IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Detection of Adulteration inFruits And Vegetables UsingMachine Learning

Chetan Raju P H[1], D Rakshitha[2], K B Likitha[3], Modala Chetan[4] ,
Mrs. Jayashubha J[5]

[*1,2,3,4]Student, Department of CSE, K S School Of EngineeringManagement(KSSEM), Bengaluru,India.

[*5]Assistant Professor, Department of CSE, K S School Of Engineering Management(KSSEM),Bengaluru, India.

Abstract - Fruits are essential for healthy life. The fruits we take should be pure, nutritious and free from any type of adulteration for proper maintenance of human health. Formalin is used as a preservative by the traders to improve the appearance of fruits and vegetables and to sustain for longer periods. Formalin is a colorless, aqueous solution of formaldehyde to preserve biological specimens. Not every case of adulteration will result in serious adverse health effects. But the chemical is highly toxic and a 30 ml of formalin containing 37 percent of formaldehyde can kill an adult. Many of the methods for detection of food adulteration require elaborate steps of sample preparation and prior analysis involving high-end technologies which makes the whole process difficult to perform and time consuming. Visual Geometry Group 16 (CNN) architecture has been incorporated in our system to accurately predict the correct concentration of formalin. The main aim of this system is to replace the manual inspection system. This helps to speed up the process, improve accuracy and efficiency. This system captures the image of the fruit and then image processing is done to get required features of fruits such as color and size. Adulterated fruit is detected based on image pixels. Sorting is done based on color and size.

Introduction

Fruits are essential for healthy life. The fruits we take should be pure, nutritious and free from any type of adulteration for proper maintenance of human health. The intake of any fruit substance is intended for the nourishment which is gained from it.

Since the fruit is into consecutive stages of production, processing and finally distribution, the nourishment in the fruit items is collapsed. For the fruit products to ripen and improved in texture, storage and appearance, a concept of adulteration is widely practiced. The nature or quality of the fruit is reduced through addition of adulterants bythe process of fruit adulteration. The adulterants is a foreign chemical substance present in fruit. In the process of fruit adulteration, little quantities of non-nutritious substances are added knowingly to improve its appearance or storage properties of the fruit. Mostly the adulteration in fruits and vegetables are caused using a harmful chemical substance called Formalin. Formalin is a colorless, aqueous solution of formaldehyde to preserve biological specimens.
India is at second number after China in the production of the fruits. In India all the preharvest and post-harvest process are done manually with help of labor. Manual processis very time consuming, less efficient so to get accurate result automation in agriculture industry is needed. The post-harvest process includes sorting and grading of fruits. Different quality factors are considered for sorting and grading of fruits. These factors are internal quality factors and external quality factors. The external quality factors are texture, shape, color, size and volume and internal quality factors are test, sweetness, flavor, aroma,nutrients, carbohydrates present in that fruit.

Related Works

Formaldehyde is naturally occurring metabolic by-product and it is used widely to improve the shelf-life of fruits and vegetables. The naturally produced formaldehyde content in fruits and vegetables are found using a spectrophotometric technique. The outcome provides the baseline data of the formaldehyde present in the food items. Based on the food types, nature of storage and its temperature, pattern of ageing, the behavior of formaldehyde varies. This experimental result could be useful for preventing food contamination and presser- vation.

- The level of formalin in a fruit or vegetable which exceeds beyond the naturally occurring formaldehyde level is detected using several approaches of machine learning. The HCHO or CH2 sensor that is used to measure formaldehyde content is used along with Arduino to find the value of formalin present in fruits orvegetables.
- A detector was developed that senses formaldehyde existence as parts per million with respect to the presence in air. It infers whether the concentration in air is permissible or harmless.

- For three trials an accuracy of 98.33 was achieved.

Overview of the System

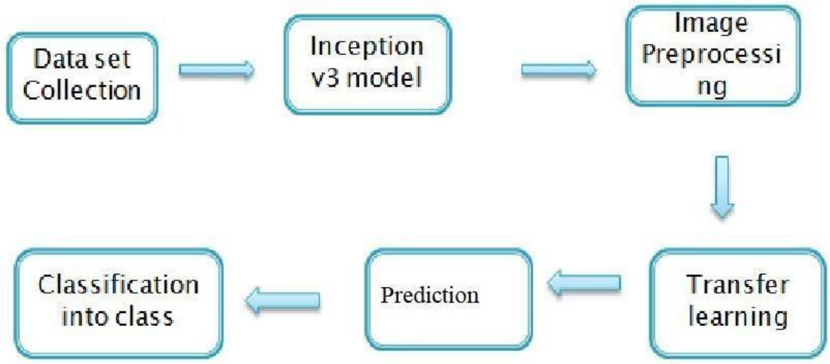

Fig-1- Proposed model for fruit recognition and grade of disease detection

TABLE I NATURALLY OCCURRING FORMALDEHYDE VALUES OF SEVERAL FRUITS.

Fruit name	Naturally occurring formaldehyde
Apple	6.3–22.3
Apricot	9.5
Banana	16.3
Cucumber	2.3–3.7
Grape	22.4
Pear	38.7–60
Plum	11.2
Tomato	5.7–13.3
Water-melon	9.2

A. Naturally Occurring Formaldehyde

Formaldehyde is naturally produced in fruits and vegetables,meats, fish, etc as a by-product of metabolism. In biological specimens, Formaldehyde is produced from methylated compounds and conversion of glycine and serine. Based on food types and conditions, the formaldehyde content produced Hyde could be found by figure out the naturally generated formaldehyde content in food. The below Table I lists the naturally occurring formaldehyde content is various fruits and vegetables.

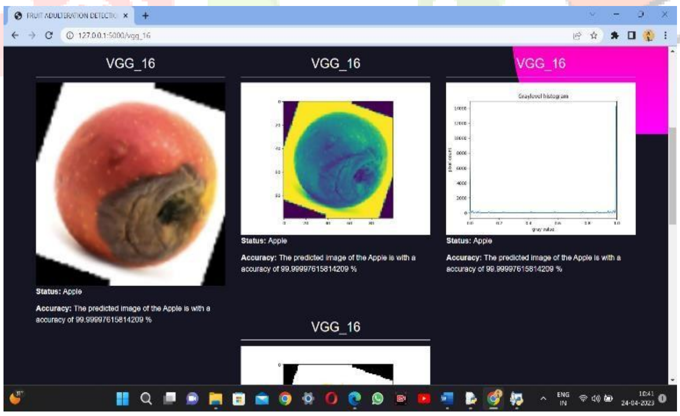

Fig-2- Adulteration detection results for unhealthy Apple

Design and Experiments

A. Dataset

A Dataset is mostly defined as the collection of data. It will be also worked out as tabular data that corresponds to 1 or more tables. Every column represents a variable and each row represents a selected record within the data set. These datasets play a significant role in Machine Learning while training a model. The more you train the model with good number of dataset that reflects within the accuracy of your model. Detecting the artificially added formalin, this naturally presentformalin in the food items may generate wrong results. To avoid this problem, the dataset is referred with the data that of naturally occurring formalin in fruits that's collected from the Centre for Food Safety (CFS). Relating to we created our own data set which consists of three columns fruit and Vegetables table, ppm(parts per million) and State in total of hundred data. There are two types of fruits in the dataset: 'apple' and 'orange'. Then we used this pre-processed dataset to train the predictive model to detect given sample fruits whether fresh or adulterated. The dataset was divided intothe training dataset and test dataset 75% for training the model and 25% of the data for testing it.

Fig-3.1- Fruit image selection by data set

B. Formalin Detection

Fruits are naturally good source for conductor of electricity due to the presence of juice in them which forms mild acid. Acids are able to conduct electricity rather like primary cell where conductivity depends on the number of chemical and water in them. More watery and chemical compound will allowmore electrical energy to meet up with. Acidic fruits are good conductor of electricity. Grove HCHO which is a semiconductor VOC sensor used for detecting the formable- Hyde combining with Raspberry-Pi. This sensor featured to detect gas concentration up to 1 ppm (parts per million). As formaldehyde is self- vaporized solution, its presence can be detected by volatile organic compound sensor. The output voltage of the sensor is exponentially proportional with the concentration of formalin present in the fruit sample.

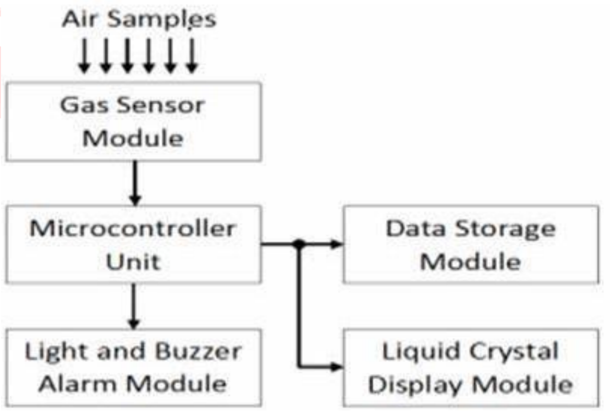

Fig-3.2- Design and development diagram of formalin detector

C **Color Detection**

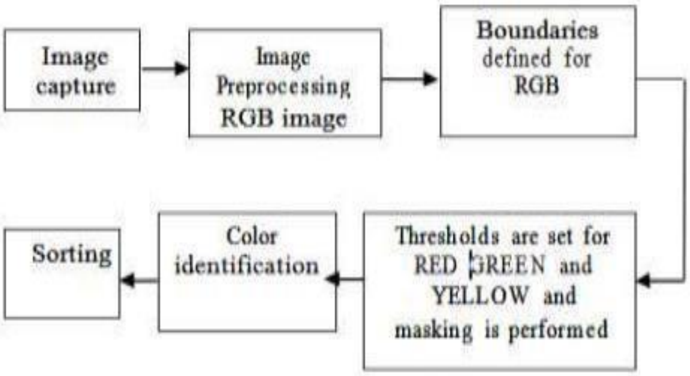

Fig3.3- Flow for Color Detection

The fig. 3.2 shows the flow of the working model of fruit grading system. Initially for the collection of data images are captured. Since the images captured are irregular, so resizing has to be done to meet the standard image size in CNN. For the pre-processing of data library, called Open CV is used. Then they are split into training and testing data. The images are fed into the CNN as input, in order to train the data. Once the images are trained, the images split for testing are fed to test the created model. When a new sample is placed on the conveyor belt, image of the sample is captured and preprocessed. Later it is compared with the trained data set by using CNN algorithm, based on the prediction fruit is classified. An open source library called keras that runson TensorFlow which is backend engine by default.
Along with that, Inutile is employed to form basic image processing functions easier with Open CV and Python.

D. Model Development

Followed by the formalin detection different voltage drops are measured for different fruit samples as each fruit contains different range of resistance. Firstly by using the extracted features From the data set we implement the rule based classification model, which first let us categorize the fruit type Later we have performed various algorithms and where results were measured. On completely considering the naturally occurring formalin in fruits we have developed this model andtrained using the data set which also depicts the naturally added ppm(parts per million) value along with the additionally added formalin. This system generates the output by making predictions whether the particular fruit item is 'dangerous 'or 'safe'. depicts the process flow.

DISEASE	Mean	Mode	S.D
Phomopsis	35	35	3.5
Blight	37	36	4.32
Alternaria Blight	25	26	2.40
Soft rot of potato	-	-	-
White Rust of Amaranthus	14	14	2.45
Anthracnose of Chilli	14	14	1.15
Anthracnose of bottle Guard	11	11	0.94
Step rot of mango	-	-	-
Step rot of guava	26	26	1.76
Step rot of Litchi	19	19	2.00
Rust of Bean	24	24	3.0
Rhizopus rot of jack fruit	-	-	-
Alternaria blight of cauliflower	21	21	1.76
Anthrancose of Amaranthus	15	15	3.05
Scab of lemon	11	11	0.94
Sigatoka of banana	24	24	1.15
Panama of banana	19	19	1.33

Fig. 3.4: Highlighting diseases associated with fruits

RESULT AND DISCUSSION

A. Performance of Various Algorithms

The features will often contain information relative to gray scale, texture, shape or context in pattern recognition. It is a vital domain of computer science in concern with recognizing patterns specifically in visual patterns and sound patterns. Several algorithms are used to discover regularities. These regularities are beneficial to perform classification of data.

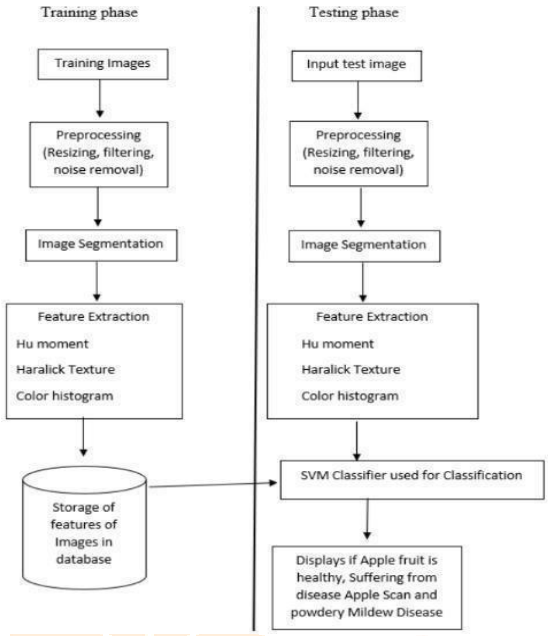

Fig. 4.1: Flow diagram for classification disease in apple fruit

The initial pattern measurement or some series of pattern measurement is converted to a new pattern feature in image processing or machine vision. Higher-level information of the object is obtained using pattern classification. Thus the information or features extracted from the metadata are assigned to the object belonging to a category or class. The development ofthe classification algorithm results in identification of the objectsin the image.

Logistic regression is a classification algorithm where in based on a set of independent variables, a binary outcome is predicted. The binary outcome is the probability of happening of an event or probability of not happening of an event. As depicted in Table II-An accuracy of 70% is obtained in the trainset of fruits named train set y and 40% is obtained onthe test set of fruits named test set y.

The support vector machine concentrates on the compli- cated points in the classification unlike other classification algorithms which defines all points. It works best in finding the best separating or uncommon line in the classification. An accuracy of 61% is obtained in the train set of fruitsnamed train set y and 33% is obtained on the test set offruits named test set y. Another algorithm which is widely

TABLE II

DEPICTION OF ALGORITHM AND THE CORRESPONDINGACCURACYIN TRAIN SET AND TEST SET.

Algorithm	Accuracy On train set y	Accuracy On test set y
Logistic Regression	70%	40%
Support Vector Machine	61%	33%
K-NN	87%	95%

used for classification is the k-Nearest Neighbors. The object x,which is unknown in the queryimage is compared with each andevery sample of similar or alternative objects that were previously in use to train the classification algorithm while the process of classification. An accuracy of 87% is obtained in the train set of fruits named train set y and 95% is obtained on the test set of fruits named test set y.

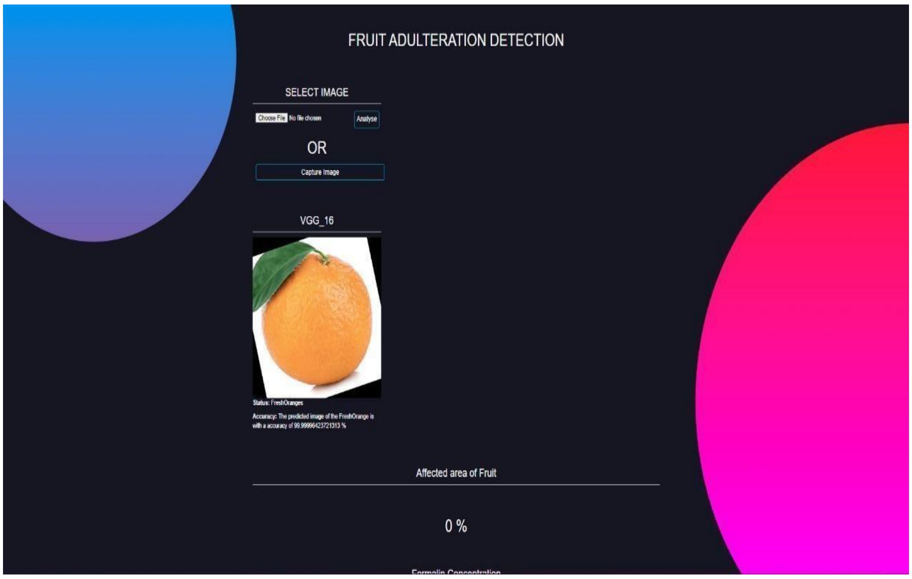

Fig 4.2: Fruit image selection UI

Fig. 4.2 shows the fruit adulteration detection page where user can select the imageof the fruit andcapture the real time image and upload it into the system

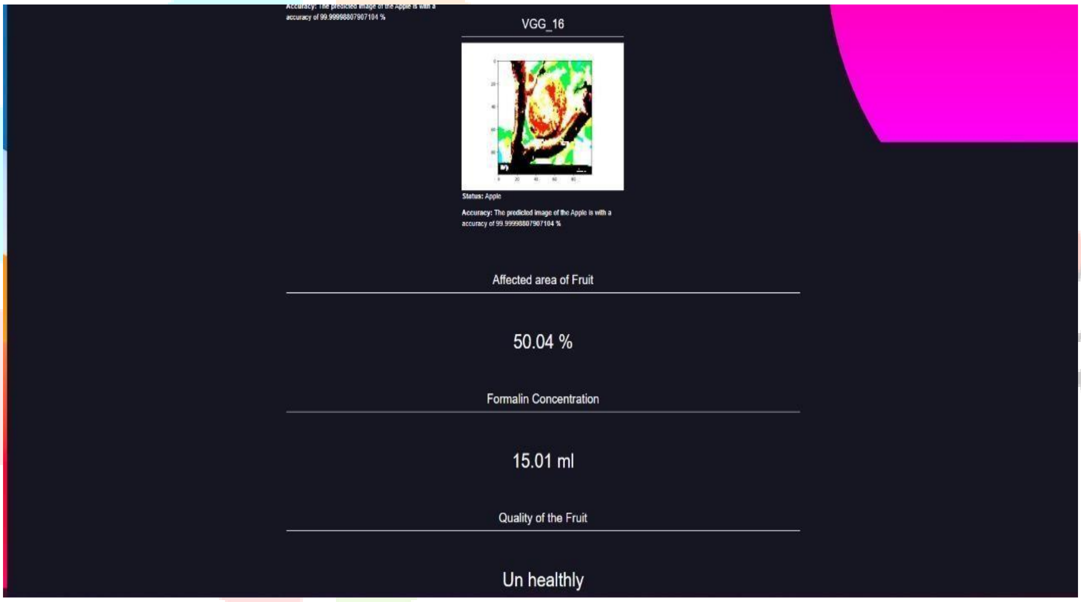

Fig 4.3: Adulteration detection results for unhealthy apple

Fig 4.3 shows the prediction results for the infected apple. It depicts the affectedarea of theapple, formalin concentration and the quality of the apple.

Conclusion

Fruit adulteration system can detect the concentration of formalin content in the fruit. The manual system fails to detect the formaldehyde level accurately and the use ofgas sensors in formalin detection is costly. This project aims to find how much percent the fruit is adulterated with the formalin and depicts the fruit is consumable or not. This model will be useful for the consumers to safe check the quality of the fruit. To get better resultsin the classification and identification of adulterated fruit, CNN architecture is used.

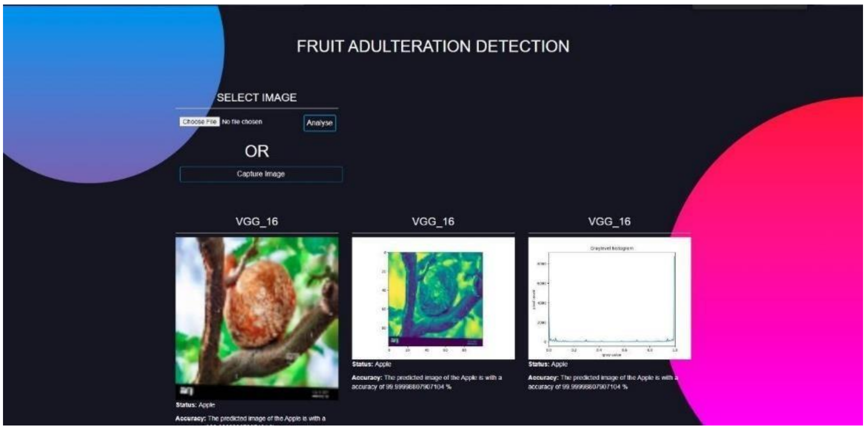

Fig 5.1: Adulteration detection stages

REFERENCES

[1] Md. Ashraf Kamal, 2 Nafiza Nusrat, 3 Nowrin Siddique Dola, "Design and Developmentof A Formalin Detection System Using IoT", International Journal of Scientific & Engineering Research Volume 10, Issue 9, September-2019 344 ISSN 2229-

5518.

[2] Shaikh Rakhshinda Nahid M.Ayyub ,Aarti Manjramkar "Fruit Disease Classification and Identification using Image Processing", Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4.

[3] Nikhitha M, Roopa Sri S, Uma Maheswari B "Fruit Recognition and Grade of Disease Detection using Inception V3 Model", Proceedings of the Third International Journal on Electronics Communication and Aerospace Technology [IEEE 2019].

[4] Anjana Purushothaman and Maruveetil, "Food Adulteration Detection Using Machine Learning" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 5, 2019.

[5] Kawaljit Kaur and Chetan Marwaha "Analaysis of diseases in fruits using image proccessing techniques" 978-1-5386-0566-0/17/$31.00 2019 IEEE.

[6] P. Kanjana Devi, Dr. Rathamani "Image Segmentation K-Means Clustering Algorithm for Fruit Disease Detection Image Processing", Fourth International Journal on Electronics, Communication and Aerospace Technology (IEEE-2020).

[7] R. Ramya, Dr. P. Kumar, K. Sivanandam, M. Babykala, " Detection and classification of fruit diseases using image processing & cloud computing", IEEE, 2020.

[8] Miss. Supriya V. Patil, Miss. Vaishnavi M. Jadhav, Miss. Komal K. Dalvi, Mr.B.P.Kulkarni, "Fruit Quality Detecction using OpenCV/Python", International Research Journal of Engineering and Technology (IRJET), May 2020.

[9] Lisha Kamala.K, Dr. Sini Anna Alex , "Development of Machine Learning based Fruit Detection and Grading system" 2021 2nd International Journal on Smart Electronics and Communication, 2021 IEEE.

[10] S. Prince Sahaya Brighty, G. Shri Harini, and N. Vishal "Detection of Adulteration in Fruits Using Machine Learning", 2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) | 978-1-6654-4086- 8/21/$31.00, 2021 IEEE