Food and Formalin Detector Using Machine Learning Approach

Kanij Tabassum, Afsana A. Memi, Nasrin Sultana, Ahmed W. Reza, and Surajit D. Barman

***Abstract*—Unethical use of formalin, in the preservation of food items posturing threat to communal nutrition. Without chemical experts accurately Formalin detection is a time consuming and complicated task. Moreover, the presence of naturally occurring formalin in food items may interfere in detecting artificially added formalin. This paper presents a dynamic and reliable food and formalin detection technique based on machine learning approaches. Different machine learning algorithms i.e., Naïve Bayes, 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 of formalin using VOC HCHO gas sensor combining with arduino-uno. Several Tests are conducted and polynomial regression has been applied to presume the application of formalin.**

***Index Terms*—Artificially added formalin, arduino uno, HCHO gas sensor, machine learning.**

I. Introduction

Food is indispensable for surviving in the world. People consume food for the nutritional support. But now these foods especially fruits are intoxicated with the most dangerous chemical named Formalin. The extensive use of formalin, in the freezing of fish, fruit and other food items became a threat for public health. Formalin is an aqueous solution that is a mixture of water and 40% of formaldehyde [1]. Some seller use this chemical as a solution for keeping row fishes look fresh and making fruits tempting [2]. In some cases this harmful chemical is used for preserving dead bodies from rotting, but now being used to preserve edible items. The exceeding level of formalin in air cause unfavorable outcomes like watery eyes; burning perceptions in one's eyes, nose, and throat; coughing; wheezing; nausea; and skin irritation [3]. The statement from NIOSH, 20 ppm of formalin is very harmful for human life [2]. Considering the harmful condition of the human, OSHA set many standards to use formalin. In the workplace like lab, office permissible exposure limit (PEL) for formalin is 0.75 parts per million parts of air (0.75 ppm) [3]. Another standard is the short-term exposure limit (STEL) that's range is 2ppm.

Manuscript received May 3, 2019; revised July 31, 2019. This work was supported in part by East West University.

Kanij Tabassum, Nasrin Sultana, Ahmed W. Reza, and Surajit D. Barman are with East West University, Dhaka, Bangladesh (e-mail: tabassumkanij@gmail.com, dinabintemisbah@yahoo.com, wasif@ewubd.edu, surajitbarman012@ewubd.edu).

A. A. Memi was with East West University, Dhaka, Bangladesh. She is now with Opus Technology Limited, Dhaka, Bangladesh (e-mail: afsana.ewu@yahoo.com).

The extensive admixture of up to date food items with malignant concentrations of formalin has become a crucial outcome among Peoples of Bangladesh, which causes health risk of generals. Due to lack of adequate observation and raids, an excessive concentration of solution has been found in most fruits of the native market. Most of our citizens frequently shop for fruits and vegetables from the edge hawkers [4]. These places are not monitored by mobile courts. Another problem is that, Formalin naturally found in foods, such as meats, fish, fruits and vegetables, dried mushroom and crustacean, as a typical metabolic by-product. It could interfere with artificially added formalin when they are going to be detected. The amount varies with the food items. For this reason, fruit identification is necessary for separating natural occurring formalin from artificially added formalin. Only sensing raw formalin without a predefine model of naturally formed formalin result could be misleading. A digitalized device with pre-trained model could be a better solution for non-technical people using their own sell phone.

Fruits and vegetables can be detected in many ways via machine-learning approaches [5]. Arduino based food detection is a simple method of detecting sample food by measuring the resistance in them. To know the concentration of formalin a detector can be used. The detector can detect the formalin by sensing the presence of formalin in the air [6]. Electronic-nose technology for sensing formalin studied in [7]. Their applications applied in food and dairy industry considered in [8], pharmaceutical industries and food application using e-noses tongues [9] are also reviewed in the paper [10]. Some paper proposes a microcontroller based electronic formalin detection system along with an Android smartphone app to show the results and status of sample food [2], [4]. In paper [4], their approach was to detect Formalin incorporating supervised learning. A related study proposed on detecting the presence of formaldehyde in air in parts per million (ppm) unit and adding SD technology [3].Another approach to sense formalin chemically in contaminated food premeditated [11]. Simple spot test quantification methods to determine formaldehyde in aqueous samples were also proposed [12]. Many researches were conducted to detect the extent of formalin use in fish available in Dhaka city [13], [14]. They measure the amount of formalin on their samples using formalin kit that is developed by Bangladesh Council of Scientific and Industrial Research (BCSIR). There are also several other methods proposed for formaldehyde detection, including (pulsed amperometric detection) PAD method [15], solid phase microextraction (SPME) extraction and gas chromatography-mass spectroscopy (GC–MS) analysis, gas chromatography-mass spectrometry with selected ion monitor (GC-MS-SIM) [16], by enhanced chemiluminescence, using cd-doped TiO_2-SnO_2 sensor [17]

doi: 10.18178/ijmlc.2019.9.5.847

and highly sensitive colorimetric detection method for gaseous formaldehyde [18]. However, sensing formalin chemically in contaminated food is not appropriate for general use.

In this paper, an IoT based food and formalin detection technique is developed to detect the presence of formalin particularly in fruits and vegetables using machine-learning approaches. Different machine learning algorithms were used to classify the fruits and vegetables based on their extracted features. Volatile compound HCHO gas sensor connected with arduino-uno were used to extract the concentration of the formalin as a function of output voltage in any fruits or vegetables. Later, rule-based classification and polynomial regression algorithms were applied to predict the concentration of the formalin from the extracted output voltage.

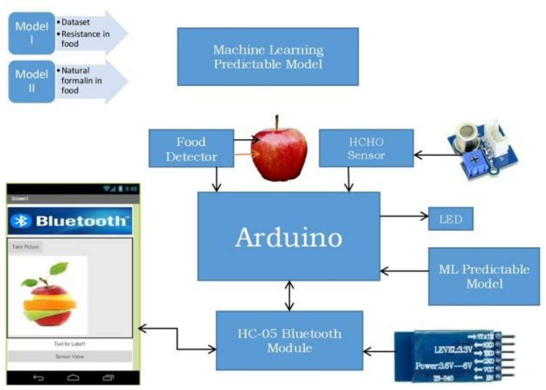

Fig. 1. Block diagram of the food and formalin detector device.

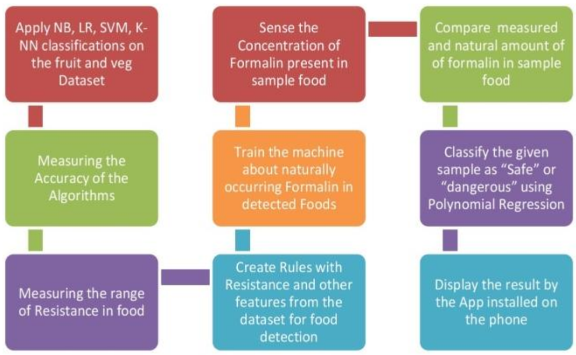

Fig. 2. Working procedure of the food and formalin detector System.

II. Overview of the System

The system detects the sample fruit by measuring the different range of resistances using serial output of Arduino. System uses the voltage drop to calculate the resistance associated with the food items. Rule based machine learning decision model is used to detect the given fruit by comparing the range of resistances in different fruits along with other features. The HCHO sensor senses the concentration of the formaldehyde from the detected fruit by placing near it. Considering naturally occurring formalin [19] in a fruits, our system identifies the level of artificially added formalin in a sample food. HC-05 Bluetooth module is used to interface the device with our self- made Android application. Apps display the food status along with concentration of formalin. A LED is used as an indicator for the presentence of high level of formalin concentration. The complete block diagram of the developed food and formalin detection system is shown in Fig. 1. The system shows status based on the detected fruit, the output voltage of the sensor and concentration of formalin in ppm value. Comparing with naturally occurring formalin in food it considers either the food is "safe" or "Dangerous" for health. Fig. 2 shows the working procedures of the overall system.

III. Design and Experiments

A. Preparing Dataset

For distinguish between different fruits we used a small fruit dataset like [24] in our experiment. The dataset has fifty-nine pieces of fruits and seven columns with fruit's label, fruit's name, fruit's subtype, mass, width, height and color score. There are four types of fruits in the dataset: 'apple' 'mandarin' 'orange' 'lemon'. To measure the performance of different algorithms, confusion matrix was introduced to represent the relation between the actual and predicted class. Feature scaling was used to the dataset for scaling all the features at the same range.

TABLE I: Confusion Matrix

		Prediction	
Actual	***Total= TP+TN+FP+FN***	***Positive***	***Negative***
	Positive	TP	FN
	Negative	FP	TN

From this confusion matrix, accuracy is calculated by using (1),

$$\text{Accuracy} = \frac{TP + TN}{Total} \tag{1}$$

Then we used this preprocessed dataset to train the predictive model to detect given sample fruits.

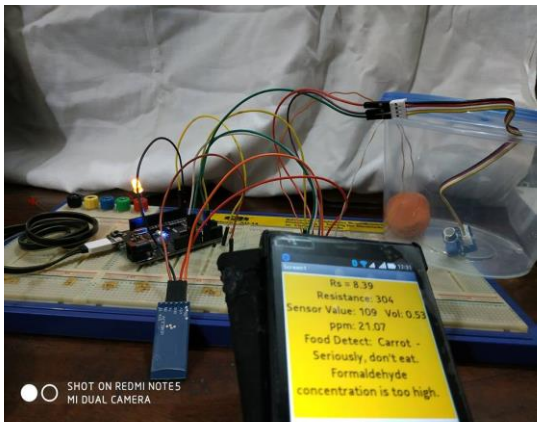

Fig. 3. Complete set-up of food and formalin detection system.

B. Food Detector

Vegetables and fruits could be good source of conductor of electricity because of the presence of juice in them which forms mild acid. Acids are able to conduct electricity just like galvanic cell where conductivity depends on the amount of chemical and water in them [20]. More watery and chemical compound will allow more electrical energy to pass through. Acidic fruits are good conductor of electricity. The designed system works based on a very simple version of ohm-meter using Arduino. Analog pin is used to measure the voltage drop across the resistance in the piece of fruit or vegetable. However, the resistance values kept changing anomalously from time to time due to the chemical changes inside the fruit.

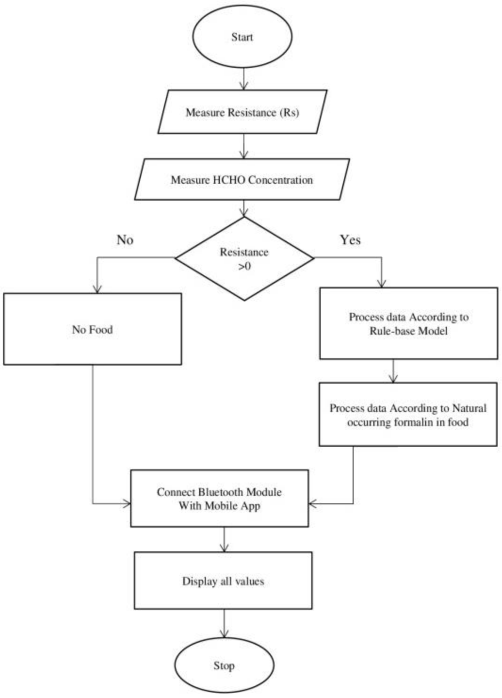

Fig. 4. Flow chart for the complete program.

C. *Formalin Detector*

Grove HCHO which is a semiconductor VOC sensor used for detecting the formaldehyde combining with arduino-uno. 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 with the concentration of formalin gas [21].

D. *Android App Development*

Android application was developed by using MIT App Inventor 2 which is a cloud based integrated Development Environment (IDE). Graphical interface is used here which is very similar to Scratch. Block-based programming language is used to build the Application [22]. The Designer, Block Editor and Android Emulator are the three section of the app inventor where the work takes place. The complete design of the developed system of food and formalin detection is shown in Fig. 3. This package is composed with arduino-uno, sensor, bluetooth module, LED and developed android app. The flow chart of the implemented system is illustrated in Fig. 4.

E. *Developing of the Models*

Different foods contain different range of resistance and arduino's analog pin was used to measure the voltage drop across that resistance in a piece of fruit or vegetable. Sample fruits e.g. apple, cucumber, carrot was selected and measure their resistances to create a range using serial output of arduino. Rule-based classification model was implemented using the extracted features e.g. output voltage, resistance, color, width, height and mass in order to classify the fruit or vegetable samples. The developed pseudo code for classification of the fruit/vegetable samples are given as

R1: IF fruit_subtype = "unknown" and Color= "green" and (Res >400 and Res <700)
THEN Fruit= Cucumber
R2: IF fruit_subtype = "braeburn" and width = 7.1 and Color= "red" and (Res >140 and Res <300)
THEN Fruit= apple
R3: IF fruit_subtype = "estern" and Color= "orange" and (Res >300 and Res <399)
THEN Fruit =Carrot
R4: IF fruit_subtype = "spanish_belsan" and Color= "yellow" or "green" and height = 0.72
THEN Fruit= lemon

F. *Experiment on Detecting Artificially Added Formalin*

Formaldehyde is one of the chemical compounds that naturally presents in many plants and animals for their normal metabolism [18], [19]. Concentration of natural occurring formalin in food items varies according to conditions and types of foods. During detecting the artificially added formalin, this naturally present formalin in the food items may generate wrong results. To avoid this interference, the developed model was trained with the data of naturally occurring formalin in fruits and vegetables collected from the Centre for Food Safety (CFS) [23].

In this experiment, a sample fruits i.e. carrot with HCHO sensor was placed inside a 1-liter closed container, added 5 milliliters of formaldehyde on the carrot, and keep it for 3 hours (Fig. 6). The experiment was conducted at a temperature of 27ºC.

The readings of the targeted gas were taken to determine the presence of artificially added formalin defined by (2). System follows basic K-NN machine learning principle.

$$DF = 1 - \left(\frac{TC - OV}{TC}\right) \tag{2}$$

where, DF = Detection of formalin
TC = Targeted concentration of formalin
OV = Observed value of the sensor

Here, the targeted concentration of the formalin is 1. Observed value would consider 1 if detected concentration of formalin is greater than natural occurring formalin and 0 for less than natural amount. The experiment was repeated for different fruit and vegetable samples with 5 ml of formaldehyde and measure the accuracy of the system. All the extracted results were transferred to the developed android app via bluetooth module.

G. *Experiment on Raw Formalin Detection*

The aim of experiment was to find out the system response for various concentration of raw formalin. At first, the temperature and humidity of the environment was measured. 14 different concentrations of constant level (1ml) of formalin were taken and response of the device was measured. Three trials were conducted at three measured temperature and humidity for testing the consistencies of the data. Relationship between output voltage and concentration of formalin can be mathematically represented by (3) which is stated as [5].

$$Y = \omega_0 + \omega_1 * x + \omega_2 * x^2 + \omega_3 * x^3 \tag{3}$$

Fig. 5. Apple and cucumber detected by the device.

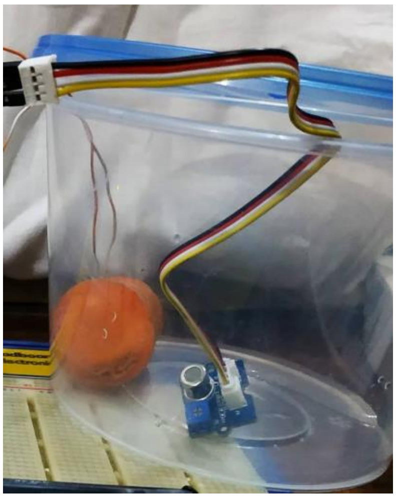

Fig. 6. Experimental set-up where sensor and carrot is placed inside the container.

IV. Result and Discussion

A. Performance of Various Algorithm

Different algorithms have provided different accuracies and their performances were measured. The accuracy of each of the algorithms was calculated from the confusion matrix given in Table II. As observed, the K-nearest neighbors (k=5) provides highest accuracy (100.00%), where the Support Vector machine performs least in terms of accuracy (36%). Table III shows the system's performance in detecting fruits and vegetables in terms of extracted values of their resistance, color and mass using rule-based machine learning approach. As seen, the applied technique effectively detects foods like apple, cucumber, carrot, red plum. Fig. 5 illustrates the experimental outcome in fruits and vegetables detection.

TABLE II: Accuracy of the Algorithms

Algorithm	Accuracy On train set y	Accuracy On test set y
Naïve Bayes	86 %	67%
Logistic Regression	70 %	40%
Support Vector machine	61%	33%
K-NN	95%	100%

Fig. 7 shows extracted value of the level of natural occurring formalin in different types of fruits and vegetables. The device was trained with these measured values of natural occurring formalin for further compare with the measured values of artificially added formalin. Table IV shows the experimented values of the concentration of formalin as a function of output voltage of the HCHO sensor in various fruits and vegetables. Food with asterisk (*) mark represents the addition of artificial formalin on them. The experimental setup of the system can appropriately detect 1-50 ppm of formalin. System predicts the presence of artificially added formalin as binary "1", otherwise "0". Based on these predictions, the system generates the status as "safe" or "dangerous" for any particular food items. Fig. 8 shows the measured output voltage of the sensor for the corresponding natural and artificially added formalin. Fig. 9 shows the outcome of the developed android app in detecting the concentration of formalin in a food sample. The developed app is capable of identify the level of either natural or artificially added formalin in ppm.

TABLE III: Training Set for Detecting Fruits and Vegetables from Second Experiment

Fruit	Resistance Range	Color	Mass (g)	Correct(1) Incorrect(0)
Apple	Res >140 and Res <300	Red	172	1
Cucumber	Res >400 and Res <700	Green	160	1
Mango	Res >410 and Res <700	Green	40	0
Tomato	Res >510 and Res <710	Red	149	1
Carrot	Res >300 and Res <399	Orange	320	1
Red plum	Res >100 and Res <279	Red	42.5	1
Banana	Res >200 and Res <500	Yellow	200	0

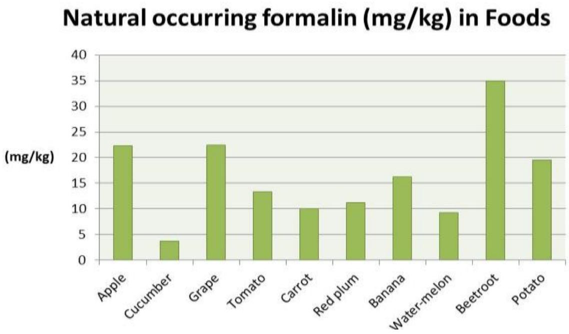

Fig. 7. Foods contain naturally occurring formaldehyde.

TABLE IV: Naturally Occurring and Artificially Added Formalin Prediction

Fruit	concentration of formalin (V)	System Prediction	Status
Apple	0.25	0	Safe
*Apple	0.32	1	Dangerous
Cucumber	0.18	0	safe
*Cucumber	0.32	1	Dangerous
Grape	22.4	0	Safe
Tomato	0.25	0	Safe
*Tomato	0.48	0 (wrong prediction)	Dangerous
Carrot	0.20	0	Safe
*Carrot	0.25	1	Dangerous
Red plum	0.19	0	Safe
Banana	0.21	0	Safe

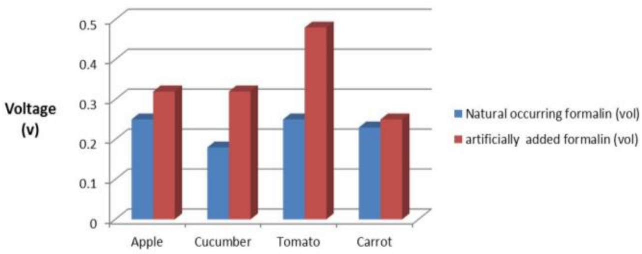

Fig. 8. Measured voltage (V) of the sensor for natural and artificially added formalin in different food items.

Fig. 10 shows the graph of the experimental result of the target gas for different concentration of raw formalin. The graph indicates the system's response for high concentration of formalin for three different trials (T1, T2, and T3). Using the polynomial regression in (3), the concentration of the formalin can be determined from the sensor's output voltage depicts in Fig. 11.

Fig. 9. Concentration of formalin in ppm (parts per million) displayed in android app.

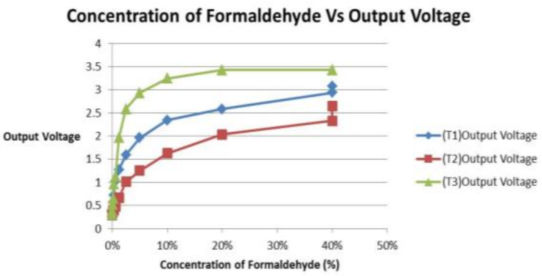

Fig. 10. Concentration (%) of formalin versus system response (V).

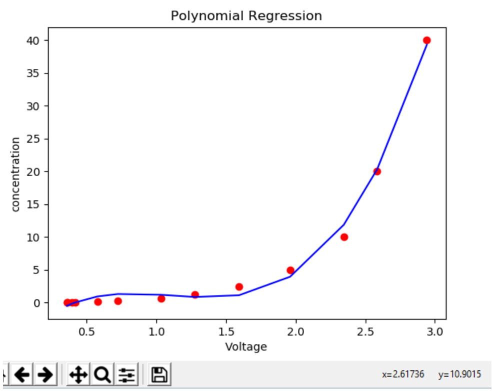

Fig. 11. Relationship between output voltage and concentration of formalin.

V. Conclusions

The development and commercial usage of different artificial technologies is increasing in rapidly throughout the world [25]. This paper demonstrates an IoT based simple food and formalin detection technique using machine learning approach. The present manual system cannot accurately detect the level of formaldehyde or even the presence of it in any comestible products. The developed formalin detection system is able to detect accurately the concentration of formalin present in any food items and also aware the users to identify the safety status of that items before consuming.

Acknowledgment

The authors would like to express gratitude to East West University for funding the presentation and publication fees of the paper. A warm thank to the support of Computer Science and Engineering Department faculty members.

References

[1] M. Rahman, M. Sultan, M. Rahman, and M. Rashid, "Food adulteration: A serious public health concern in Bangladesh," *Bangladesh Pharmaceutical Journal*, vol. 18, no. 1, p. 1, Dec. 2015.

[2] A. Mamun, A. A. Mamun, A. Begum, M. S. Alam, and M. Asaduzzaman, "Microcontroller-based formalin detection: Hardware design and android app development," BSc. Thesis, Dept. of Elec. and Electronic Eng., University of Dhaka , Dhaka, 2017.

[3] C. Cortez, F. C. Bato, and T. J. G. Bautista, "Development of formaldehyde detector," *International Journal of Information and Electronics Engineering*, vol. 5, no. 5, p. 385, 2015.

[4] S. S. Saha, M. Siraj, and W. B. Habib, "FoodAlytics: A formalin detection system incorporating a supervised learning approach," in *Proc. 2017 IEEE Region 10 on Humanitarian Technology Conference (R10-HTC)*, Dhaka, Bangladesh, 2018, pp. 26-29.

[5] W. C. Seng and S. H. Mirisaee, "A new method for fruits recognition system," in *Proc. International Conference on Electrical Engineering and Informatics*, Aug. 2009, vol. 1, pp. 130-134.

[6] C. Cortez and J. Santos, "Evaluation of the developed formaldehyde detector," *International Journal of Environmental Science and Development*, vol. 7, no. 6, pp. 449-452, 2016.

[7] A. Wilson, "Diverse applications of electronic-nose technologies in agriculture and forestry," *Sensors*, vol. 13, no. 2, pp. 2295-2348, 2013.

[8] P. R. Chaudhari and A. L. Judal, "Sensory instruments and its application in dairy and food industry," in *Proc. International Conference of Advance Research and Innovation, ICARI'15*, India, 2015.

[9] E. A. Baldwin, J. Bai, A. Plotto, and S. Dea, "Electronic noses and tongues: Applications for the food and pharmaceutical industries," *Sensors*, vol. 11, no. 5, pp. 4744-66, 2011.

[10] B. Amol and M. Thirugnanam, "Quality measurements of fruits and vegetables using sensor network," in *Proc. the 3rd International Symposium on Big Data and Cloud Computing Challenges*, Feb. 2016, pp. 121-130.

[11] M. Adlim, Z. Fitri, Y. Amri, M. Sari and S. Mahya, "Preparations of chemical sensors for simple formalin detection in contaminated food (A model for final project works for pre-service teachers in chemistry subject)," in *Proc. the Annual International Conference*, *Syiah Kuala University-Life Sciences & Engineering Chapter*, 2011, vol. 1, no. 2.

[12] A. Dar, J. Amara, U. Shafique, J. Anwar, and A. Naseer, "A simple spot test quantification method to determine formaldehyde in aqueous samples," *Journal of Saudi Chemical Society*, vol. 20, pp. S352-56, 2016.

[13] R. Uddin, M. Wahid, T. Jasmeen, N. Huda, and K. Sutradhar, "Detection of formalin in fish samples collected from Dhaka City, Bangladesh," *Stamford Journal of Pharmaceutical Sciences*, vol. 4, no. 1, 2011.

[14] F. Bianchi, M. Careri, M. Musci, and A. Mangia, "Fish and food safety: Determination of formaldehyde in 12 fish species by SPME extraction and GC–MS analysis," *Food Chemistry*, vol. 100, no. 3 pp. 1049-53, 2007.

[15] S. Ngamchana and W. Surareungchai, "Sub-millimolar determination of formalin by pulsed amperometric detection," *Analytica Chimica Acta*, vol. 510, no. 2, pp. 195-201, 2004.

[16] O. Lau and S. Wong, "Determination of plasticisers in food by gas chromatography-mass spectrometry with ion-trap mass detection," *Journal of Chromatography A*, vol. 737, no. 2, pp. 338-342, 1996.

[17] W. Zeng, T. Liu, Z. Wang, S. Tsukimoto, M. Saito and Y. Ikuhara, "Selective detection of formaldehyde gas using a Cd-doped TiO_2-SnO_2 sensor," *Sensors*, vol. 9, no. 11, pp. 9029-9038, 2009.

[18] P. Wahed, M. Razzaq, S. Dharmapuri, and M. Corrales, "Determination of formaldehyde in food and feed by an in-house validated HPLC method," *Food Chemistry*, vol. 202, pp. 476-483, 2016.

[19] F. Nowshad, M. Islam, and M. Khan, "Concentration and formation behavior of naturally occurring formaldehyde in foods," *Agriculture & Food Security*, vol. 7, no. 1, p. 17, 2018.

[20] Galvanic Cell. [Online]. Available: https://en.wikipedia.org/wiki/Galvanic_cell
[21] Grove - HCHO Sensor - Seeed Wiki. (2017). [Online]. Available: http://wiki.seeed.cc/Grove-HCHO_Sensor
[22] D. Walter and M. Sherman, "Learning MIT App Inventor a hands-on guide to building your own android apps," U S. River, NJ: A. Wesley, 2015.
[23] Formaldehyde in Food. (2018). [Online]. Available: https://www.cfs.gov.hk/english/programme/programme_rafs/programme_rafs_fa_02_09.html
[24] Fruit data with colors. [Online]. Available: https://github.com/susanli2016/Machine-Learning-with-Python/blob/master/fruit_data_with_colors.txt
[25] N. Ivkovic, D. Jakobovic, and M. Golub, "Measuring performance of optimization algorithms in evolutionary computation," *International Journal of Machine Learning and Computing*, vol. 6, no. 3, pp. 167-171, 2016.

Kanij Tabassum is currently finishing her bachelor of science in computer science and engineering at East West University, Bangladesh. Her major field of research interest concentrates on data science and robotics. She has also strong interests in cloud computing, cyber security and Internet of Things. She is a member of Bangladesh Robotics Foundation since August 15.

Afsana Azad Memi pursued her bachelor of science in computer science and engineering from East West University, Bangladesh. She mainly focuses on trending problems for research and currently working on IoT based projects. She has also strong interests in cloud computing and machine learning. She is currently working as a project manager in Opus Technology Limited, Bangladesh.

Nasrin Sultana is currently finishing her bachelor of science in computer science and engineering from East West University, Bangladesh. Her research interest includes machine learning and artificial intelligence. She is a self-motivated person. She loves to work with new challenges.

Ahmed Wasif Reza is currently working as an associate professor and chairperson in East West University, Faculty of Engineering, Department of Computer Science and Engineering, Dhaka, Bangladesh. Previously, he was attached with University of Malaya, Department of Electrical Engineering, Faculty of Engineering, Malaysia for almost 8 years where his last position was senior lecturer (equivalent to assistant professor). His research interest includes radio frequency identification (RFID), radio wave propagation, internet of things, wireless sensor network, wireless communications, biomedical image processing, and cognitive radio and electromagnetic research, both in industrial exposure and academically research valued work. He has authored and co-authored a number of Science Citation Index (SCI) journals and conference papers. He has research collaboration with leading international companies/industries, like Motorola, Silterra, Hexamatics, Malaysian Nuclear Agency, Keysight, Agilent, Huawei, etc.

Surajit Das Barman is currently working as a senior lecturer in East West University, Faculty of Engineering, Department of Computer Science and Engineering, Dhaka, Bangladesh. He received his M.Eng.Sc. degree in electrical engineering at University of Malaya, Malaysia in 2016. In 2009, he completed his bachelor in electrical and electronic engineering from Chittagong University of Engineering & Technology (CUET), Bangladesh. His research interests include wireless power transfer, power system analysis and internet of things.