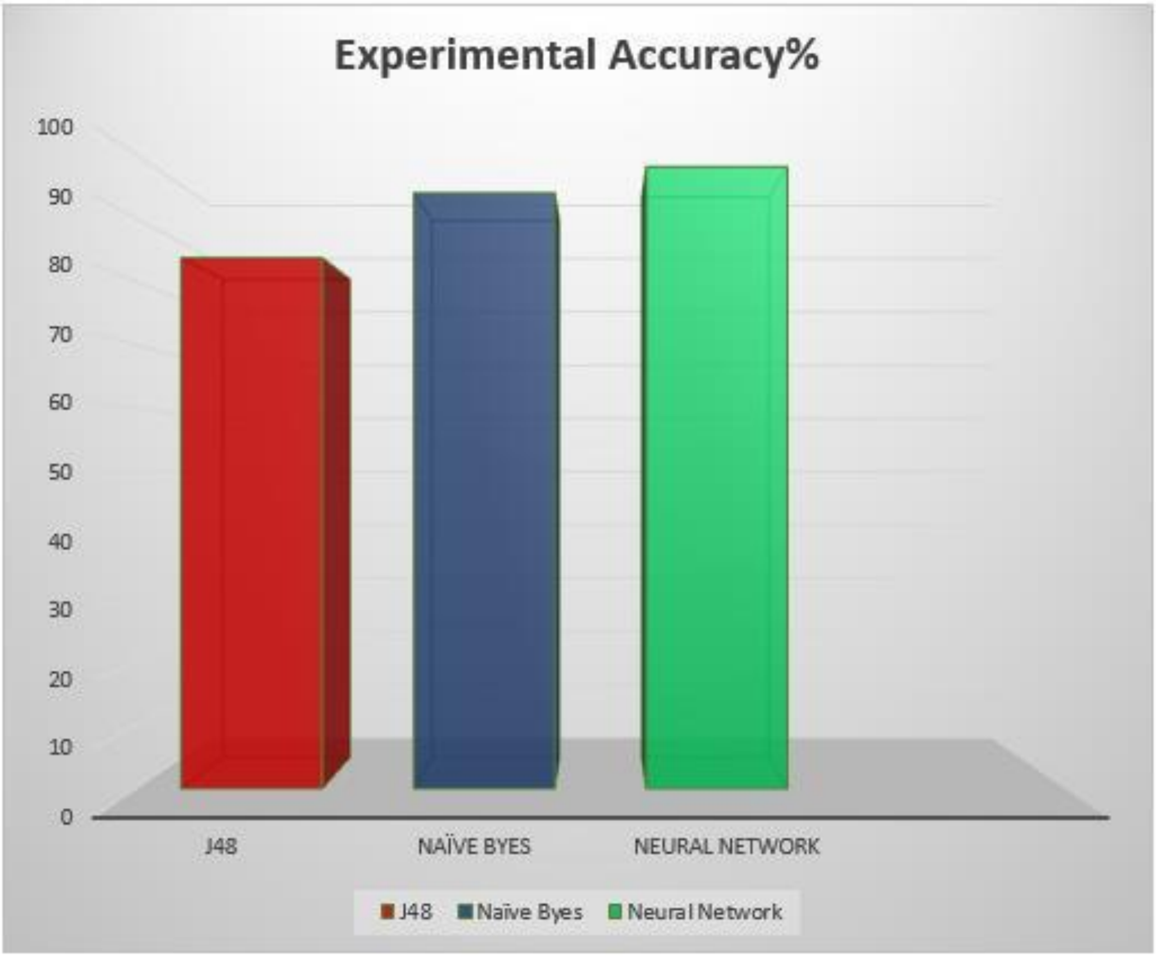


Figure 4.3.1: Bar chart showing the accuracy of J48 decision tree, Naïve Byes and Neural Network

If I consider both run time of each method and the corresponding accuracy then ultimately Naïve Bayes technique shows more promising results than Decision Tree and Multilayer Perceptron (Fig.4.6).

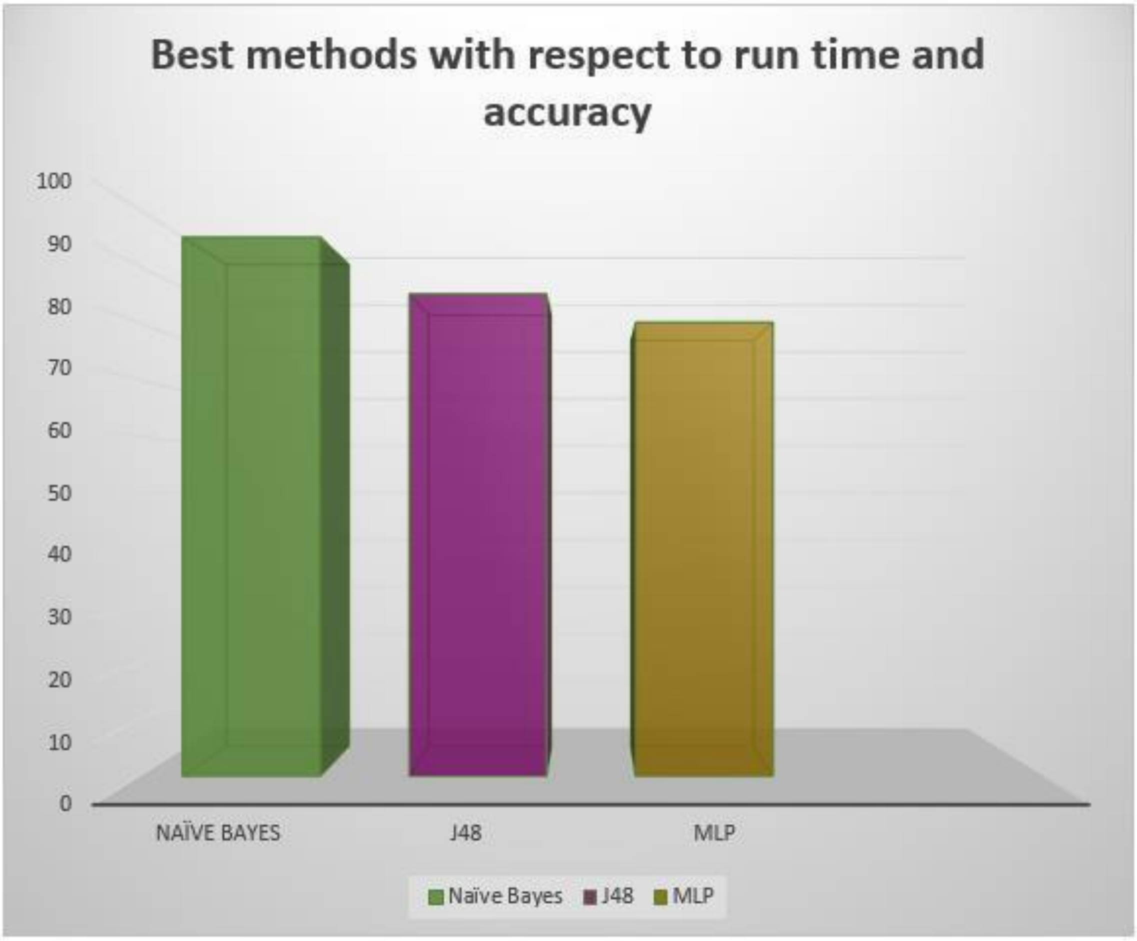


Figure 4.3.2: Ranking of three methods with respect to accuracy and run time

When I used 55 sample mangoes for classification using J48 decision tree in WEKA, the accuracy was only 43.786 % for both CfsSubsetEval and InfoGainAttributeEval by which the top 3 features were selected. While we increased the number of sample images to 120, the accuracy was amplified. However, it can be said that training data need to be huge in order to obtain more accurate and error free result. Picture quality plays a great role in defining the accuracy of classification specially the intensity of light affects the results most. Some of misclassification happens because reason. The same mango with dissimilar brightness has the chance to be classified in different stages. For this reason, the influence of illumination should be considered seriously and special precautions must be taken during image acquisition in order to avoid this effect. Another concern is the dark spots on the mangoes which actually appear in case of diseased mangoes.

4.4 Summary

In this chapter, the results and accuracy using our methods has been discussed and compared. A best method according to the comparison of accuracy has also been discussed in this chapter.

CHAPTER 5

CONCLUSION

5.1 Summary of the Study

The objective of our research is to implement a digital analysis technique by which the mangoes can be classified effectively according to the presence of chemical in the fruit body. My research is about categorizing popular mango variet “Amrapali. Skin color of mangoes play a vital role in the estimation of chemicals in fruit body. The color of Amrapali varies considerably stage by stage till it gets totally matured. However, the classification is done based on the RGB and HSI color model of those mango images. The analysis of classification clearly showed that there are significant changes in red, green and blue color components of RGB color model and hue-saturation-intensity (HSI) value between artificially ripened mangoes and normally ripened mangoes.

5.2 Conclusions

Health is valuable more than anything in the world for a human being. No one wants to compromise with their health. Artificially ripened fruits especially mangoes are literally a threat to human health as they contain carcinogenic properties. So it is very essential for us consumers to identify which mangoes are artificially ripened and which are normally ripened.

Generally, it is kind of impossible for naked eye to identify artificially ripened mangoes among thousands of mangoes. Also this process is susceptible to error due to the distraction, tiredness and tedium of experts. For these reasons, an automated system of mango categorization is necessary so that industries can be able to process huge amount of mangoes within very short time and within very low cost. That automated system should be applied to mangoes before that mangoes reach the consumers. It should be strictly maintained that no mango would enter into market before be checked by that automated system.

Therefore the objective of my research is to implement a digital analysis technique by which the mangoes can be classified effectively according to the presence of chemical in the fruit body. My research is about categorizing popular mango variet “Amrapali. Skin color of mangoes play a vital role in the estimation of chemicals in fruit body. The color of Amrapali varies considerably stage by stage till it gets totally matured. However, the classification is done based on the RGB and HSI color model of those mango images. The analysis of classification clearly showed that there are significant changes in red, green and blue color components of RGB color model and hue-saturation-intensity (HSI) value between artificially ripened mangoes and normally ripened mangoes.

Several computer vision and image processing approaches are explored in this thesis. My research is consisted of three main portions that are (1) image acquisition), (2) feature subtraction and feature selection, (3) training and classification. In the first step, sample images of both artificially and normally ripened mangoes were collected. The second step was acquired by extracting 24 features from the sample images and selecting the top most informative features based on information gain with ranker search method and classifier subset evaluator with Best First search. Finally in the third step, images were classified between artificially ripened mango and normally ripened mango using decision tree and multilayer perceptron. Though it can be observed that by using MLP I got more accuracy than decision tree and naïve bayes, the MLP took more time to build the model than the decision tree and naïve bayes. Now among decision tree and naïve bayes, naïve bayes got more accuracy than decision tree. Here running time got more precedence than the accuracy. By observing all these, it can be said that overall naïve bayes is more advantageous than multilayer perceptron and decision tree for my research data.

5.3 Recommendations

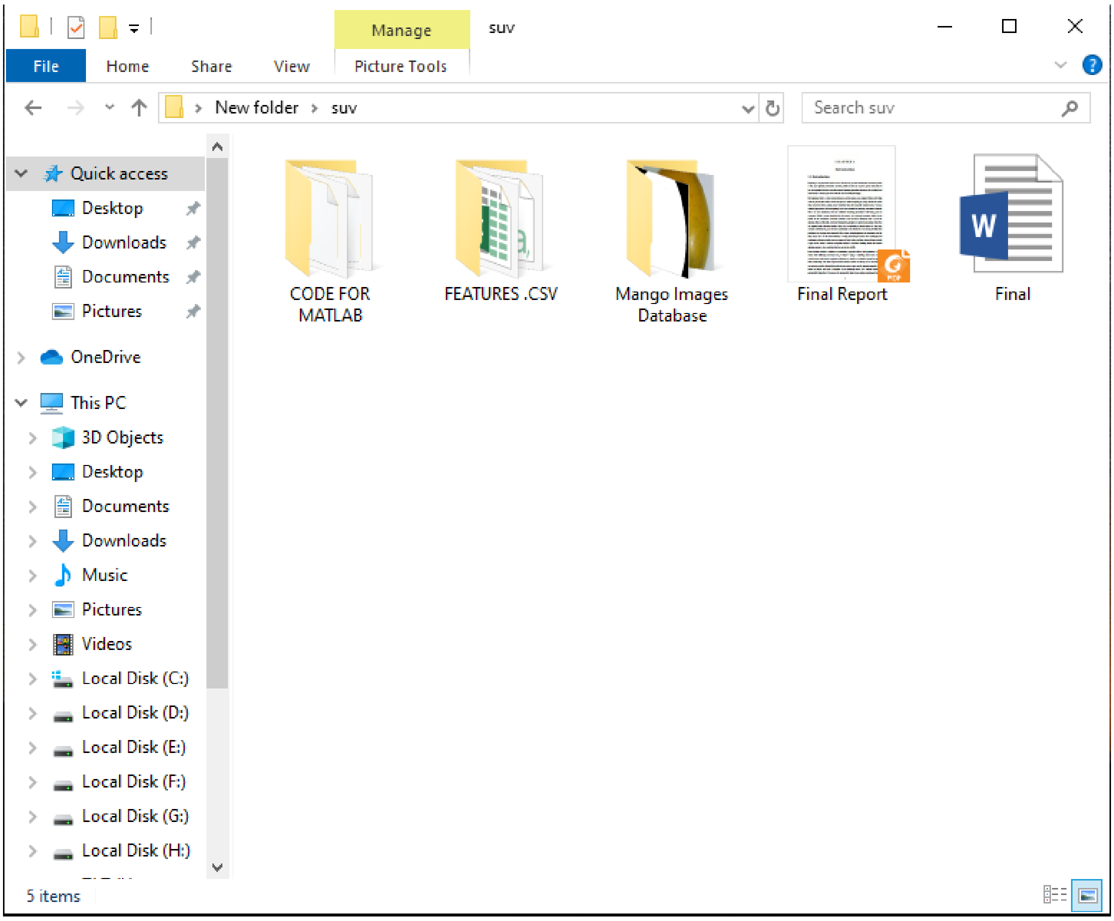
As it is proved that naïve bayes classification got more accuracy and less running time than any other classification, this classification should be taken into account and should be implemented on different species of fruits that are endangered with artificially ripening agents. In addition, medical sections should implement antidotes for calcium carbide effects and techno sections should improvised the proposed method that I’ve implemented.

5.4 Implication for Further Study

In future, my aim will be collecting huge amount of data so that the percentage of accuracy can be extended. Also, in future research I would take necessary precautions to avoid the effect of dissimilar light intensity which is foremost responsible for the errors.

In addition, I would like to implement some classification methods for the mangoes which do not change colors with the ripening process that is “Fajli” mangoes.

APPENDIX


Manage
suv
File
Home
Share
View
Picture Tools
New folder
suv
Search suv
Quick access
Desktop
Downloads
Documents
Pictures
OneDrive
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
Local Disk (C:)
Local Disk (D:)
Local Disk (E:)
Local Disk (F:)
Local Disk (G:)
Local Disk (H:)
CODE FOR MATLAB
FEATURES .CSV
Mango Images Database
Final Report
Final
5 items

REFERENCES

[1] ClaudlaLINTAS.,"Nutritional aspects of fruit and vegetable consumption", Istituto Nazionaledella Nutrizione, Rome(Italy),1992.

[2] Mohammad Asif, "Physico-chemical properties and toxic effect of fruit-ripening agent calcium carbide", Annals of Tropical Medicine and Public Health Vol.5, No.1, pp 150-156, 2012.

[3] DAILY TRUST. Available: https://www.dailytrust.com.ng/health-implications-of-eating-fruits-ripened-with-calcium-carbide.html [Accessed 14 August 2018].

[4] THE LOGICAL INDIAN. Available: https://thelogicalindian.com/story-feed/awareness/artificial-ripening-of-fruits/ [Accessed 15 September 2018].

[5] Rahul, Aniket, "Image Processing for Mango Ripening Stage Detection: RGB and HSV method", Third International Conference on Image Processing pp 362-365, 2015.

[6] MehnazMursalat, AsifHasanRony, AbdulHasnat, Md.SazedurRahman, Md.Nazibul. "A Critical Analysis of Artificial Fruit Ripening: Scientific, Legislative and Socio-Economic Aspects". Chemical Engineering & Science Magazine Vol. 4 No 1. 2013

[7] Asil Hossain, M. F Akhtar, Anwar. "Health hazards posed by the consumption of artificially ripened fruits in Bangladesh". International Food Research Journal 22(5) pp 1755-1760, 2015

[8] Mansor,A.R., Mahmod,o.,Nazari, A.B.M., "Fuzzy ripening mango index using RGB colour sensor model" Res. World J. Art Sci. Commer. Vol.120, pp-6387-6396, 2014.

[9] Nandi, C.S., Tudu,B., Koley, C, "A machine vision technique for grading of harvested mangoes based on maturity and quality", International Conference on Sensing Technology ,pp-6387-6396, 2016

[10] Sapan Naik, Patel, "Machine Vision based Fruit Classification and Grading – A Review" International Journal of Computer Applications(0975-8887),Vol.170 No.9,July 2017.

[11] Vanloot, Bertrand, Pinatel, Artaud, Dupuy "Artificial vision and chemo metrics analyses of olive stones for varietal identification of five French cultivars" Computer Electronics Agriculture Vol.102, pp-98-105, 2014.

[12] Karthika,Maragaevi,Asvini. "Detection of artificially ripened fruits using image processing", International journal of Advance Science and Technology Research Vol.2 No.1, 2017

[13] S Maheswaran, S.Sathesh, P. Priyadharshini, B. Vivek. "Identification of Artificially Ripened Fruits Using Smart Phones", International Conference on Intelligent Computing and Control (I2C2) pp 1-6, 2017.

[14] Mim,Md. Galib,Jerin. “An automatic detection of mango ripening stages an application of information technology to botany” ScientiaHolticulturae 237 pp 156-163, 2018.

[15] Hall, M.A., Smith,L.A., “Practical Feature Subset Selection for Machine Learning”, 1998.

[16] G.M. Foody. “Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes”, International Journal of Remote Sensing Vol.25 No.15 pp 3091-3104, 2000.

[17] ALL ABOUT CIRCUITS. https://www.allaboutcircuits.com/technical-articles/understanding-color-models-used-in-digital-image-processing/ [Accessed 26 September 2018]

[18] WIKIPEDIA https://en.wikipedia.org/wiki/Artificial_neural_network [Accesses 26 September, 2018]

Turnitin Originality Report

Processed on: 02-Nov-2019 20:00 +06
ID: 1205501002
Word Count: 9872
Submitted: 1

Shu By Nusrat Jahan

Similarity Index	Similarity by Source	
28%	Internet Sources:	23%
	Publications:	21%
	Student Papers:	14%

6% match (publications)
Farjana Sultana Mim, Syed Md. Galib, Md. Fakhrul Hasan, Sanjida Akther Jerin. "Automatic detection of mango ripening stages – An application of information technology to botany", Scientia Horticulturae, 2018

6% match (Internet from 26-Jul-2018)
https://www.omicsonline.org/open-access/digital-measurement-of-maturity-indices-of-mangoes-using-selected-image-features-2157-7110-1000701-95698.html?view=mobile

1% match (Internet from 31-Jul-2017)
https://www.rroij.com/open-access/advanced-multimedia-answer-generation-byscraping-information-through-web.php?aid=47532

1% match (Internet from 28-May-2019)
https://www.ijcaonline.org/archives/volume170/number9/naik-2017-ijca-914937.pdf

1% match (Internet from 02-May-2015)
http://research.ijcaonline.org/volume98/number22/pxc3897433.pdf

1% match (student papers from 27-May-2019)
Submitted to Ateneo de Davao University on 2019-05-27

< 1% match (Internet from 12-Jul-2017)
http://docplayer.net/10272060-More-data-mining-with-weka.html

< 1% match (student papers from 09-May-2016)
Submitted to CSU, San Jose State University on 2016-05-09

< 1% match (Internet from 26-Oct-2019)
https://www.futurelearn.com/courses/more-data-mining-with-weka/0/steps/29142

< 1% match (Internet from 30-Jun-2015)
http://chethoughts.com/?p=1190

< 1% match (student papers from 18-May-2018)
Submitted to University of Bradford on 2018-05-18

< 1% match (Internet from 06-Nov-2018)
http://organicexpressindia.com/blog/view/47

< 1% match (student papers from 23-Mar-2019)
Submitted to National Institute of Technology, Silchar on 2019-03-23

< 1% match (Internet from 04-Jul-2014)
http://euacademic.org/UploadArticle/98.pdf

< 1% match (Internet from 25-Sep-2019)
http://technobium.com/decision-trees-explained-using-weka/

< 1% match (Internet from 05-May-2019)
https://generalfraud.blogspot.com/2015/06/calcium-carbide-box-of-artificially.html

< 1% match (publications)
"Improved Diabetic Data Analytic Model for Complication Prediction", International Journal of Engineering and Advanced Technology, 2019

< 1% match (Internet from 26-Nov-2018)
https://www.ftapcci.com/source/ftapcci/FR%202017/FR%202017%2007%2026.pdf

< 1% match (Internet from 07-May-2016)
http://www.dailypioneer.com/city/pil-seeks-enforcement-of-ban-on-calcium-carbide-use-to-ripen-fruits.html

< 1% match (student papers from 01-Apr-2015)
Submitted to Northeastern Illinois University on 2015-04-01

< 1% match (student papers from 09-Aug-2016)
Submitted to National University of Singapore on 2016-08-09

< 1% match (Internet from 30-Oct-2019)
https://medium.com/@chiragsehra42/decision-trees-explained-easily-28f23241248