Detection of Adulteration in Fruits Using Machine Learning

Mrs Babitha Poojary [1], Chinmayee S[2], Girish M K[3], Jayanth G[4]

[1] *Alva's Institute of Engineering and Technology, babitha@aiet.org.in*

[2-4] *Students, Department of Computer Science and Engineering, Alva's Institute of Engineering and Technology*

Abstract— **Summary - food is necessary for life. The food we use must be clean, nutritious and** without additives **to maintain human health. This paper develops an IoT-based** detection of **food and formalin technology identify formalin using machine learning methods. A volatile compound HCHO gas sensor connected to a Raspberry pi3 was used to extract the formalin content of any fruit or vegetable as a function of output voltage, and various machine learning algorithms were used to classify the fruit or vegetable based on the extract. characteristics. Our system includes supervised machine learning algorithms which accurately predict the correct concentration of formalin at all temperatures, which can also be correctly classified between artificially added and naturally formed form.**

Index Terms - Adulteration, Formalin, Naturally Occurring Formalin, Formaldehyde, Machine Learning, Internet of Things, HCHO Gas Sensor, Raspberry Pi3, Supervised Machine Learning Algorithms. *Index Terms—* **Component, formatting, style, styling, insert.**

I. INTRODUCTION

The consumption of any food is for the nutrition it provides. As food goes through the successive during production, processing and finally marketing, the nutrients contained in the food are broken down. The concept of adulteration is widely used to improve the composition, shelf life and appearance of food. The nature or quality of food deteriorates when degrading substances are added or any essential substance is removed during food adulteration. Degradants can foreign or non-standard chemical substances contained in the food that damage the food. Food adulteration involves the intentional addition of small amounts of non-nutritive substances to improve the appearance or shelf life of the food.

Most fruits and vegetables are adulterated with a harmful chemical called formalin. Formalin is a colorless aqueous solution of formaldehyde for preservation of biological specimens. This chemical is used to prevent corpses from decomposing. This function is used for decomposition. Not all cases of adultery lead to serious health problems. However, the chemical is very toxic and 30 ml of formalin containing 37 percent formaldehyde can kill an adult [7]. Merchants use formalin as a preservative for treatment fruity appearance and vegetables and keep them longer. SEM secretary Abdus

Sobhan found in a survey where 115 samples of mangoes and other fruits were collected from more than 50 organic shops. stores and allows formalin-treated chemical-free fruit. Consuming any fruit substance is for the nutrition, nutrition and minerals of the fruit. Because the crop object was collapsed by prefab, adjustment and closing areas. The idea of adulteration is used to keep fruits healthy for their appearance. Besides, he's used to it Keep fruit or vegetables nice and storable. Fruit adulteration can reduce the fruit by adding crumbles or using ingredient removal. The chemical in the fruit may be foreign or worse. Non-nutritive elements can be added in small doses to improve the shelf life and brightness of the fruit.

All components of fruits are consumed for sustenance. Nutritional value of fruits fell because the grain went through several stages of production, processing and distribution. Adulteration is a term often used to improve the texture, shelf life and appearance of fruit products. Fruit adulteration is the alteration of the character or quality of a fruit by adding or removing essential parts. One of them could be a foreign or inferior chemical in the fruit an adulterer During adulteration, small amounts of non-nutritive chemicals are intentionally added to fruit to improve the appearance or shelf life of the fruit.

India is a country based on agriculture. India produces a wide variety of fruits and vegetables. In terms of fruits, India is second only to China. All pre-harvest and post-harvest operations are done manually in India. The agricultural sector requires manual automation, which takes a long time and is inefficient. The fruits are sorted and graded in the post-harvest process. Fruits are sorted and graded according to several quality parameters. Both internal and external quality criteria are applied to these variables. Internal quality criteria are taste, sweetness, flavor, aroma, foods from fruits and carbohydrates. Quality indicators of the external surface are structure, shape, color, size and volume. Modified algorithms such as YOLOv3 with better accuracy than YOLOv2 and VGG16 method under Convolutional neural network algorithm were added to correctly predict formalin concentration at all temperatures in our system.

A chemical called formalin is harmful to fake fruits and vegetables. Formalin is a colorless aqueous solution of formaldehyde for preservation of organic specimens. This

chemical is used to prevent our body from breaking down unnecessarily. All cases of adultery no longer lead to serious, dangerous health consequences. But the chemical is clearly toxic, and 35 ml of formalin, which contains 0 percent formaldehyde, can kill a life. Formalin can be used as a preservative to keep fruits and vegetables clean for a long time. India is an agricultural country, special fruits and vegetables are produced in India. India is second only to China in fruit production. In India, the entire pre-harvest and post-harvest approach is done manually, with labor. The manual method is very time-consuming, much less efficient, so the cultivation must be automated to get the right result. The post-harvest approach involves sorting and grading the fruit. several key factors are considered when sorting and grading fruit.

II. RELATED WORKS

Formaldehyde is a naturally occurring metabolic byproduct and is widely used to improve the shelf life of fruits and vegetables. Naturally produced formaldehyde content in fruits and vegetables is determined by spectrophotometric technology. The result provides basic information about formaldehyde in food. The behavior of formaldehyde varies depending on the type of food, the nature and temperature of storage and the aging method. This experimental result can be useful for food prevention and preservation [1]. Levels of formalin in a fruit or vegetable that exceed the concentration of naturally occurring formaldehyde are detected using several machine learning methods. The HCHO or CH2 sensor used to measure formaldehyde content is used with Arduino to determine the formalin content of fruits or vegetables [2]. A detector has been developed that detects the presence of formaldehyde in parts per million relative to air. This determines whether the air concentration is acceptable or harmless. Three tests achieved an accuracy of 98.33 [3].

Fruits can be identified based on objects found in Tree by consulting an expert or using an already successfully identified specimen (herbarium). Similar images from a monograph or plant book allow identification. Presentation of summaries is used as an identification key. Fractal and Euclidean distance methods are used in this work develop the results of previous studies. Fractal Dimension and Fractal Code Detection Systems, with an accuracy of 68% and 51%. The main goal of this study is to increase accuracy. The proposed technique combines the median filter with the texture analysis method. When using the filter image enhancement method, texture analysis is used as a feature extraction tool. Based on the tests, the detection of the intermediate filter and texture analysis increased to 78%. The application of the identification system, which is tested in the medicinal fruit information systems of the network.

III. OVERVIEW OF THE SYSTEM
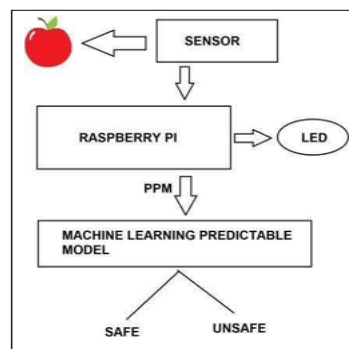
A. Block Diagram

Fig. 1. Depiction of the formalin detection in fruits.

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

B. Naturally Occurring Formaldehyde

Formaldehyde is found naturally as a byproduct of metabolism in fruits and vegetables, meat, fish, etc. In biological samples, formaldehyde is produced from methylated compounds and from the conversion of glycine and serine. The amount of formaldehyde produced varies depending on the type of food and conditions in different foods. The dose of external formaldehyde was discovered by finding out the naturally formed formaldehyde content of the food. Table I below lists the naturally occurring concentrations of formaldehyde in various fruits and vegetables.

IV. LITERATURE SURVEY

S. Prince Sahaya Brighty1 et. al., [1] IoT-based food and formalin detection technology has been developed to detect formalin. A volatile formalin sensor attached to raspberry pi3 was used to extract the formalin content of any fruit as a function of production voltage.

M. Nikhitha et. al., [2] The proposed system classifies the fruits based on the infection rate. The system is developed for the tensor flow platform. Some features can be added to the model to make the combination of image processing and deep neural network useful for fruit disease detection.

P. Kanjana Devi et. al., [3] Clustering and fruit image segmentation algorithms were used to detect fruit diseases. The datasets used here are still images, and k means clustering cannot handle noisy data and outliers, resulting in less accurate results.

Kawaljit Kaur et. al., [4] This paper presents a machine vision-based system suitable for picking mangoes according to their post-harvest use-by date. Image processing is very expensive, depending on the system used and the number of detectors required.

K Lisha Kamala et. al., [5] Various image processing techniques are involved, such as image acquisition, image pre-processing, feature extraction, image segmentation and classification. Only works with some hydroponic fruits. SVM does not perform very well when there is more noise in the dataset, ie. overlapping target classes.

Md.Ashraf Kamal et. al., [6] This system has developed an automated IoT-based device that detects formalin applied to food. The device must be connected to the Internet and placed in a closed area to give an accurate result.

Shaikh Rakhshinda Nahid M.Ayyub et. al., [7] Image processing using different combinations of color, texture, and shape features is proposed for the detection and classification of apple fruit diseases. SVM provides less accuracy than CNN.

Jijesh J.J et. al., [8] The proposed system captures fruit placed on a conveyor belt and then compares the captured image with a CNN dataset that extracts fruit characteristics such as texture, color and size. Even if the quality of the grain is good, the accuracy will change. is low because it considers only appearance.

Miss. Supriya V. Patil et. al., [9] This paper presents automatic fruit quality detection system to sort and classify fruit and damaged fruit. Although the fruit quality is good, the accuracy is shown to be low because it only considers the appearance.

R.Ramyal et. al., [10] This article mainly deals with the detection and analysis of the infection of fruits available in planting areas, and the storage of agricultural records and the data of farmers into a database, and data extraction by cloud computing. The SVM algorithm is not suitable for large datasets.

V. DESIGN AND EXPERIMENTS

A. Dataset

A dataset is usually defined as a collection of data. It is also produced as tabular data corresponding to one or more tables. Each column represents a variable and each row represents a record selected from the dataset. These datasets play an important role in machine learning when training a model. The more you train a model with a large amount of data, it will reflect the accuracy of your model. Detection of artificially added formalin, the formalin that occurs naturally in food, can give false results. Avoid this problem, the material refers to information collected by the Food Safety Center about formalin naturally occurring in fruits. In this regard, we created our dataset, consisting of three columns fruit_label, ppm (parts per million) and State, into hundreds of datasets. There are two types of fruit in the dataset: "apple" and "orange". We then used this pre-processed data to train a predictive model to identify specific fresh or adulterated fruit. The data (Figure 2) was divided into training data and test data with −75% for

model training and 25% for test data. They are abstractly defined. Abbreviations such as IEEE and SI do not need to be defined. Do not use abbreviations in the title or headings unless necessary.

B. Formalin Detection

Fruits are a naturally good source of electrical conductors because they contain juice that forms a mild acid. Acids are able to conduct electricity instead of a primary element, whose conductivity depends on the amount of chemicals and water they contain [6]. More water and chemical compounds make it possible to obtain more electrical energy [7]. Acidic fruits are good conductors of electricity. Grove HCHO what is

	A	B	C	D	E
1	fruit_labe	ppm	State		
2	1	23.42	0		
3	1	16.39	1		
4	1	15.77	1		
5	3	52.3	1		
6	3	63.3	0		
7	3	56.7	1		
8	3	58.7	1		
9	3	61.13	1		
10	1	22	0		
11	3	64.2	0		
12	1	18.15	1		
13	1	16.09	1		
14	1	20.3	0		
15	3	63.04	0		
16	3	61.47	1		
17	3	68	0		
18	1	18.83	1		
19	3	63.37	0		
20	1	20.66	0		
21	1	20.36	0		
22	1	21.08	0		
23	1	23.46	0		
24	1	15.53	1		
25	3	63.39	0		

fruit_data   ⊕

Fig. 2. Dataset of fruits that are mapped to different ppm levels.

a a semiconductor VOC sensor used to detect formaldehyde attached to a Raspberry-Pi. This sensor detects gas concentrations down to 1 ppm (parts per million). Since formaldehyde is a self-evaporating solution, its presence can be detected with a VOC sensor. The output voltage of the sensor is exponentially proportional [8] to the formalin content of the fruit sample.

C. Model Development

After formalin detection, different voltage drops are measured for different fruit samples because each tree contains a different resistance range. First, using the features extracted from the dataset, we apply a rule-based classification model [a set of IF-THEN rules], which allows first to classify the type of fruit [9]. Then we made different algorithms and where the results were measured. Taking full account of naturally occurring formalin in fruit, we developed and trained this model using a dataset that also describes naturally occurring ppm (parts per million) and added formalin. This system produces a result that predicts whether a particular fruit is "dangerous" or "safe". describes the flow of the process.

VI. RESULT AND DISCUSSION

A. *Performance of Various Algorithms*

Features often include grayscale, texture, shape, or context information in pattern recognition. It is an important branch of computer science that deals with pattern recognition, especially visual and auditory patterns. Several algorithms are used to search for patterns. These regularities are useful for classifying data. The initial pattern measurement or series of pattern measurements are transformed into a new pattern feature in image processing or machine vision. Automation of more advanced information is an important feature of the daily life of industry. In India, more than 1/2 of the population depends on agriculture. Their important source of income is agriculture.
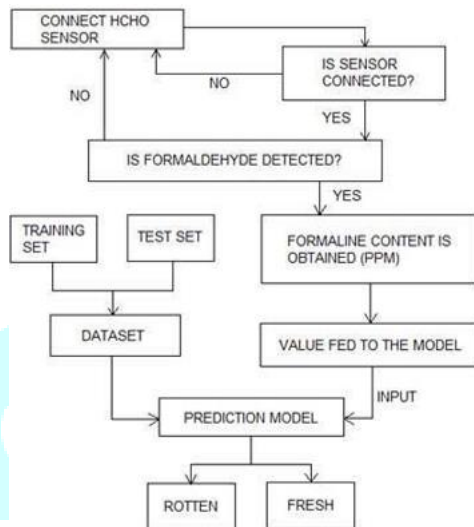

Fig. 3. Flowchart of the proposed system.

object is obtained by pattern classification. Thus, information or properties extracted from metadata are assigned to a class or an object belonging to a class. The development of a classification algorithm[10] leads to the detection of objects in the image. Logistic regression is a classification algorithm which a binary outcome is predicted based on a set of independent variables. A binary outcome is the probability of an event or the probability that the event will not occur. As shown Table II, 70% accuracy is achieved for a set of fruits called set y and0% accuracy for a set of fruits called test y. Support Vector Machine focuses on complex classification points unlike other classification algorithms that define all points. This is best for finding the most distinctive or unusual row in the classification. The accuracy is 61% of the set of fruits called The training set y and 33% are obtained from the test set of fruits called test set y. Another widely used classification algorithm uses k-nearest neighbors. In the query image, the unknown object x is compared to each sample of a similar or alternative object previously used to train the classification algorithm during the classification process. 87% accuracy is achieved with a set of fruits called Trainset y and 95% accuracy with a test set of fruits called testset y.

VII. PROPOSED METHOD

The main purpose of this laptop is to replace the driver's control system. This helps to speed up the efficiency of the system decoration and the accuracy of time limit. This machine collects graphics from a digital camera placed on a conveyor belt. Image processing is then performed to obtain the required characteristics of the fruit, such as color and size. Fruit ripening is detected based on photo pixels. Sorting is done entirely by color and size. This proposed device is a dynamic and reliable method for the identification of fruits and formalin based entirely on the knowledge of bench-based approaches. Faking a disease using arduino is one of the easiest strategies to stop using the durability of the fruit. Identifying crude formalin without a predetermined mimic as the end result of formalin in its natural form can also be misleading.

The main purpose of this system is to replace the manual inspection system. This helps reduce the length of the process, increasing accuracy and efficiency. The data set that this system uses as input comes from a local record. Image processing is then performed to obtain the required fruit characteristics such as color and size. YOLOv3 detects fruit fakes based on image pixels. The algorithm is more efficient than YOLOv2 and combined with CNN (Convolution Neural Network). method VGG16. This model achieves an accuracy of 92.7 percent in five tests on the ImageNet dataset. consists of 14 million images belonging to 1000 categories. Sorting is done by color. And educated scales use an identification method to detect the color of the chemical components formed in the fruit with different white fruit layers.

VII. CONCLUSION

Thus, the paper describes formalin detection based on machine learning approaches. A manual system cannot accurately detect formaldehyde levels. This proposed system is a dynamic and reliable food and formalin detection technique based on machine learning methods. Detecting food contamination using arduino is one of the simple methods that gives the result by measuring the resistance of the food. Identifying crude formalin without a predefined pattern of naturally occurring formalin result can be misleading. Therefore, this system monitors artificially added formalin as stored binary "1 machine learning algorithm ie. logistic regression, support vector machine, K-NN classifier is applied to experimental data set to build prediction model. Basic functions were used, the designed system can detect 1-50 ppm (ppm) formalin VOC HCHO gas sensor connected to Raspberry Pi, this shape detection machine learning method can detect formalin content in any food and consumption based on detected safety status food. Instead of checking the fruit guide, a laptop is used with an imaginative and predictive calculator that provides a genuine, fair and non-destructive estimate. An automatic creative and predictive main based machine was mentioned to sort and classify the fruits completely according to their color and accordingly. When inspecting bananas for defects, defective fruit is identified. And for the three unique properties, good, medium and low. conveyor speed and light variation, digital camera decision affects the system.

REFERENCES

[1] M. Samad Khan, F. Nowshad, and M. d. N. Islam, "Concentration and formation behavior of naturally occurring formaldehyde in foods," *Nowshad et al. Agric & Food Secur*, 2018.

[2] A. W. Reza, S. D. Barman, K. Tabassum, A. A. Memi, and N. Sultana, "Food and formalin detector using machine learning," *International Journal of Machine Learning and Computing*, 2019.

[3] F. C. Bato, T. J. G. Bautista, and C. D. Cortez, "Development of formaldehyde detector," *International Journal of Information and Elec- tronics Engineering*, 2015.

[4] Formaldehyde in Food [2018]. [Online] Available: https: //www.cfs.gov.hk/english/programme/programme_rafs/programme_ rafs_fa_02_09.html.

[5] Fruit data with colors. [Online] Available: https://github.com/susanli2016/Machine-Learning with Python/blob/master/fruit_data_with_colors.txt.

[6] S. Mahya, M. Adlim, Z. Fitri, Y. Amri, and M. Sari, "Preparations of chemical sensors for simple formalin detection in contaminated food", *Annual International Conference*, 2011. A. Wilson, "Diverse applications of electronic-nose technologies in agriculture and forestry," *Multidisciplinary Digital Publishing Institute (MDPI)*, 13.

[7] A. Naseer, A. Dar, J. Amara, U. Shafique, and J. Anwar, "A simple spot test quantification method to determine formaldehyde in aqueous samples," *Journal of Saudi Chemical Society*, 2016.

[8] S. Dea, E. A. Baldwin, J. Bai, and A. Plotto, "Electronic noses and tongues: Applications for the food and pharmaceutical industries," *Sensors*, 2011.

[9] W. C. Seng and S. H. Mirisaee, "A new method for fruits recognition system," in *Proc. International Conference on Electrical Engineering and Informatics*, 2009.

[10] M. Siraj, W. B. Habib, and S. S. Saha, "Food alytics: A formalin detection system incorporating a supervised learning approach," *Human- itarian Technology Conference (R10-HTC)*, 2018.

[11] Sari, Yuita Arum, R V HariGinardi, RiyanartoSarno. "Assessment of Color Levels in Fruit Color Chart Using Smartphone Camera with Relative Calibration". Information Systems International Conference (ISICO), 2013: 631-636

[12] Raid, Richard Neil, and J. C. Comstock. Sugarcane adulterated. University of Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, EDIS,

[13] Camargo, A., and J. S. Smith. "Image pattern classification for the identification of adulterated causing agents in fruits." Computers and Electronics in Agriculture 66.2 (2019): 121-125.

[14] Asraf, H. Muhammad, M. T. Nooritawati, and M. S. B. Rizam. "A Comparative Study in Kernel-Based Support Vector Machine of Oil Palm Leaves Nutrient Adulterated." Procedia Engineering 41 (2020): 13531359.

[15] Phadikar, Santanu, Jaya Sil, and Asit Kumar Das. "Rice adulterateds classification using feature selection and rule generation techniques." Computers and Electronics in Agriculture 90 (2021): 76-85.

[16] Li, Daoliang, Wenzhu Yang, and Sile Wang. "Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine." Computers and electronics in agriculture 74.2 (2018): 274279.

[17] Ginardi, R. V Hari, RiyanartoSarno, and Tri AdhiWijaya. "Sugarcane Fruit Color Classification in Sa*b* Color Element Composition". 2019 International Conference on Computer, Control, Informatics and It's Application, pp:175-178

[18] Li-jie, Yu, Li De-sheng, and Zhou Guan-ling. "Automatic Image Segmentation Base on Human Color Perceptions." International Journal of Image, Graphics and Signal Processing (IJIGSP) 1.1 (2019): 25.

[19] Shivakumar, G, P.A Vijaya. "Face Recognition Using Geometric Attributes". International Journal of Computational Intelligence Research Volume 6, Number 3 (2020), pp. 373–383.