DETECTION OF ADULTERATION IN FRUITS USING MACHINE LEARNING

G.Shri Harini [1],P.Swathi [2],N.Vishal [3],S.Prince Sahaya Brighty [4]

[1] UG Student, Sri Ramakrishna Engineering College, Coimbatore 641022, India

[2] UG Student, Sri Ramakrishna Engineering College, Coimbatore 641022, India

[3] UG Student, Sri Ramakrishna Engineering College, Coimbatore 641022, India

[4]Assistant Professor, Sri Ramakrishna Engineering College, Coimbatore 641022, India.

Email :[1] shriharini.1601225@srec.ac.in ,[2] swathisruthip@gmail.com , [3]vishal.1601253@srec.ac.in ,[4] brighty.s@srec.ac.in

ABSTRACT

Food is essential for life. The food we take should be pure, nutritious and free from any type of adulteration for proper maintenance of human health. In this paper, an IoT based food and formalin detection technique is developed to detect the presence of formalin using machine-learning approaches. Volatile compound HCHO gas sensor connected with Raspberry pi3 were used to extract the concentration of the formalin as a function of output voltage of any fruit or vegetable and different machine learning algorithms were used to classify the fruit or vegetable based on their extracted features. Supervised machine learning algorithms have been incorporated in our system to accurately predict the correct concentration of formalin at all temperatures which is also able to correctly classify between artificially added and naturally formed formalin.

Keywords：Adulteration, formalin, naturally occurring formalin, formaldehyde, machine learning, IoT, HCHO gas sensor, Raspberry Pi3, supervised machine learning algorithms.

1. Introduction

The intake of any food substance is intended for the nourishment which is gained from it. Since the food is into consecutive stages of production, processing and finally distribution, the nourishment in the food items is collapsed. For the food products to remain improved in texture, storage and appearance, a concept of adulteration is widely practiced. The nature or quality of the food is reduced through addition of adulterants or removal of vital substance by the process of food adulteration. The adulterants may be a foreign or inferior chemical substance present in food that causes harm to the food. In the process of food adulteration, little quantities of non-nutritious substances are added knowingly to improve its appearance or storage properties of the food.

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. This chemical is used to prevent the dead bodies from decaying. This property is used to decay. Not every case of adulteration will result in serious adverse health effects. But the chemical is highly toxic and a 30ml of formalin containing 37 percent of formaldehyde can kill an adult. Formalin is used as a preservative by the traders to improve the appearance of fruits and vegetables and to sustain for longer periods.SEM Secretary Abdus Sobhan found in a research where in 115 samples of mangoes and other fruits were collected from over 50 shops which was organic shops and promises chemical free fruits was treated with formalin.

2. 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 preservation [1].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 or vegetables [2].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[3].

3. Overview of the system

3.1. Block diagram

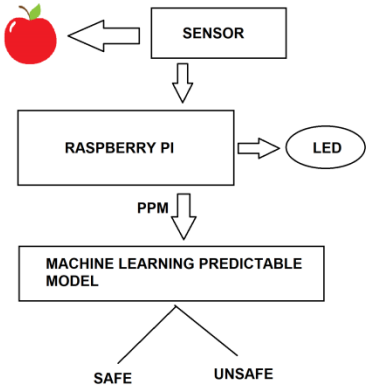


Fig 1: Depiction of the formalin detection in fruits.

3.2. 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 varies in different food items. The dosage of external formaldehyde could be found by figure out the naturally generated formaldehyde content in food. The below table 1 lists the naturally occurring formaldehyde content is various fruits and vegetables.

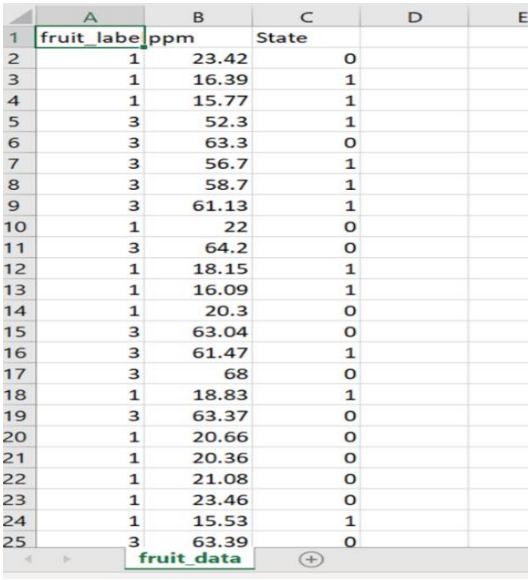
Fruit name	Naturally occurring formladehyde
Apple	6.3 – 22.3
Apricot	9.5
Banana	16.3
Cucumber	2.3 – 3.7
Grape	22.4
Pear	38.7 – 60
Plum	11.2
Tomato	5.7 – 13.3
Water-melon	9.2

Table 1: Naturally occurring formaldehyde values of several fruits.

4. Design and Experiments

4.1. 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 data sets 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 present formalin 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) [4]. Relating to [5] we created our own data set which consists of three columns fruit_label, ppm 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 (Fig 2) was divided into the training dataset and test dataset - 75% for training the model and 25% of the data for testing it.



	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: Output Voltage dataset

4.2. 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 [6]. More watery and chemical compound will allow more electrical energy to meet up with[7]. Acidic fruits are good conductor of electricity. Grove HCHO which is a semiconductor VOC sensor used for detecting the formaldehyde combining with Raspberry-Pi. This sensor featured to detect gas concentration up to 1 ppm. 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[8] with the concentration of formalin present in the fruit sample.

4.3. 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 [set of IF-THEN rules], which first let us categorize the fruit type[9]. 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 and trained using the data set which also depicts the naturally added ppm value along with the additionally added formalin. This system generates the output by making predictions whether the particular fruit item is 'dangerous' or 'safe'. Fig3 depicts the process flow.

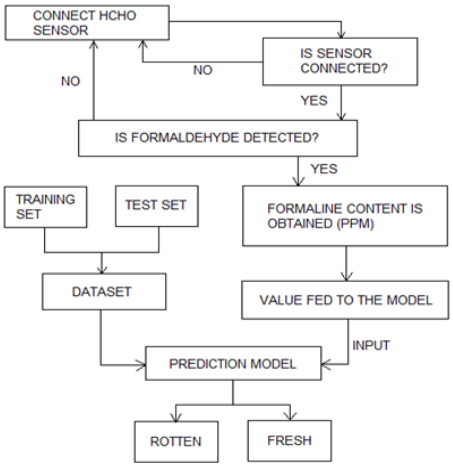


Fig 3: Flowchart of the proposed system

5. Result and Discussion:

5.1. 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. 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 of the classification algorithm[10] results in identification of the objects in 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 2-An accuracy of 70% is obtained in the train set of fruits named train set y and 40% is obtained on the test set of fruits named test set y.

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

Table 2: Depiction of algorithm and the corresponding accuracy in train set and test set.

The support vector machine concentrates on the complicated 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 fruits named train set y and 33% is obtained on the test set of fruits named test set y.

Another algorithm which is widely used for classification is the k-Nearest Neighbors.The object x, which is unknown in the query image is compared with each and every 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.

6. Conclusion:

Thus the paper depicts formalin detection based on machine learning approaches. The manual system fails to detect the formaldehyde level accurately. This proposed system is a dynamic and reliable food and formalin detection technique based on machine learning approaches. Detection of food contamination using arduino is one of the simple methods which produces result by measuring the resistance present in the food. Sensing raw formalin without a predefine model of naturally formed formalin result could be misleading. Hence this system traces the artificially added formalin as a preservative binary "1", otherwise "0". Different machine learning algorithms i.e, Logistic regression, Support Vector Machine, K-NN Classifier are applied to the experimental dataset to build a predictive model. Conductive properties were used to detect the type of foods. The designed system is able to detect 1-50 ppm(parts per million) of formalin using VOC HCHO gas sensor combining with Raspberry-Pi.This machine learning approach for formalin detection detects the formalin concentration in any food item and the consumption could be based on the safety status detected on the food items.

Acknowledgement

We express our deepest gratitude to our Project

Guide and Project Coordinator who gave constant support and also our college for encouragement throughout the work.

References

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

[2] Reza.A.W ,Barman.S.D, Kanij Tabassum, Memi.A.A and Nasrin Sultana “Food and Formalin Detector Using Machine Learning” International Journal of Machine learning and computing,October 2019

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

[4] Formaldehyde in Food[2018] (Online)Available: https://www.cfs.gov.hk/english/programme/programme _rafs/program me_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] Mahya.S, Adlim.M, Fitri.Z, Amri.Y and Sari.M, “Preparations of chemical sensors for simple formalin detection in contaminated food", Annual International Conference,2011. Wilson.A, "Diverse applications of electronic-nose technologies in agriculture and forestry”, Multidisciplinary Digital Publishing Institute (MDPI),’13.

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

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

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

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

AUTHOR PROFILES

[1] **Shri Harini G** is pursuing her final year in Computer Science and Engineering in Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India. She have a large interest to work in AI.

[2] **Swathi P** is also pursuing her final year B.E Computer Science and Engineering in Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India. She have a great interest to learn new technology.

[3] **Vishal N** is pursuing his final year B.E Computer Science and Engineering in Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India. He is working to help the society with the help of new technologies.

[4] **Prince Sahaya Brighty S** is working as an Assistant Professor (Sr grade) in the Department of Computer Science and Engineering in sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India. Her research interest include DataScience, AI and Machine Learning.