IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 4, 2021, accepted March 24, 2021, date of publication April 2, 2021, date of current version April 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3070558

Development of Spectroscopic Sensor System for an IoT Application of Adulteration Identification on Milk Using Machine Learning

N. SOWMYA AND VIJAYAKUMAR PONNUSAMY, (Senior Member, IEEE)
Department of ECE, SRM Institute of Science and Technology, Chennai 603203, India
Corresponding author: Vijayakumar Ponnusamy (vijayakp@srmist.edu.in)

ABSTRACT Adulteration in milk is a common scenario for gaining extra profit, which may cause severe harmful effects on humans. The qualitative spectroscopic technique provides a better solution for detecting the toxic contents of milk and foodstuffs. All the available spectroscopic methods for milk adulterant detection are based on laboratory-based with costly equipment. This laboratory-based detection takes a long time and is more expensive, which may not be afforded by a common man. To overcome this issue, this research work involves the design and development of a low-cost, portable, multispectral, AI-based, non-destructive spectroscopic sensor system that can be used to detect the milk adulterant in real-time. The designed sensor system uses the spectroscopic method with wavelength ranges from (410-940nm) which consists of three different bands Ultraviolet (UV), visible, and Infra-Red(IR) spectrum to improve the accuracy of detection. The sensor system is connected to the internet via the developed IoT application module, which displays the detected adulterant results in a dedicated web page designed for this purpose. This IoT application enables the adulterant detected results published on the internet immediately with location information for bringing transparency. Adulterant detection problem is formulated as a classification problem and solved by machine learning algorithms of a decision tree, Naive Bayes, linear discriminant analysis, support vector machine and neural network model. The average accuracy of linear discriminant analysis, support vector machine, Naive Bayes, decision tree and neural network model are obtained as 88.1%, 90%, 90%, 91.7% and 92.7% respectively. Genetic algorithm framework is formulated for hyperparameter tuning of neural network model which improved the accuracy from 92.7% to 100%. The model is trained for five different classes of four adulterants, namely Sodium Salicylate, Dextrose, Hydrogen Peroxide, Ammonium Sulphate, and one pure milk sample.

INDEX TERMS Back propagation algorithm, K- means clustering, machine learning, milk adulteration, multispectral spectroscopy, neural network.

I. INTRODUCTION

Nowadays,food adulteration is a common scenario among shopkeepers in gaining extra immediate profits. Adulteration is added in foods like ripening mangoes, adding chalk powder on turmeric, starch on curry powder, blending papaya seeds on black pepper, etc. This adulteration venture leads to harmful effects for humans on a long-term basis. The fluid cow milk consumption is about 77.68 million metric tons in India. In December 2019, statistics showed that India plays the best role in cow milk consumption [1]. Milk has powerful nutrients such as lactose, fat, proteins, minerals, and vitamins in a significant proportion, which helps to provide a better human diet [2]. Since the peoples consume the milk daily in their diet, detection of adulterants in the milk is one of the significant research to ensure the health safety of the humans. Spectroscopy-based adulteration detection is one of the methods to tackle the above problem, which is widely used.

The Spectroscopy system offers the measurement, production, and interpretation of spectra due to the interaction of electromagnetic radiation and matter. The analysis methods may differ with respect to molecules or atoms and the type of matter-radiation interaction, such as diffraction, absorption,

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Zhang.

and emission. Based on emission or absorption of radiation in the ultraviolet (UV), Visible(Vis), Infrared(IR), and radio Nuclear- Magnetic Resonance (NMR) frequency ranges are utilized in traditional food analysis [3]. For a soluble, thin-film coating material, UV-VIS spectroscopy would give good results [4]. Various spectroscopic methods are available to know the molecular and atomic levels of interaction for the liquid contents.

The relationship between the concentration of absorption and absorbance of a substance whose chemical constituents are being measured, identified, by Beer lambert's law as,

$$A = \in CL \tag{1}$$

where, A- Absorbance, $\in$ -Molar Absorption coefficient, C - Concentration of substances, L -Path length (cm) Absorbance has no unit; it is the log ratio of beam powers. Molar Extinction coefficient varies for each molecule and it is represented as a constant which has the unit of $Lmol^{-1}cm^{-1}$.

The spectrometer uses a cuvette to measure the sample,which is 1 cm in width, L is the path length, assumed to be equal to 1 cm. The concentration C of the sample was expressed by mole/L.

The spectroscopic methods detect the adulterants, which are a bit expensive,laboratory-based and provide a non-destructive methodology. When considering the adulteration in milk, few adulterants can be blended into the milk to increase its quantity fraudulently.

There are many more literature works reported for adulterant detection, apart from those listed in table 1. Table 1 format is prepared to explore the possible wavelength of response for the milk adulteration. A milk adulterant detection system that can detect adulteration by receiving signal analysis at the frequency of 0.5MHz is presented [5]. A transmitter is used to pass this signal through milk samples of adulterated and the received signal is analyzed using a digital storage oscilloscope. The result indicates that with an increase in adulteration, the voltage level of the signal increases.

TABLE 1. Milk adulterants and its responsive wavelength.

Adulterant	Responsive Wave-length(nm)	Method	Results Achieved	Related Work
Water	1200-1450	NIR	Accuracy -99%	R.de Souza, et.al(2016) [11]
Hydrogen peroxide	2700	FTIR	limit of detection is 200 (parts per million)	P.W.Hansen,et al(2019). [12]
Ammonium sulphate	3420	FTIR	Correlation coef-ficient =83%	P.W.Hansen, et.al(2019). [12]
Sodium hydroxide	2700	FTIR	limit of detection-1%	P.W.Hansen, et.al(2019). [12]
water and whey	1450 -2310	NIR	accuracy- 99%	Kasemsumran, et.al(2007) [13]
Melamine	1100	NIR	Classification Accuracy-89%	De Marchi, et.al(2018) [14]
Hydrogen peroxide	2780	MIR	Correlation-90%	P.M.Santos, et al(2013) [15]

One of the studies indulges in quantifying and detecting the mixture of cow alpha-lactalbumin adulteration in buffalo milk mozzarella-(dairy product) by capillary electrophoresis method [6]. The capillary electrophoresis methodology produces output with the linearity of R2 = 0.968, and for different bovine whey mixtures, ratios obtained various relative standard deviation. Results concluded with detectable fraud on cow milk were 1%, and quantification was 3.1%.

Another work presents a quantitative identification of urea, tap water and starch adulteration in milk using electrical impedance measurements [7]. Impedance analysis shows that adulteration detection is found mostly in the resistive part of the electrical impedance. In this work, 90 KHz signal frequency was used for the milk adulteration identification. The milk sample is analyzed through an equivalent electrical modelling series circuit. This system provides a quantitative measurement for adulterant detection. The results also state that as the concentration of adulteration rises which in turn decreases its pH, voltage value, and conductivity.

A study on the Electrical Impedance Spectroscopy (EIS) technique employed to identify and quantify the percentage of soap weight concentration in three various cow milk samples is carried out [8]. The result shows a steady variation present in the impedance, current, conductance, and capacitance with the milk's measured addition of soap. This method can also detect the presence of >= 0.1% soap in the milk, and it is applicable effectively to about 0.9%.

An electro-analytical cyclic voltammetry test with chemometrics analysis was employed to detect and quantify fresh milk's adulteration with reconstituted skim milk powder [9]. Principal Component Analysis gives a better differentiation between the reconstituted skim milk powder and fresh milk samples. Soft Independent Modelling Class Analogy (SIMCA) and Partial Least Square (PLS) regression deliver better prediction in classification between the two classes. The fresh cow milk samples tested under Multivariate analysis and adulterant are predicted with 100% of accuracy. One of the commonly used adulterants everywhere around the world is water. Normally fresh raw cow milk itself has 75- 80% of water [1]. To increase the volume of the milk, water is added to the milk by the shopkeepers. Hydrogen peroxide is adulterated in milk to activate the enzymes and also it is used where cooling is not necessary to maintain the dairy products. It may cause irritation or redness side effects to human beings. It can be measured by Fourier Transform Infrared (FTIR) non-destructive spectroscopic methods.

Ammonium sulphate is blended in milk since it is a nitrogen-rich component and it can be identified by analyzing the milk through lactometer reading though it can maintain the density of the milk. Sodium salicylate is used to increase lactation and is also used as a preservative. For mature dairy cows' milk to show an increased fat amount, this adulterant was utilized. Dextrose is added into the milk as an adulterant, which is also called as vegan pastry. Specifically, it is made

up of corn, often used for baking products. There are many spectroscopic techniques available to detect adulteration in milk on its specific wavelength. A literature survey is conducted to identify the responsive wavelength of our targeted adulteration in the milk and summarized in Table. 1

From the Table 1. It can be understood that various spectroscopic methodologies are applied to detect different kinds of adulterants. But those techniques are not applicable for the real-time analysis, takes more time to deliver the results, and those mechanism tests are carried out on the lab environment and do not support the field test. Moreover, the equipments are not portable. Those approaches are only capable of detecting one or two adulterants only. But in practice, there will be more number of adulterants mixed in the milk which to be detected. Our research work is motivated to solve the above drawback of the existing mechanism by design and development of multispectral sensor system. Our proposed system is portable, capable of simultaneous delivery of accurate results for multiple adulterants in real-time. The adulterant detected results could be visualized from anywhere through developed IoT applications. The system cost is optimized to 8624/- INR. Since our proposed model does not require any raw material for testing and there is no limitation of testing sample. The flow of the work in forthcoming sections is discussed as follows. Section II describes the adulteration detection related to simulation techniques and cost analysis. Section III delineates the sample preparation, equipment design and experimentation. Section IV discusses the results obtained from the processed spectral data, followed by the conclusion.

II. RELATED WORK

Milk is a nutritious dairy product and highly nutritious in its natural form. It contains carbohydrates, minerals, proteins, amino acids, and minerals. Nowadays, adulteration in raw milk usual to make extra profit and increase the quantity. Some of the adulterants may cause internal infection in adults and infants. Another way of adulteration in species milk is carried out by mixing camel milk with goat milk and cow milk adulteration in buffalo milk. Another paper [16] illustrates the detection of adulterants in milk surveyed with ten different adulterants in different quantities, and the test was conducted using MID- IR spectroscopy and Soft Independent Modeling of Class Analogy (SIMCA) technique. Multiclass modeling implemented for five effective adulterants and results were obtained with 82% correct classification, 17% of inconclusive and 1% misclassified.

The freezing point test is the test for identifying added water in the raw milk [11]. In this work, an instrument was developed to measure substance concentration based on NIR diffuse reflection. It is a non-destructive, fast, and widespread method. To validate this developed mechanism, an experiment was conducted to detect adulteration in milk, especially water-diluted milk content. The results showed 99% of accuracy in identifying the water diluted milk. But this system is not validated for other adulterants.

A prototype able to measure the intensity of IR scattered light from the raw milk samples is presented [17]. This method aims to identify the added water from the raw milk, resulting in a determination index of 0.932 and Root Mean Square Error of Prediction (RMSEP) = 0.0267.

One more classification made between authentic adulterant skim, and nonfat dry milk powder has done through solution state, high field 1 Hydrogen nuclei- Nuclear Magnetic Resonance (1H NMR) spectroscopy combined with conformity index, non-targeted method-based analysis [18]. It supports finding the root cause analysis of suspicious results and supporting milk powder authentication and adulterant detection. The results show that the detection of adulteration was high for the lowest 0.05% weight concentration of a solution containing dicyandiamide samples, melamine, and nitrogen-rich samples.

Another study analyzed to overcome the blending effect on the detection of adulteration in milk powder samples with adulterants: ammonium sulphate, cornstarch, and semi carbazide hydrochloride, a nitrogen-rich organic compound [19]. A fraction of skim milk powders were spiked with these unstable potential milk adulterants. Each adulterant at an amount of 50g weigh is mixed with 100ml of milk powder. But this approach requires Pre- sample preparation for doing analysis. Wet and dry- blended adulterated samples tested on Mid- Infrared (MIR) measurement and classified on one-class SIMCA model for reconstituted skim milk. The results show that $< 5\%$ of adulterant levels could be detected.

A fast, non-destructive, quantitative, and simple Fourier Transform Infrared spectroscopy technique is used to determine sucrose in raw cow milk [20]. Also, it is analyzed using Partial Least squares and chemometrics methods. The results show that the Limit Of Detection (LOD) is about 1.25gram/Litre, and the relative Prediction error is below 5% with the spectral range of $920cm^{-1}$ to $1800cm^{-1}$.

Enhanced ν- Support Vector Machine was applied in discriminating the adulterants in milk like Pseudo proteins and solid contents, and analysis made through the spectral analysis of NIR spectra with various sample sets [21]. The NIR spectroscopy combined with improved ν- Support vector machine produces output with maximum validation correct ratio,which Correlates positively with the adulterant solution's content and discrimination adulteration equals or exceeds 5%. A new method of identifying formalin in raw cow milk was analyzed using NIR (700-2500 nm) combined with multivariate analysis [22]. The PLS (Partial Least Square) model obtained a good prediction with a root mean square error prediction value of 1.45% and quantified formalin adulteration with less than 2%.

Another study was conducted using NIR spectroscopy coupled with multivariate analysis to discriminate urea in the raw cow milk samples [23]. All the samples were tested in absorption mode with wave number ranging from $10,000cm^{-1}$ to $4,000cm^{-1}$. This study revealed that the Partial least square Discriminant Analysis model (PLS-DA) could produce better urea discrimination in the milk of about 96.8%.

Further studies identified four adulterants, namely Glucose, formaldehyde, Hydrogen Peroxide, and urea in milk using FT-MIR spectroscopy [24]. FT-IR PLS method produces output with relative errors of 0.0014, 0.03162, 0.03557, and 0.3924, with the spectral range from 4000 and $400cm^{-1}$.

Portable milk adulteration testing system includes impedance spectroscopy and microelectrode sensor is developed [25] named as Hi tester. This Hi tester 3522-50 is used to test the impedance of the samples. This system requires 0.2ml of milk for testing. The frequency range assigned for testing is 0.05- 5.0 KHZ. For the addition of 10% of detergent in the testing sample, the impedance value changes with 56% from the reference value. The results found that if the concentration of adulterants, (starch and detergent) increases the impedance of the tester also increases.

A 785-nm diode laser incorporated Raman spectroscopy employed [26] for the quantitative detection of urea content in the milk sample. The spectral samples are obtained and evaluated with Partial Least Square based chemo metric algorithm. Various concentrations of urea are mixed with milk and the results are analysed. The result shows that the detection accuracy is greater than 90% for the prediction of urea concentration in the milk.Some rapid, non- destructive machine learning and deep learning techniques are employed to detect this adulteration and addition of chemical substances.

The methods reported in the literature are listed in Table. 2 The sample data were extracted in different regions with various standards of milk proportions. The results obtained for Limit of Detection in adulterants identification is $30 - 50\%$. Some of the Processing techniques give classification accuracy of 98% and avoid delay in getting the result.

These methods carried their testing analysis in the lab environment, not in real-time. The on-field analysis makes the system to be more effective and appreciable in decision making. Therefore, in real-time, reliable detection is necessary for the prediction of adulterants on the spot. To overcome these shortcomings, this research work contributed to the following aspect.

- The proposed research work involves designing and developing portable, low-cost sensor device, which can support non- destructive testing in real-time.
- Design of multispectral sensor system is carried out to capture data from three different bands called UV, Visible and NIR with six channels for each group in a total of 18 channels to improve detection accuracy.
- Adulterant detection is formulated as classification problem and solved by applying the naive Bayes, Linear discriminant analysis, decision tree, support vector machine and neural network machine learning algorithms to detect four different adulterants and pure milk. Those models achieves average accuracy of 90%, 88.1%, 91.7%, 90% and 92.7% respectively.
- The neural network model hyper parameter is tuned to improve the accuracy from 92.7% to 100% by formulated optimization problem and solved by genetic algorithm.

TABLE 2. Milk adulterant identification using machine learning techniques.

Adulteration detected	Dataset/Machine learning technique	Accuracy (%)
Detergent, ammonium sulphate,sodium hydroxide,sodium bi-carbonate,common salt in milk [10]	spectral data/ ANOVA analysis	LOD= 0.3%
Formalin in cow milk [27]	spectral samples/ATR-FTIR,SIMCA, PCA,PCR, PLSR	Sensitivity= 0.5%
Water and hydrogen Peroxide in Bovine milk [28]	EIS spectral data/ Neural Network	94.6%
Sucrose, Soluble Starch, Sodium bicarbonate, hydrogen peroxide and formaldehyde in milk [29]	spectral Samples/FTIR/ Deep and ensemble decision tree	98.76%
Camel milk in goat milk [30]	Spectral data /NIR / Multivariate analysis-PLS-DA	97.4%,LOD -0.5%
Cow milk adulteration in buffalo milk [31]	spectral samples/ fluoroscence spectroscopy/PLS	Correlation coefficient = 99%
Goat milk adulteration with cow milk [32]	spectral samples/NIR spectroscopy/ PLS	RMS = 99%
Goat milk adulteration in bovine milk [33]	Six reagents Electric potential signals/ LDA	Sensitivity =97%.
Sodium bicarbonate, sodium hydroxide, hydrogen peroxide, starch, sucrose and urea in milk [34]	spectral data/ATR-FTIR/ MLR	Coefficient ratio= >76.6%
Goat milk purity detection [35]	Pulse excitation signal/ artificial fish swarm	Prediction coefficient = 99.8%.
Hydrogen peroxide in milk [36]	Chromatographic data/liquid chromatography with diode array detection	Pasteurized milk/ Correlation coefficient =99%
Melamine in milk Powder [37]	Laser spectral samples/ Neural networks, PCA	Correlation coefficient =99%
Pseudo proteins and thickeners in Cow milk [38]	NIR spectral samples/ Improved-SVM and KNN	Calibration Correct Ratio=95%
Soybean flour, Rice flour in milk powder [39]	FTMIR-spectral samples/ SVM ,extreme Learning Machine	Sensitivity = 90%
Sweet whey and Acid whey in milk powder [40]	laser spectral samples/ PCM analysis	Accuracy (Whey powder)=80.5%, (Sweet Whey) =98.5%

- The sensor system is made as an IoT device to update the detected adulterant in the web server to access the result from anywhere.

III. MATERIALS AND METHODS

The design of AI-enabled multispectral spectroscopic sensors, sample preparation, spectral data collection, processing

of them and neural network algorithms employed for analysis were discussed in the following sections. Section A describes the developed multispectral sensor system design with Artificial Intelligence methodology for the identification of milk adulterants. The cost analysis of hardware design is also presented. Section B delineates the sample preparation process,which involves collecting milk samples, adulterant varieties and proportion of adulterants blending in the raw pure cow milk. Section C expresses the spectral data acquisition for the adulterant and non -adulterant samples.

A. DESIGN OF AI-ENABLED MULTISPECTRAL SENSOR SYSTEM FOR MILK ADULTERANT DETECTION

The Prototype of the designed AI-enabled sensor system is shown in Fig 1. which consists of Triad Spark fun Spectroscopic system, Arduino blackboard, Qwiic cable, Wireless module(ESP8266), PC is used for data logging and neural network AI software module implemented.

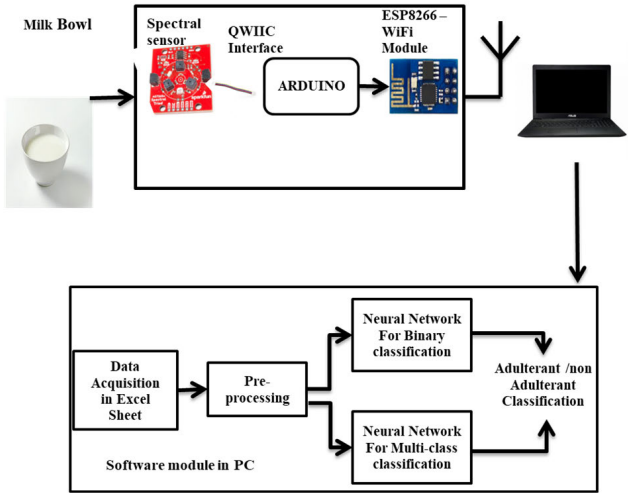


FIGURE 1. The AI Enabled Multispectral Sensor System.

Fig. 1 shows the AI-enabled multispectral sensor system. The AS7265x sensor chipboard is connected serially to the 3.3V compatible Arduino uno blackboard through QWIIC cable interface, which comprises a spectral sensor. The spectral sensor data will be communicated serially to the Arduino microcontroller and transmitted to a Personal computer through a Wi-Fi module. The Wi-Fi module enables the user to communicate the spectral collected data to the PC. The acquired spectral data is stored for pre-processing. The pre-processed sample data are transmitted to the non- linear neural network for further analysis. Those pre- processed data are used to train, test and validate the neural network. One of the network outputs will classify/detect as an adulterant /non- adulterant. Another neural network model also trained for five class classification purposes to classify /detect the four types of adulterants: sodium salicylate, hydrogen peroxide, dextrose, ammonium sulphate, and pure milk samples. The following section describes the individual block of the proposed sensor system.

1) MULTISPECTRAL SENSOR

The design of the sensor system uses the AS7265X triad spectroscopic hardware module from SPARKFUN. The Spark fun triad spectroscopy consists of three optical sensors: i) AS72651 ii) AS72652 iii) AS72653.

Each sensor has six independent on-device optical filters,18 channels of its spectral wavelength range from 410nm to 940nm with a Full-width Half Maximum [FWHM] of 20nm spectral width and 10nm wavelength accuracy. The AS72651 plays a master role having Ultraviolet (UV) wavelengths ranging from 610nm to 860nm. The slave AS72652 in the Visible (Vis) region having spectral response starts from 560nm to 940nm, and the slave AS72653 in the Near Infra-Red (NIR) region having spectral response starts from 410nm to 535nm.

The intensity of light is precisely measured upto $28.6nW/cm^2$ with an accuracy of $\pm 12\%$. It has a light source with LEDs of 405nm UV,875nm IR, and white LED. This optical sensor is 3.3V compatible [40]. It communicates over I2C /serial interface via Qwiic cable by 115200/9600 bps (Bits per second) transfer rate.

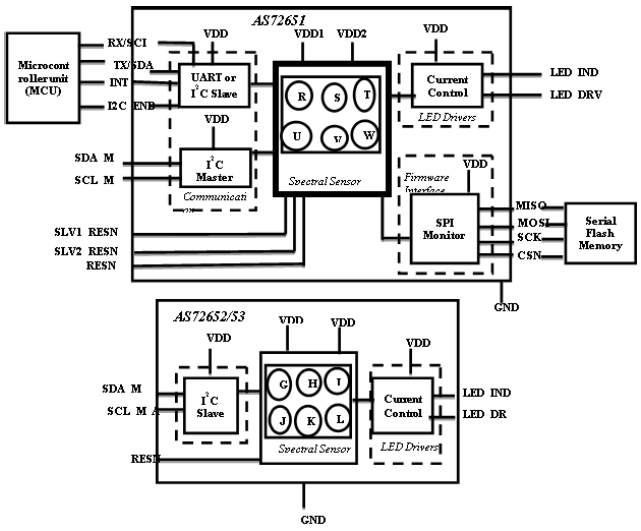


FIGURE 2. Multispectral AS7265X Sensors Chipset Diagram.

Fig. 2 shows the multispectral sensors chipset diagram, which can be utilized for spectral identification in the range from Ultraviolet, visible to Near-Infrared Red [NIR]. This sensor is placed one inch above the testing sample before observing the readings. The AS72651 provides a smart UART interface with the microcontroller, which acquires the sensor's data. This sensor system produces the spectral data based on the received reflected light from the target sample. It has a flash memory to store the data. It provides serial UART communication for communicating spectral data to the computing device, which is the Arduino blackboard in our system. Each chip contains an analog- to- digital converter. Each AS7265X chipset has two integrated LED drivers and a programmable current level circuit to enable this chipset to operate in electronic shutter applications. The package field

TABLE 3. Sparkfun triad Sensor Channel and its wavelength ranges.

Regions	Channel	Wavelength(nm)
Ultraviolet	I,K,M,N,O,P	610,680,730,760,810,860
Visible	G,H,J,L,Q,R	560,585,645,705,900,940
Near Infra-Red	A,B,C,D,E,F	410,435,460,485,510,535

of view is ±20.5. Moreover, this spectroscopic sensor system allows our system to operate in portable mode during the analysis effectively. The Sparkfun Triad sensor has 18-channels wavelength of operation, which is listed in Table 3.

2) ARDUINO BLACKBOARD

The low-cost Arduino blackboard is used to collect the triad sensor's spectral data and communicate those spectral data to the server PC for further analysis. The sensor is connected to the Arduino blackboard via the QWIIC interface. The Arduino blackboard consists of 14 digital Input/output pins with 6 Pulse width modulation pins, 4 Analog Inputs, UART (Universal Asynchronous Receiver and Transmitter) and Serial Peripheral Interface (SPI). It again transmits the collected spectral data to the WiFi module (ESP8266) single-chip pin.

3) DATA ACQUISITION

The Collected spectral data are received in the server PC from the Arduino WiFi module. Microsoft Excel data streamer software module is used to acquire data from the USB port where the WiFi module is connected.This block delineates the data acquisition through an excel sheet. Data streamer provides two-way data transfer that streams the live data from a microcontroller into an excel sheet.

4) PRE-PROCESSING

The collected 18 channel data are formatted in a form that Machine Learning algorithms can accept. Some outlier data and excess data above the standard size were also removed.

5) NEURAL NETWORK MODEL

The pre-processed spectral data are given as an input to the neural network architecture. Fig. 3 shows the neural network architecture with 18 input layers followed by ten hidden layers and a multilayer output for multi- classification.

The spectral pre-processed data fed into the input layer has 18 inputs, since our spectroscopic design has 18 channels with different wavelengths. In this network layout i_1, i_2, ... i_{18}. Represents the input neurons, W_{ij} represents the weights connecting the input to the hidden layer, W_{jk} and W_{kl} represents the weights present in the interconnected hidden layers and W_{lm} represents the weights connecting between the hidden layer to output layer. Initially, the weight and bias values are randomly assigned and processed along with the input data in a forward propagation manner. The output error value is calculated by comparing the actual value and the predicted value, such that the output error will be minimized by using a gradient descent algorithm. Cross entropy is used

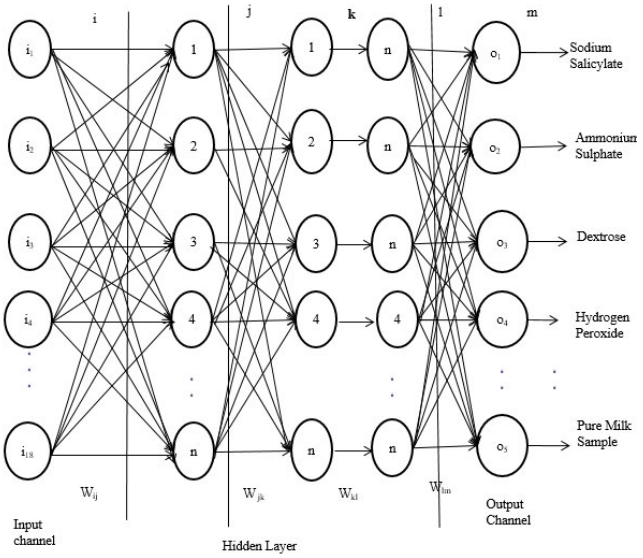


FIGURE 3. Neural Network Architecture.

as a loss function, and it is represented by

$$CE = -\sum_{(i=1)}^{n} Y_i log_2(p_i) \tag{2}$$

where, p_i -probability value;Y_i- one hot encoded value; cross-entropy loss function computes the loss value between the predicted value and actual value, which is used to update the weight value. The weight update function is given as

$$W = w + \delta w \tag{3}$$

where W-updated weight; w- old weight. The sigmoid function is used as an activation function in binary classifiers. It is given by

$$F(x) = 1/(1 + e^{-\beta x}) \tag{4}$$

where, x-input; F(x) = Output.

The Soft-max layer is used at the output layer and the sigmoid function is used in all the other layers. It is represented by

$$\sigma(z_i) = e^{z_i} / \sum_{(j=1)}^{K} e^{z_j} \tag{5}$$

Z_i Elements of the input vector; e_i^z - Exponential function to the input vector, $\sum_{j=1}^{k} e^{z_j}$ Normalization term.

6) HYPERPARAMETER TUNING BY GENETIC ALGORITHM

Hyperparameter tuning is one of the significant processes in the deep learning model design. The machine learning/deep learning model's hyperparameter needs to be selected optimally to provide the best accuracy. In this work, four hyperparameters namely the number of hidden layers L, Initial learning rate μ, maximum number epochs EP_{max}, and the number of neurons in the hidden layer n_h are selected to

provide high accuracy.The parameter selection optimization problem is defined as follows.

$$\underset{L,\mu,EP_{\max},n_h}{Max} \quad Accuracy(A) \tag{6}$$

$$C1 : A > 90\% \tag{7}$$

$$C2 : L \leq 1024 \tag{8}$$

$$C3 : \mu \in \{0-1\} \tag{9}$$

$$C4 : \ n_h \in \{0-100\} \tag{10}$$

In the above optimization problem, the four parameters are selected such that the selected parameter will give the highest accuracy. Constrains C1 to C4 ensure that the selected parameters will be bounded with a range of typical values. The above problem is solved by using a genetic algorithm to find out the optimal hyperparameters. The genetic algorithm is designed using the setting values given in Table. 4. The binary chromosome design is given in Table. 5, where 4 bits are used to encode 16 different hidden layer sizes, 8 bits are used for all other parameters to cover broad range values. Accuracy value is used as a fitness function to be maximized while selecting the best chromosome solution. The genetic algorithm is executed using the above setting, and the optimal parameters are obtained as $L_{opt} = 10$; $\mu_{opt} = 0.477$; n_h opt = 64; $EP_{(maxopt)} = 26$.

TABLE 4. Genetic algorithm parameter setting.

Parameter	Value
Number of Generations	20
Population	50
Chromosome size (binary)	28 bits
Crossover probability	85%
Mutation	3 point
Selection method	Roulette wheel selection

TABLE 5. Chromosome design- bit allocation.

L	μ	n_h	$EP_{(max)}$
4 bits	8 bits	8 bits	8 bits

7) NEURAL NETWORK TRAINING ALGORITHM

Training phase:

Collection of sample in all the three bands:

For Channel = 1 to 18,

Collect the photon counts $\alpha(channel) = \gamma - \gamma \times \in CL$

where γ amount of incident light on the sample End

Split the data into α (training), α (testing) and α (validation)

For all training set

1. For k = 1 to EP_{max}

1 a. Apply α (training set) on neural network and pass through all the layer

2. Compute loss function value at output layer using cross entropy CE = $-\sum_{(i=1)}^{n} Y_i log_2(p_i)$

TABLE 6. Spectroscopic sensor cost details.

Name Of the Component	Cost
The Spark fun Triad sensor	Rs.6,250/-
Arduino Black board	Rs. 1850/-
Qwiic I2C cable	Rs. 150/-
Wi-Fi Module[ESP8266]	Rs. 374/-
Net total	Rs.8,624/-

3. Compute the output error and perform weight adjustment at output layer using optimal learning rate which is found from the genetic optimization mechanism.

Compute the activation hj of each neuron j in the hidden layer

$h_j = \sum_{(i=1)}^{n} \alpha$ (channel)

Compute the output Y_k

$Y_k = g(h_j) = 1/(1 + e^{-\beta h_j})$

Compute Error $\delta_o(k)$ at output layer

$\delta_o(k) = (Y_k - t_k)\, Y_k(1 - Y_k)$

t_k - target output

Adjust weight $W_o(k+1)$ at output layer

$W_o(k+1) \leftarrow W_o - \mu_{opt}\delta_o(k)Y_k$

4. Back propagate the error, compute the error proportion component $\delta_h(k)$for every layer and do weight adjustment $W_h(k+1)$ according to the proportion component and the optimal learning rate μ_{opt} found from genetic algorithm optimization problem.

$\delta_h(k) = Y_k(1 - Y_k)\sum_{(j=1)}^{n} W_k\delta_o(k)$

$W_h(k+1) \leftarrow W_h(k) - \mu_{opt}\delta_h(k)\alpha(channel)$

5. Repeat the above step 4 until the input layer.

6. Update epoch number k=k+1 and go to 1a.

Testing phase:

1. Apply α (testing set) to neural network and pass through all the layer

2. Compute output value Y_k at output layer

Compute the activation h_j of each neuron j in the hidden layer $h_j = \sum_{(i=1)}^{n} \alpha(channel)$

Compute the output Y_k

$Y_k = g(h_j) = 1/(1 + e^{-\beta h_j})$

3. Compute the output error and accuracy

Error = $Y_k - t_k$

Accuracy = number of $(Y_k \neq t_k)$ / Total number of test samples × 100

8) COST EVALUATION

Triad optical sensor plays a significant role in obtaining the real-time spectra, which costs around Rs. 6250/-.The designed Spectroscopic sensor system expense details are as followed in Table 5. The cost of Arduino blackboard is around Rs.1850/-. Moreover, the I2C connector / Qwiic cables help transmit and receive the data. The overall designed equipment cost is very much less expensive around Rs. 8,624/-.

From the Table 7. It is clear that our proposed sensor system design is low-cost compared with other equipment designs. Various types of equipment cost comparison are discussed in Table 7, used for the milk adulteration identifica-

TABLE 7. Cost comparision.

Name of the equipment	Adulteration detection	price/device
Fluoroscence Spectroscopy	Milk Powder	Rs. 13,00000/-
HP4192ALF Impedance analyser	Water	Rs. 2,41,994/-
E-tongue	Water	Rs. 30,000/-
Ultrasonic transmitter and receiver	Water	Rs. 28,128/-
Osmometry	Freezing point, Water	Rs. 4, 12,984/-
Master Eco milk analyser (Milky mist)	Test the freezing point and Solid non-fat content	Rs. 26,500/-
Ksheer tester	Urea, salt, Detergent, Boric acid, Liquid soap, Caustic soda, Hydrogen Peroxide	Rs. 5000/-
Milk Tech kit	Sodium chloride, Sodium bicarbonate	Rs.20,000
Our Proposed Multispectral sensor	Sodium Salicylate, Dextrose, Ammonium Sulphate, hydrogen peroxide	Rs. 8,624/-

tion, they are laboratory-based, involve destructive analysis, and take more time to deliver the results. But our proposed system is portable, non-destructive and capable of providing real-time results.

B. SAMPLE PREPARATION

The commercial raw cow milk samples were utilized for this experiment. Four kinds of adulterants namely Sodium Salicylate, Ammonium sulphate, Dextrose,and Hydrogen Peroxide, were taken for this experimental analysis. Adulterants are added in two steps to get two levels of adulterants mixed samples. The preparation design is given in Table 8. In Table 8. The tested samples and their measured amount of consumption was listed in milliliter and equivalent gram units. The Samples were taken on the sample tube and measured. The first step, the pure raw cow milk sample, weighed about 110ml/110 g taken in a silver bowl, the sensor system was placed on the top of the milk bowl such that the distance between the milk and sensor system is 1 inch away to get the better observations. Now the corresponding spectral readings are observed for the pure raw milk sample. Four adulterant samples were prepared as in Table 8.the first adulterant mixture, namely sodium salicylate measured at a rate of 5*g in the sample tube and stirred in the pure milk sample,then spectral measurements are taken on the same set up which is maintained for pure milk sample. The second sample is prepared by again adding 5ml of sodium salicylate adulterant to get the 10ml adulterated sample. The various milk variety has

TABLE 8. Design of sample preparation in 110ml of pure milk.

Test Cases	Nature of Samples	Adulterant level mixture in pure milk (ml)
Case I	Raw cow Milk Sample	-
Case II	Sodium Salicylate	5*ml and10ml
Case III	Dextrose	5*ml and 10ml
Case IV	Hydrogen Peroxide	5*ml and 10ml
Case V	Ammonium Sulphate	5*ml and 10ml

TABLE 9. Types of samples taken for training.

Types of Milk sample	Milk Fat	Solid Non-Fat
Skimmed milk-No Fat	Not more than 5%,	8.7
Full cream milk- Rich in Protein	6.0	9.0
Toned milk (Mixture of cow/buffalo/bot with fresh skimmed milk)	3.0	8.5
Double Toned Milk	1.5	9.0
Standardized Milk	4.5	8.5
Buffalo Milk	5.0-6.0	9.0
Pure Cow milk	3.0-4.0	8.5-9.0

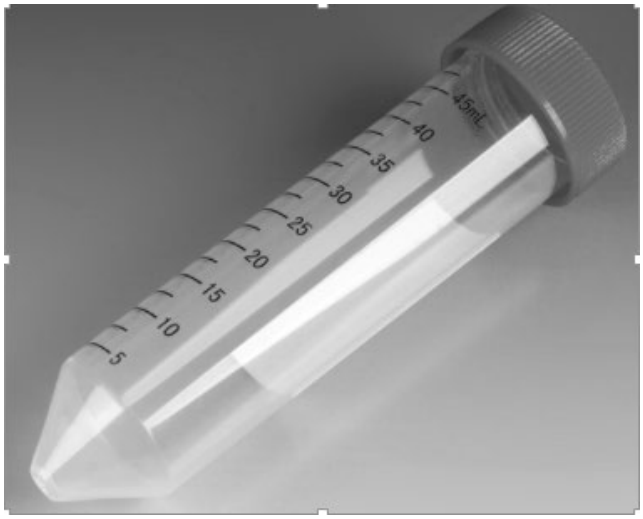


FIGURE 4. Sample Tube.

taken as sample and the sample preparation is also described in Table 9.

In order to provide proper mixing of adulterants on the pure cow milk samples, the measuring unit called sample tube is used for the experimental testing, which has scale units in ml as shown in Fig 4.

C. SPECTRAL DATA COLLECTION

This section delineates the sample spectral data collection for the adulterants and pure milk samples. Samples are

prepared as per earlier discussion. Once the experimental setup connections are completed, the spectral data are collected through the sensor by executing the acquisition program from the Arduino Integrated Developed Environment (IDE) library. The I2C bus (QWIIC interface) is used to communicate data to the Arduino board microcontroller, where the spectral data are transmitted to the Personal computer through a WiFi module. The spectral data were collected for all the five classes of four adulterants and one pure milk sample on the same setup.

The Table 10. delineates the number of samples collected during experimentation. There are four adulterants and one pure raw cow milk sample was made for the experimental observation. Samples obtained from each class are 1800 for pure milk, 1800(Sodium Salicylate), 1800 (Dextrose), 1800(Ammonium Sulphate),1800(Hydrogen Peroxide). There are a total of 16,200 samples tested. All the adulterants / pure milk sample spectral observations are taken for a minimum of 5 minutes. The experiments were made in the presence of ambient light. The multispectral sensor system's sampling period is 150ms at a sampling rate of 6 samples per second.

TABLE 10. Spectral sample collection.

S.no	Cases	Sampling rate /Total no of samples acquired
1.	Pure Milk	6/1800
2.	SodiumSalicylate(5ml) and(10ml)	6/1800
3.	Dextrose(5ml),and (10ml)	6/1800
4.	HydrogenPeroxide(5ml), and (10ml)	6/1800
5.	Ammonium Sulphate(5*ml,and (10ml)	6/1800

IV. RESULTS AND DISCUSSION

A. SPECTRAL DATA ANALYSIS

The acquired spectral data are analyzed for the classifier design. The acquired spectral samples are plotted to explore the relationship among the variables, identifying the group of similar intraclass datasets.

Fig. 5 depicts the photon count Vs. Channel wavelength spectral response curve. It can be understood that the wavelength region from 730nm to 760nm lies on the Ultraviolet region able to distinguish the sample classes for the pure cow milk and 10ml/g added ammonium sulphate in the milk sample. From 810nm to 900nm, the spectral response lies on the same line for both classes, indicating those regions cannot be used for classification.

Fig. 6 shows the NIR spectral response for pure cow milk and 10ml/g mixture of ammonium sulphate in the pure cow milk sample. The NIR wavelength ranges from 410nm to 535nm can discriminate the differences between the sample classes with respect to the photon count. It can be observed from the photon count graph in Fig. 6. that the entire range of

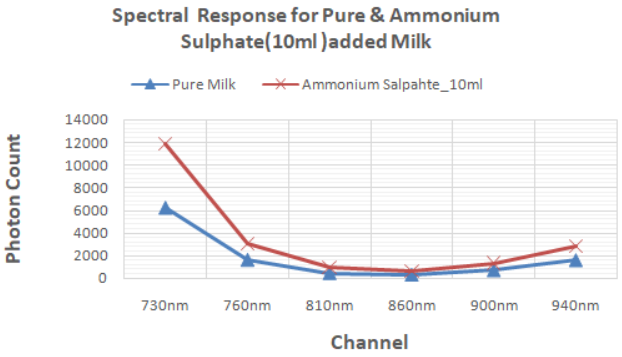


FIGURE 5. Spectral Response (730nm - 940nm) for Pure Cow Milk and 10ml Added Ammonium Sulphate.

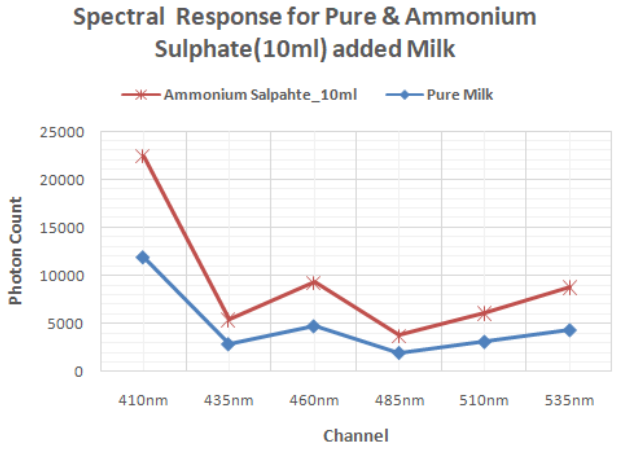


FIGURE 6. Spectral Response(410nm-535nm) for Pure Cow Milk Vs 10ml/g Added Ammonium Sulphate.

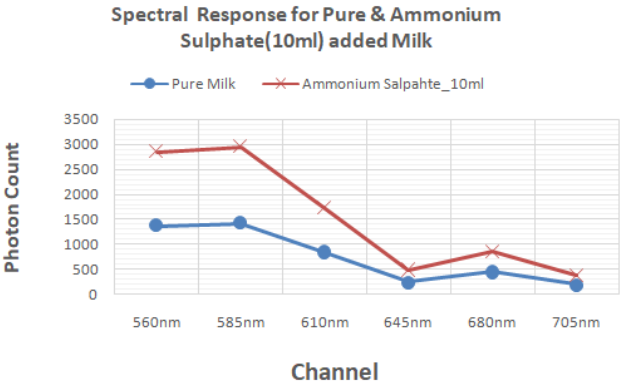


FIGURE 7. Spectral Response (560nm -705nm)for Pure Cow Milk and 10ml/g Added Ammonium Sulphate.

spectral sensor wavelength produces different photon counts, which will enable accurate classification of data.

The Visible and UV region spectral response for pure cow milk and 10ml/g blended ammonium sulphate illustrated in Fig. 7 The spectral response 560nm has higher variation in the absorption between the sample classes and the wavelength 705nm have tiny variation in distinguishing between the

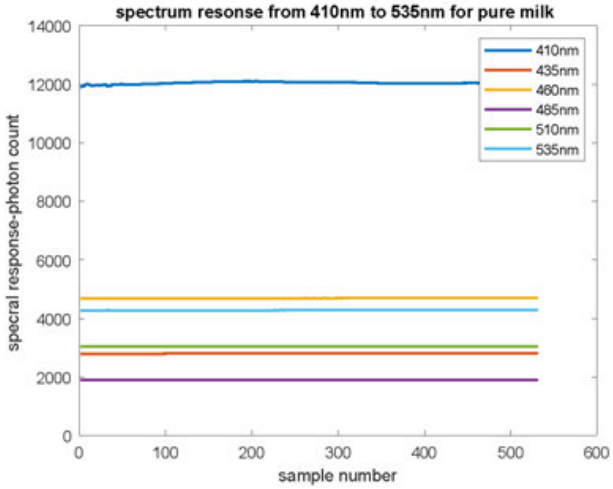


FIGURE 8. Arduino Serial Plotter for Pure Cow Milk Spectral Response.

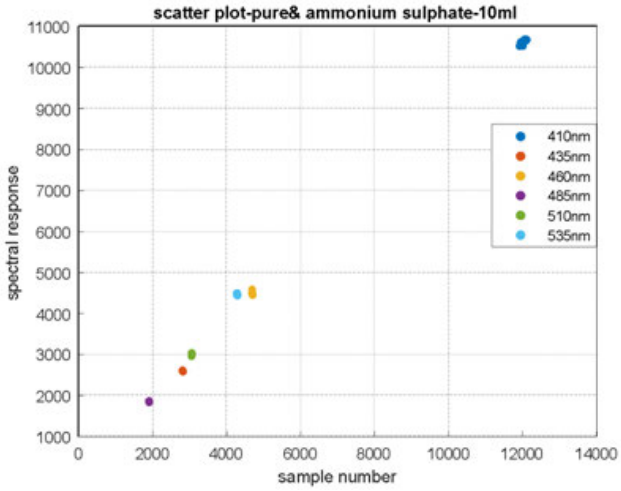


FIGURE 10. Scatter Plot for Pure Cow Milk and 10ml/g Added Ammonium Sulphate.

classes. It is visualized that the pure milk sample class and the adulterated sample class can be distinguished clearly. Fig. 8 visualizes the Pure raw milk NIR spectral response obtained on the serial plotter of Arduino. The wavelength observed in the graph is extracted by the AS72653 Near InfraRed sensor with the wavelength from 410nm to 535nm. We can observe that the 410nm wavelength has the highest peak response for the pure cow milk sample.

Fig. 9 shows deviations between the pure raw milk spectra and a mixture of 10ml ammonium sulphate added to it. Response analysis shows that it is possible to predict the adulterant and also differentiate it from the pure milk sample. It can be noted that the wavelength for this sample deviation is from 410nm to 535nm, which lies in the NIR spectral response of the sensor. Fig. 9. it is observed that the small dotted line present on the 12000 spectral response photon count axis shows the peak response for pure milk sample.

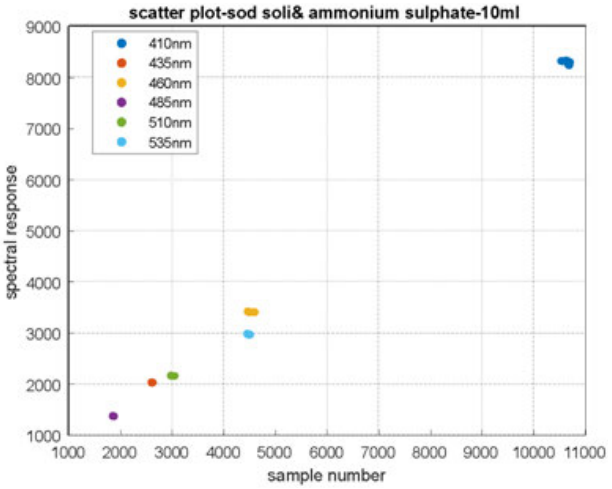


FIGURE 11. Scatter Plot for 10ml/g Added Sodium Salicylate and Ammonium Sulphate.

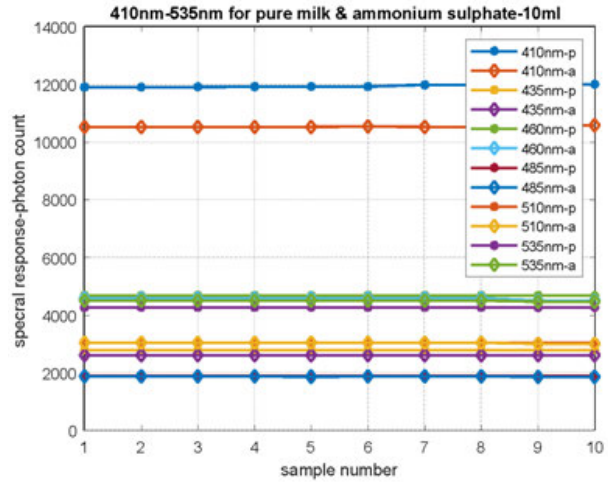


FIGURE 9. Arduino Serial Plotter for Pure Cow Milk and 10ml/g Added Ammonium Sulphate.

For ammonium sulphate mixture, the diamond line appears at below 12000 spectral response photon count. So it is possible to discriminate the variation between the adulterant and non-adulterant samples on the NIR wavelength region.

Fig.10 is made to visualize the classifiability of different adulterants in milk. The scatter diagram in Fig. 10 shows the spectral response of pure raw milk samples and a mixture of 10ml ammonium sulphate on the raw milk samples. It is observed that the spectral wavelength 410nm lies on the diagonal axis but far away from the other sample spectral responses. The adulterant spectral response is clustered at different cluster centers, which proves that this spectral data is possible to successfully classify the adulterant.

The scatter plot Fig. 11 shows the relationship between a 10 ml/g mixture of sodium salicylate and ammonium sulphate in the milk samples. The plotted dots shows a strong relationship between the variables of the samples. Clustering

of the point shows that it is possible to classify them successfully. The CSV data of multi-spectrum is pre-processed to make it suitable for neural network training. One of the pre-processing steps involves cluster analysis, which identifies the similarity between the data points and groups the data points which belong to a category. Intraclass variation was performed based on the certain feature which they share.

Clustering analysis also aids in visualizing and ensure whether it is possible to classify the adulterant using raw spectral samples. K - Means clustering algorithm is used for this purpose. Calculating the centroid point among the dataset and using the group's mean points, the new groups are formulated in the K-means clustering. The cluster assignment graph in Fig. 12 depicts that the dotted points lie between 4000 and 8000 sample numbers corresponding to pure cow milk from a group, cross mark as its centroid. The dotted points lie from 0 to 4000 sample numbers corresponding to the ammonium sulphate, which is making another cluster with a cross mark as its centroid. It can be observed that the data points for each individual sample are closely aligned to its respective centroid. Fig. 12 shows the centroids of the two classes are far away from each other. This proves the highest reliability and accuracy of classification using raw spectral data.

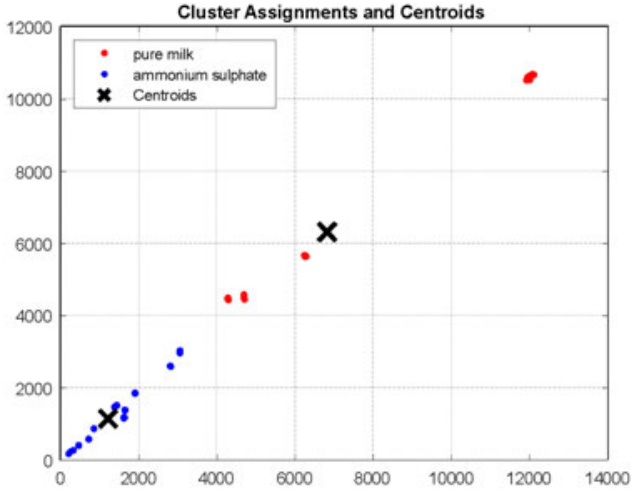


FIGURE 12. K-Means Clustering Analysis for Pure Cow Milk and Ammonium Sulphate.

The cluster assignment in fig. 13. depicts that the dotted points lie from 3000 to 6500 sample numbers corresponding to pure milk samples can form a group with a cross mark as its centroid. The dotted points present from sample number 1 to 2000 corresponding to ammonium sulphate can make another cluster in the same way. Another dotted point on the right top corner belongs to another cluster of sodium salicylate. It can be clearly observed that the data points for each individual sample are closely aligned to their respective centroids. This fig.13 shows that the centroid of the three classes is far away from each other. This proves the maximum reliability and accuracy of classification using raw spectral data.

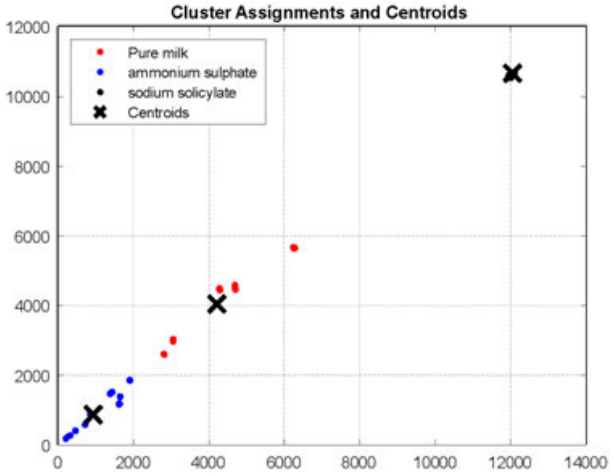


FIGURE 13. K Means Clustering Analysis on Pure Cow Milk, Ammonium Sulphate and Sodium Salicylate.

B. NEURAL NETWORK RESULTS AND DISCUSSION

The 16200 spectral data samples are organized so that the training dataset will be 70% of the model. The validation dataset will be 15% of the dataset from the remaining. The test dataset will be 15% of the unsettled dataset. Two neural network models are developed for the two different classification problems. i) Binary classification model - To classify whether the adulterant present or not.

ii) Multiclass classification model - To classify five-class classification problems namely pure milk, Sodium Salicylate, Dextrose,Ammonium Sulphate and Hydrogen Peroxide.

1) BINARY CLASSIFICATION

The binary classifier neural network model is trained for two-class problems of pure milk and adulterant milk. The adulterant milk class is designated as 0 and the pure milk class is designated as 1.

2) VALIDATION METRICES

The performance metrics of the confusion matrix, cross-entropy error Vs. Epoch validation parameters are presented. Cross entropy is a loss function that optimizes the classification model. The precision, performance measure, sensitivity, the specificity of a neural network model can be validated using the confusion matrix. In the confusion matrix, the column values correspond to actual values and the row corresponds to predicted values that are known. The performance is given in Table 11. Those validation parameters can be calculated from the confusion matrix.

Where, TN- True Negative; TP -True Positive; FN- False Negative; FP - False Positive.

Fig. 14 binary class confusion matrix (after parameter tuning of neural network model) shows very clearly the performance metrics for the adulterant /non- adulterant two classes. The row value represents the predicted values and the column value holds the actual values. The diagonal axis

TABLE 11. Performance metrices.

Validation Parameter	Definition
Sensitivity(Recall)	TP/TP+FN
Specificity	TN/TN+FP
Accuracy /classification rate	(TN+TP)/(TN+TP+FN+FP)

Where, TN- True Negative; TP -True Positive; FN- False Negative; FP - False Positive.

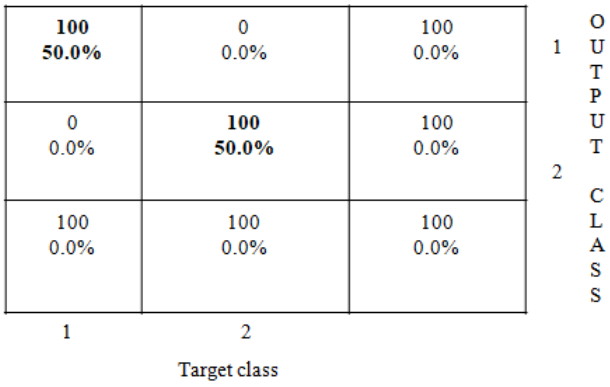


FIGURE 14. Binary Class Confusion Matrix after Hyper Parameter Tuning.

shows that the true negative and true positive labels are classified as an adulterant and non- adulterant, respectively. The true positive,false negative and true negative, false positive of the confusion matrix ratio produces 100% classification accuracy.

The training, testing, validation error analysis is carried out to find optimal epoch. The cross-entropy, gradient and validation fail during the neural network training can be visualized in Fig. 15. the best validation accuracy is achieved on epoch 16. The error rate got minimized and reached zero on the sixteenth epoch during the validation check as it can be a loss function to train the model. The cross-entropy loss goes down as the number of epochs increases, which implies that the model gets trained using the data very well. It minimizes the distance between the predicted and actual value of the sample. The validation and test data curves are too close and so the model is generalized enough. From all the error analysis, we can conclude that the model requires only ten epochs to train and set optimal weight value, which will take only a small amount of training time

Fig.16 shows the confusion matrix obtained for the training, validation and testing dataset. From the confusion matrix, we can prove that the true positive and true negative values are classified correctly and the false positive, false negative values are misclassified correctly, which implies that the sensitivity and specificity of the confusion matrix achieve 100% accuracy.

The error histogram in Fig.17 shows the model is well optimized by taking less error value.

Fig. 18 shows the receiver operating curve for the training, validation and test data set. By selecting the threshold limit, the ROC(Receiver Operating Curve) is calculated for the True positive rate and the false positive rate (i.e.)sensitivity and

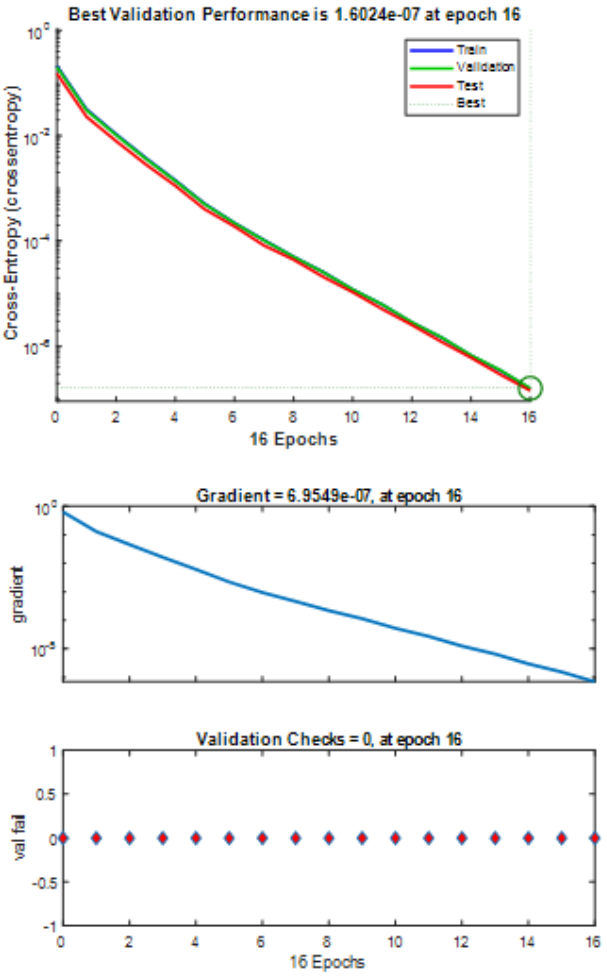


FIGURE 15. Best Validation Performance for Binary Classification for Epoch 16.

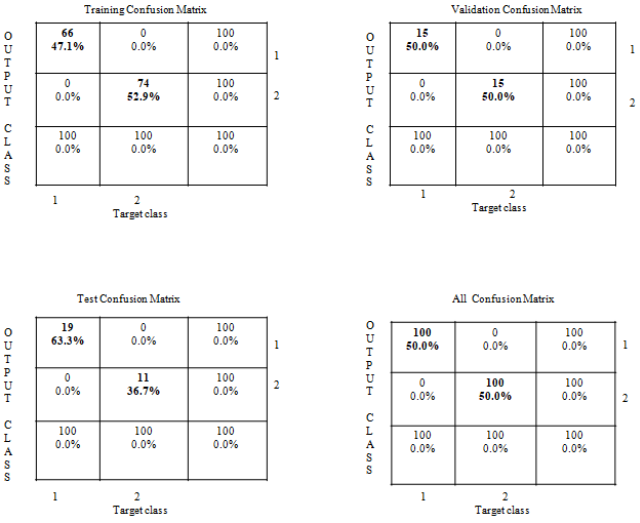


FIGURE 16. Overall Confusion Matrix for Binary Class.

false positive, false negative (i.e.) specificity. It is a graphical plot between the True Positive Rate and False Positive Rate (FPR) at different threshold settings. The ROC curve

FIGURE 17. Error Histogram for Two Class.

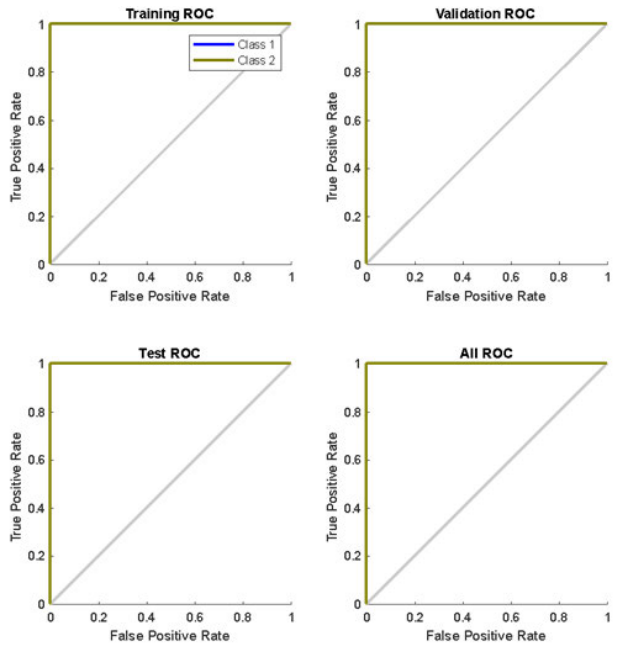


FIGURE 18. Receiver Operating Curve(ROC) for Binary Class.

shows the trade-off between specificity (FPR) and sensitivity (TPR). The curve achieved for two classes lies on the top left corner for all training, validation and test slots. The top left corner lines define that the system works better and its performance is 100% for two-class classification models.

3) FIVE CLASS CLASSIFIER FOR FOUR DIFFERENT ADULTERANT DETECTION

The multiclass classification employed for the dataset having more than two classes. The model of the neural network is developed for five classes. There are five different classes namely pure milk, sodium salicylate, dextrose, hydrogen peroxide and Ammonium sulphate. Each of the classes was trained, validated and tested iteratively with the whole labeled sample dataset and the performance of the model is analyzed.

The confusion matrix in Fig.19 represents the performance metrics of five class classification models. The diagonal matrix value says that the samples of five classes were identified correctly. The sensitivity and specificity for the neural network model are 100%.The performance measure of the five-class neural network model shows 100% classification accuracy.

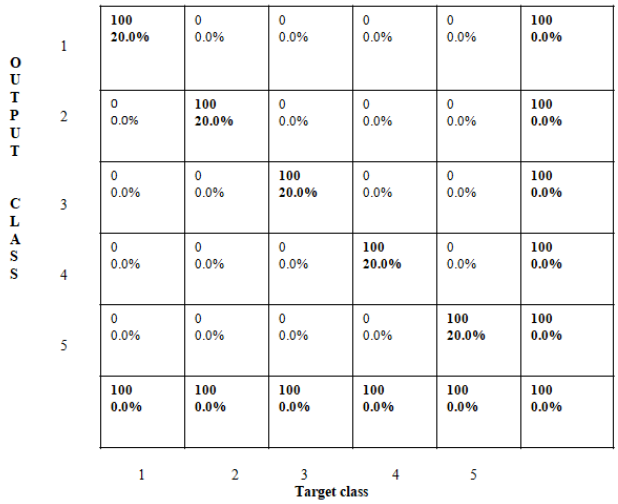


FIGURE 19. Five class Confusion Matrix after Hyper Parameter Tuning.

The training performance of the system model can be validated using cross-entropy.Fig. 20 shows the overall validation check and its corresponding performance result. Once the training step is finished, it is necessary to validate the input samples with the target values. The validation step evaluation analyzes the parameters on which the model produces less error on the assigned validation data set. Here, from the graph, the minimum error rate achieved at $6.4261e^{-07}$ on the 26^{th} epoch during the validation check.Fig.20 shows the best validation performance at $3.7058e^{-07}$ on the 26^{th}epoch during training.

The error histogram of fig. 21 shows the error histogram during training, the model is well optimized by taking less error value.

Fig. 22 describes the performance metrics for the five different classes of adulterant and non- adulterant milk samples. The evaluation was made on training, validation, test data and the system model's overall performance. From the confusion matrix, the sensitivity and specificity of the trained model are calculated. The results acquired between the predicted value and actual value show an accuracy of about 100% after hyperparameter tuning by genetic algorithm.The confusion matrix was obtained for the training, validation and testing dataset. From the confusion matrix, we can prove that the true positive and true negative values are classified correctly and the false positive,false negative values are misclassified correctly, which implies that the sensitivity and specificity of the confusion matrix achieve 100% accuracy.

Fig. 23 visualize ROC graph was plotted for the training, validation test and overall dataset. The curve ensures that the

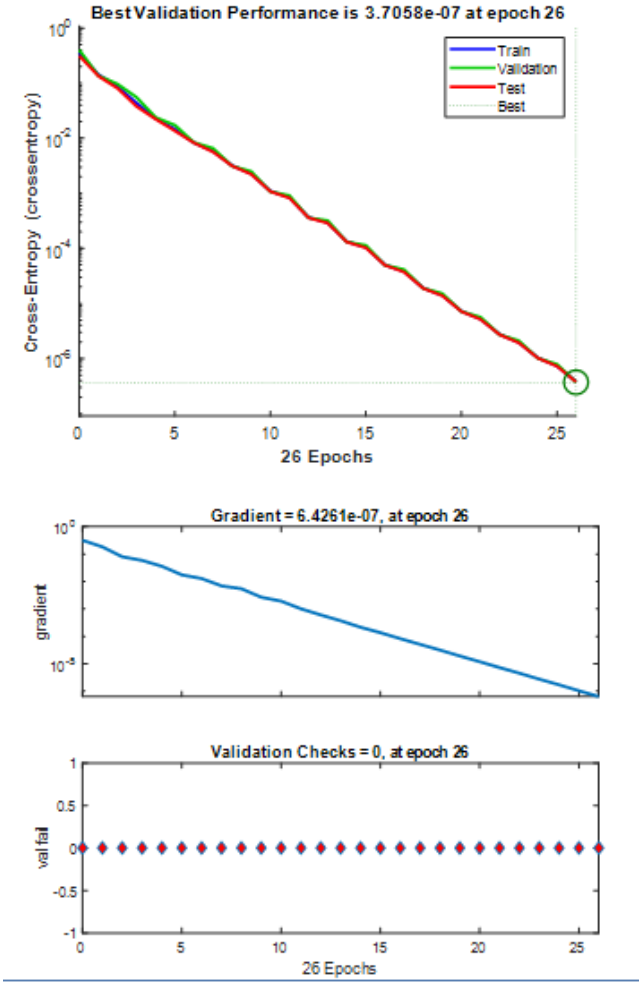


FIGURE 20. Best Validation Performance and Gradient Descent and Validation Check.

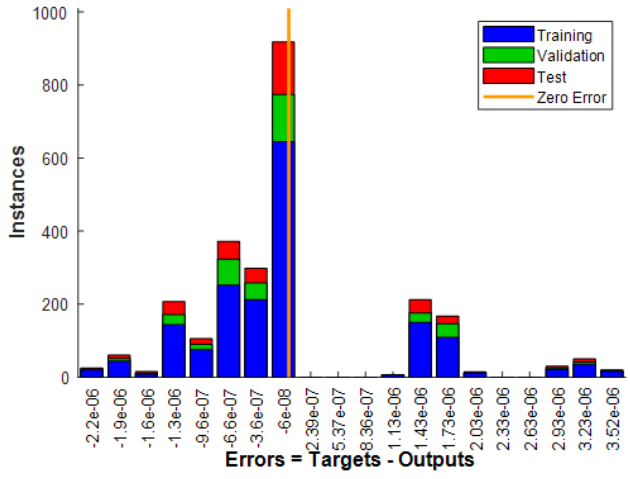


FIGURE 21. Error Histogram for Five Class Performance.

curve is above the diagonal axis for all the cases and merged with the X and Y axis, which proves the highest accuracy of classification results. Some more machine learning algorithms like decision tree, naive Bayes, linear discriminant

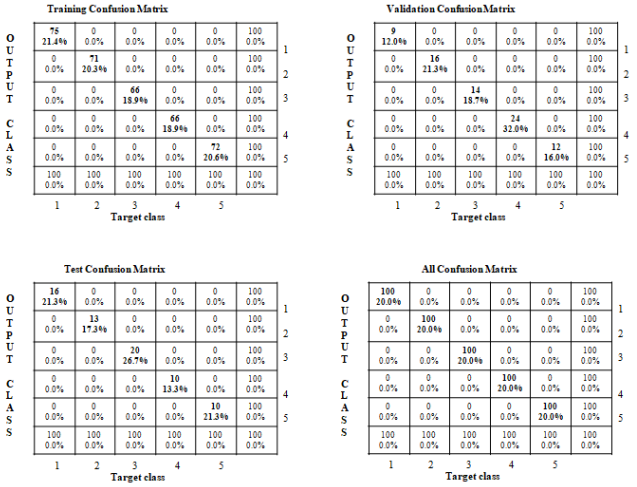


FIGURE 22. Overall Confusion Matrix for Five Class Classification after Hyper Parameter Tuning.

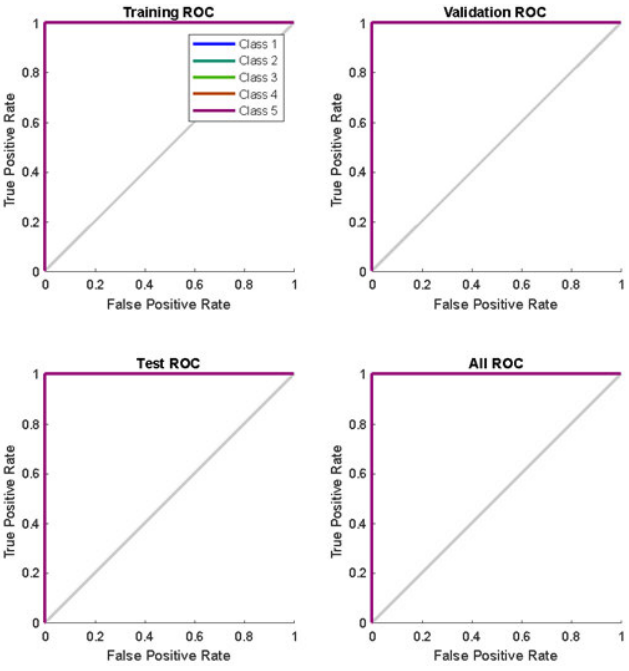


FIGURE 23. Receiver Operating Curve for Five Class Classification.

analysis and support vector machine are applied and the results are tabulated in Table 12. From the table, it is evident that the mentioned machine learning algorithms like Naive Bayes, Linear discriminant analysis, support vector machine, decision tree and neural network model are applied to the spectral data and achieved an accuracy of 90%, 88.1%, 90%, 91.7% and 92.7%respectively. After hyperparameter tuning, we are able to achieve performance improvement from 92.7% to 100%. This is happening because of the strong linear relationship between the adulterant content and the spectral photon response. Since we are capturing 18 different channel spectral responses, it is possible to have a strong linear relationship in three channels for any given adulterant.

TABLE 12. Accuracy performance of machine learning algorithms.

Algorithm Applied	Skimmed milk	Full cream milk	Toned milk	Double Toned Milk	Standardized Milk	Buffalo Milk	Average accuracy
Decision tree	91%	91.9%	92%	91.5%	91.8%	92%	91.7%
Navie Bayes	89.6%	89.8%	90.2%	89.5%	90.3%	90.8%	90%
Linear Discriminant analysis	88.2%	88%	88.3%	88%	88.1%	88.3%	88.1%
Support Vector Machine	90%	90.1%	90.3%	89.8%	90%	89.9%	90%
Neural Network without hyperparameter	92.4%	93%	93.2%	92.1%	92.7%	92.8%	92.7%
Neural Network with hyperparameter	100%	100%	100%	100%	100%	100%	100%

The response time of the proposed system model is calculated. It is found that the system takes around 0.4 seconds to analyse the data and classify it.But there is 1.5 second overhead for collecting data (minimum of 10 samples) because of sensor system sampling rate is 150ms. So with this overhead the overall response time of the system will be 1.9 second. This 1.9 second response time is more than enough to produce real time detection of adulterant on the field.

The proposed model performance is compared to the similar literature works and tabulated in Table 13.

In [29] the milk adulterant testing using the Fourier Transform spectroscopy spectral analysis was carried out for five adulterants. Though it is non- non-destructive and rapid method, the experiments were tested in the lab set up and the equipment was expensive, non- portable and of higher wavelength region. The results were obtained using deep and ensemble learner methodology and achieved the classification accuracy of about 98.7 In [34] FTIR-ATR(Fourier Transform InfraRed -Attenuated Total Reflectance) spectroscopic analysis identified the presence of five adulterants in milk. This method achieves classification accuracy of about 76.6%. This technique, applied equipment is expensive and it was not portable. Real-time analysis cannot be carried out.

In [36] the presence of hydrogen peroxide was identified and analysed in milk. Systematic evaluation is done through liquid chromatography with a diode array detection. The correlation coefficient achieved is 99% with the detectable limit of 0.28mg/L presence of adulterant in milk. In comparison with the listed similar literature work, our proposed system is superior in the following aspects,

- Other methods involve single spectral wavelength, our system works at multispectral wavelength.

TABLE 13. Comparision result.

Literature work employed	Type of spectroscopy	Methodology adapted	Adulterant Detected	Accuracy achieved
H.A.Neto,et al.[29]	FTIR	Deep and ensemble learners	Sucrose, soluble starch (Amylose and amylopectin), sodium bicarbonate, hydrogen peroxide and formaldehyde	98.76%
F.F.D.Hora, etal.[34]	FTIR-ATR(Fourier Transform infrared attenuated total Reflectance)	Multiple Linear Regression	Sucrose,urea, sodium hydroxide,hydrogen peroxide,sodium bi- carbonate	76.6%.
A.S.Ivanova, etal.[36]	Diode array Detector	High performance Liquid Chromatography	Hydrogen Peroxide	Correlation Coeff=99%
Proposed work	Multispectral sensor system	i,Binary class Neural Network-classification ii,Five class Neural Network classification	Sodium Salicylate, Dextrose, Ammonium Sulphate,hydrogen peroxide	100%(after hyper parameter tuning)

- Navie bayes, Linear discriminant analysis, support vector machine, decision tree and neural network algorithms are applied to the spectral data and accuracy of 90%, 88.1%, 90%, 91.7% and 92.7% are achieved.
- After hyperparameter tuning of neural network highest accuracy of 100% is achieved for identifying adulterants. whereas other techniques are able to achieves only less than 100%.
- The proposed system able to detect five different adulterants with 100% of accuracy whereas, mostly other methods able to detect one or more adulterants.even [29] able to detect six adulterant, whose accuracy only 98.76%
- Our developed system is portable, rapid, non-destructive and less expensive when compared to other systems.

C. WEB INTERFACE

Our proposed sensor system design to provide a web interface application for real-time visualization of adulterant detected results. With this web interface, the milk adulteration results can be accessed from anywhere at any time. The neural network classifier output is communicated to a web page for real-time updates. The following methodology describes the workflow of the web interface.

1) METHODOLOGY OF WEB INTERFACE

IoT combined spectral sensor system design makes the user to know the adulteration detection in real-time without delay. The webpage is designed such a way only authorized user

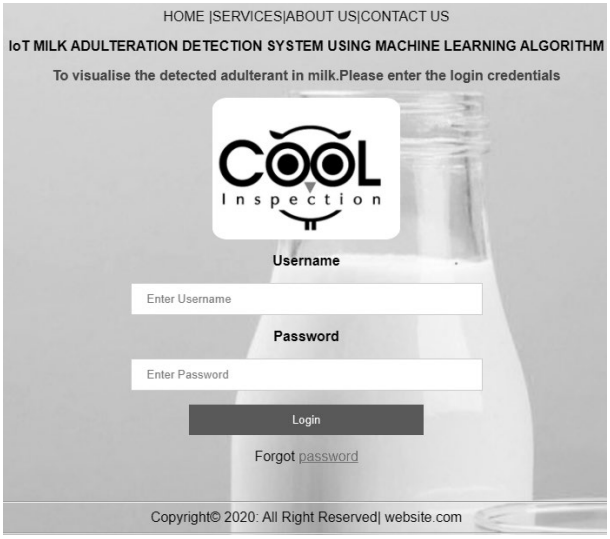
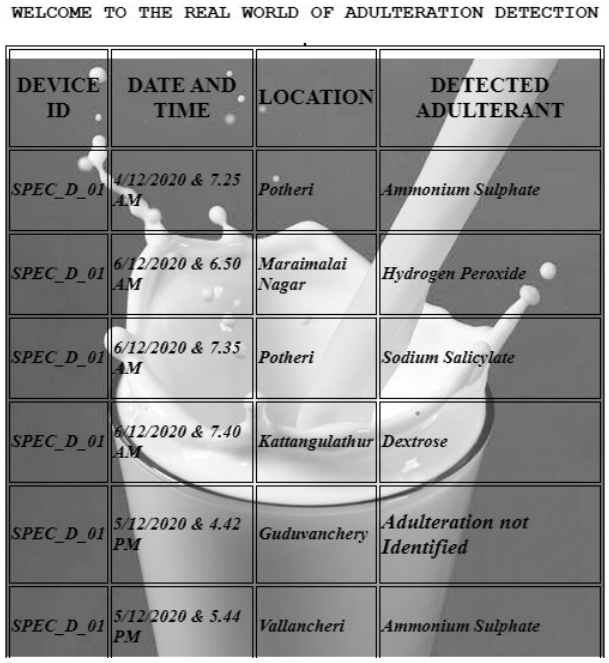


FIGURE 24. Login Page of Web Interface.



DEVICE ID	DATE AND TIME	LOCATION	DETECTED ADULTERANT
SPEC_D_01	4/12/2020 & 7.25 AM	Potheri	Ammonium Sulphate
SPEC_D_01	6/12/2020 & 6.50 AM	Maraimalai Nagar	Hydrogen Peroxide
SPEC_D_01	6/12/2020 & 7.35 AM	Potheri	Sodium Salicylate
SPEC_D_01	6/12/2020 & 7.40 AM	Kattangulathur	Dextrose
SPEC_D_01	5/12/2020 & 4.42 PM	Guduvanchery	Adulteration not Identified
SPEC_D_01	5/12/2020 & 5.44 PM	Vallancheri	Ammonium Sulphate

FIGURE 25. Web Display of Detected Adulterant.

can access. The webpage is designed using an HTML tag document. The home page consists of login details for authorization purposes. Once the consent person entered, a valid username and access will be granted. After logged in, the user can access device information, location of sample collection, date and timings during which the adulterant is detected and the detected adulterant result. Fig. 24. shows the home page details and Fig. 25. provides detailed information of detected adulterant

V. CONCLUSION

Milk adulteration is one of the major problems around the globe. Although, different spectroscopic methodologies entailed for the identification of adulterants in milk. The proposed model with Artificial Intelligence presents the four adulterant identification namely Hydrogen peroxide, sodium sulphate, dextrose Ammonium sulphate and pure milk sample by employing the designed multispectral sensor model. This developed sensor model extracts the sample's spectral information in a non-destructive way. Various machine learning algorithms like Naive Bayes, Linear discriminant analysis, support vector machine, decision tree and neural network model are applied to the spectral data and accuracy of 90%, 88.1%, 90%, 91.7% and 92.7% are achieved. Optimal parameter selection/parameter tuning of the neural network is carried out using the Genetic algorithm framework. By applying optimal parameter setting, the neural network performance is improved from 92.7% to 100%. A dedicated webpage with an IoT application is designed to enable the user/ authorized people to visualize the detected adulterant from anywhere. The proposed system has the following limitations such that

i) If the distance between the sensor and milk sample is greater than one inch the model performance may be degraded.

ii) The model performance will be degraded, if the distance between sensor and the milk sample is greater than one inch.

iii) There are few more adulterants which will be mixed in the milk are not explored.

REFERENCES

[1] *Annual Consumption of Fluid Cow Milk Worldwide in 2020*. Accessed: Dec. 1, 2020. [Online]. Available: https://www.statista.com/statistics/272003/global-annual-consumption-of-milk-by-region/

[2] M. Guetouache, B. Guessas, and S. Medjekal, "Composition and nutritional value of raw milk," *J. Issues Biol. Sci. Pharm. Res.*, vol. 2, no. 10, pp. 115–122, Dec. 2014.

[3] S. K. Chang and Y. Zhang, "Protein analysis," in *Food Analysis*. Cham, Switzerland: Springer, 2017, pp. 315–331.

[4] A. Ravindran, F. P. Nesamani, and D. Nirmal, "A study on the use of spectroscopic techniques to identify food adulteration," in *Proc. Int. Conf. Circuits Syst. Digit. Enterprise Technol. (ICCSDET)*, Dec. 2018, pp. 1–6.

[5] V. K. Verma, P. Mustajab, and A. Sadat, "Determination of adulteration in milk using ultrasonic technique," in *Proc. Int. Conf. Electr., Electron. Comput. Eng. (UPCON)*, Nov. 2019, pp. 1–3.

[6] F. Trimboli, N. Costanzo, V. Lopreiato, C. Ceniti, V. M. Morittu, A. Spina, and D. Britti, "Detection of buffalo milk adulteration with cow milk by capillary electrophoresis analysis," *J. Dairy Sci.*, vol. 102, no. 7, pp. 5962–5970, Jul. 2019.

[7] M. Sude and K. Ghodinde, "Electrical impedance sensor for real-time detection of urea and starch in milk," in *Proc. 3rd Int. Conf. Trends Electron. Informat. (ICOEI)*, Apr. 2019, pp. 431–434.

[8] C. Das, S. Chakraborty, A. Karmakar, and S. Chattopadhyay, "On-chip detection and quantification of soap as an adulterant in milk employing electrical impedance spectroscopy," in *Proc. Int. Symp. Devices, Circuits Syst. (ISDCS)*, Mar. 2018, pp. 1–4.

[9] P. Nikolaou, E. Deskoulidis, E. Topoglidis, A. T. Kakoulidou, and F. Tsopelas, "Application of chemometrics for detection and modeling of adulteration of fresh cow milk with reconstituted skim milk powder using voltammetric fingerpriting on a graphite/SiO_2 hybrid electrode," *Talanta*, vol. 206, Jan. 2020, Art. no. 120223.

[10] M. Chakraborty and K. Biswas, "Limit of detection for five common adulterants in milk: A study with different fat percent," *IEEE Sensors J.*, vol. 18, no. 6, pp. 2395–2403, Mar. 2018.

[11] L. de Souza Ribeiro, F. A. Gentilin, J. A. de França, A. L. D. S. M. Felício, and M. B. D. M. Frana, "Development of a hardware platform for detection of milk adulteration based on near-infrared diffuse reflection," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 7, pp. 1698–1706, Mar. 2016.

[12] P. W. Hansen and S. E. Holroyd, "Development and application of Fourier transform infrared spectroscopy for detection of milk adulteration in practice," *Int. J. Dairy Technol.*, vol. 72, no. 3, pp. 321–331, Apr. 2019.

[13] S. Kasemsumran, W. Thanapase, and A. Kiatsoonthon, "Feasibility of near-infrared spectroscopy to detect and to quantify adulterants in cow milk," *Anal. Sci.*, vol. 23, no. 7, pp. 907–910, 2007.
[14] M. De Marchi, M. Penasa, A. Zidi, and C. L. Manuelian, "Invited review: Use of infrared technologies for the assessment of dairy products Applications and perspectives," *J. Dairy Sci.*, vol. 101, no. 12, pp. 10589–10604, Dec. 2018.
[15] P. M. Santos, E. R. Pereira-Filho, and L. E. Rodriguez-Saona, "Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis," *Food Chem.*, vol. 138, no. 1, pp. 19–24, May 2013.
[16] C. de Souza Gondim, R. G. Junqueira, S. V. C. de Souza, I. Ruisánchez, and M. P. Callao, "Detection of several common adulterants in raw milk by MID-infrared spectroscopy and one-class and multi-class multivariate strategies," *Food Chem.*, vol. 230, pp. 7–68, Sep. 2017.
[17] L. da Silva Dias, J. C. da Silva, Jr., A. L. de Souza Maudeira Felicio, and J. A. de Franca, "A NIR photometer prototype with integrating sphere for the detection of added water in raw milk," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 12, pp. 2812–2819, Dec. 2018.
[18] M. M. Bergana, K. M. Adams, J. Harnly, J. C. Moore, and Z. Xie, "Non-targeted detection of milk powder adulteration by 1H NMR spectroscopy and conformity index analysis," *J. Food Composition Anal.*, vol. 78, pp. 49–58, May 2019.
[19] R. R. R. Gonzalez, L. Cobuccio, and T. Delatour, "Reconstitution followed by non-targeted mid-infrared analysis as a workable and cost-effective solution to overcome the blending duality in milk powder adulteration detection," *Food Chem.*, vol. 295, pp. 42–50, Oct. 2019.
[20] M. Bassbasi, M. De Luca, S. Souhassou, A. Hirri, M. Berkani, F. Kzaiber, G. Loele, G. Ragno, and A. Oussama, "Determination of milk adulteration by sucrose using FT-MIR spectroscopy and chemometrics methods," *Agricult. Sci. Res. J.*, vol. 4, no. 11, pp. 175–180, Nov. 2014.
[21] L. J. Ni, L. G. Zhang, M. Y. Tang, Z. B. Xue, X. Zhang, X. Gu, and S. Huang, "Discrimination of adulteration cow milk by improved ν-support vector machines (ν-SVM) and near infrared spectroscopy," in *Proc. 8th Int. Conf. Natural Comput. (ICNC)*, May 2012, pp. 69–73.
[22] F. Mabood, J. Hussain, A. N. Moo, S. A. Gilani, S. Farooq, Z. Naureen, F. Jabeen, M. Ahmed, Z. Hussain, and A. A. Harrasi, "Detection and quantification of formalin adulteration in cow milk using near infrared spectroscopy combined with multivariate analysis," *Adv. Dairy Res.*, vol. 5, no. 1, p. 167, 2017.
[23] F. Mabood, L. Ali, R. Boque, G. Abbas, F. Jabeen, Q. M. I. Haq, J. Hussain, A. M. Hamaed, Z. Naureen, M. Al-Nabhani, M. Z. Khan, A. Khan, and A. Al-Harrasi, "Robust Fourier transformed infrared spectroscopy coupled with multivariate methods for detection and quantification of urea adulteration in fresh milk samples," *Food Sci. Nutrition*, vol. 8, no. 10, pp. 5249–5258, Oct. 2020.
[24] K. Kishor and R. Thakur, "Analysis of milk adulteration using mid-IR spectroscopy," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 3, no. 10, pp. 5890–5895, 2015.
[25] R. Gupta, S. Wadhwa, A. Mathur, and A. K. Dubey, "Designing of a microelectrode sensor-based label free milk adulteration testing system," *IEEE Sensors J.*, vol. 17, no. 18, pp. 6050–6055, Sep. 2017.
[26] K. M. Khan, H. Krishna, S. K. Majumder, and P. K. Gupta, "Detection of urea adulteration in milk using near-infrared Raman spectroscopy," *Food Anal. Methods*, vol. 8, no. 1, pp. 93–102, Jan. 2015.
[27] B. Balan, A. S. Dhaulaniya, R. Jamwal, Amit, K. K. Sodhi, S. Kelly, A. Cannavan, and D. K. Singh, "Application of attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy coupled with chemometrics for detection and quantification of formalin in cow milk," *Vibrational Spectrosc.*, vol. 107, Mar. 2020, Art. no. 103033.
[28] P. Bertemes-Filho, L. H. Negri, and A. S. Paterno, "Detection of bovine milk adulterants using bioimpedance measurements and artificial neural network," in *Proc. 5th Eur. Conf. Int. Fed. Med. Biol. Eng. (IFMBE)*. Berlin, Germany: Springer, 2011, pp. 1275–1278.
[29] H. A. Neto, W. L. F. Tavares, D. C. S. Z. Ribeiro, R. C. O. Alves, L. M. Fonseca, and S. V. A. Campos, "On the utilization of deep and ensemble learning to detect milk adulteration," *BioData Mining*, vol. 12, no. 1, p. 13, Dec. 2019.
[30] F. Mabood, F. Jabeen, M. Ahmed, J. Hussain, S. A. A. Al Mashaykhi, Z. M. A. Al Rubaiey, S. Farooq, R. Boqué, L. Ali, Z. Hussain, A. Al-Harrasi, A. L. Khan, Z. Naureen, M. Idrees, and S. Manzoor, "Development of new NIR-spectroscopy method combined with multivariate analysis for detection of adulteration in camel milk with goat milk," *Food Chem.*, vol. 221, pp. 746–750, Apr. 2017.
[31] R. Ullah, S. Khan, H. Ali, and M. Bilal, "Potentiality of using front face fluorescence spectroscopy for quantitative analysis of cow milk adulteration in buffalo milk," *Spectrochimica Acta A, Mol. Biomolecular Spectrosc.*, vol. 225, Jan. 2020, Art. no. 117518.
[32] E. V. D. S. Pereira, D. D. D. S. Fernandes, M. C. U. de Araújo, P. H. G. D. Diniz, and M. I. S. Maciel, "Simultaneous determination of goat milk adulteration with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms," *LWT*, vol. 127, Jun. 2020, Art. no. 109427.
[33] L. A. Dias, A. M. Peres, A. C. A. Veloso, F. S. Reis, M. Vilas-Boas, and A. A. S. C. Machado, "An electronic tongue taste evaluation: Identification of goat milk adulteration with bovine milk," *Sens. Actuators B, Chem.*, vol. 136, no. 1, pp. 209–217, Feb. 2009.
[34] D. Conceição, B.-H. Gonçalves, F. da Hora, A. Faleiro, L. Santos, and S. Ferrão, "Use of FTIR-ATR spectroscopy combined with multivariate analysis as a screening tool to identify adulterants in raw milk," *J. Brazilian Chem. Soc.*, vol. 30, no. 4, pp. 780–785, Apr. 2018.
[35] H. Han, Z. Wang, C. Li, Z. Ma, Z. Yang, and X. Ma, "Purity detection of goat milk based on electronic tongue and improved artificial fish swarm optimized extreme learning machine," *IFAC-PapersOnLine*, vol. 52, no. 30, pp. 391–396, 2019.
[36] A. S. Ivanova, A. D. Merkuleva, S. V. Andreev, and K. A. Sakharov, "Method for determination of hydrogen peroxide in adulterated milk using high performance liquid chromatography," *Food Chem.*, vol. 283, pp. 431–436, Jun. 2019.
[37] S. Moncayo, S. Manzoor, J. D. Rosales, J. Anzano, and J. O. Caceres, "Qualitative and quantitative analysis of milk for the detection of adulteration by laser induced breakdown spectroscopy (LIBS)," *Food Chem.*, vol. 232, pp. 322–328, Oct. 2017.
[38] L.-G. Zhang, X. Zhang, L.-J. Ni, Z.-B. Xue, X. Gu, and S.-X. Huang, "Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy," *Food Chem.*, vol. 145, pp. 342–348, Feb. 2014.
[39] L. Feng, S. Zhu, S. Chen, Y. Bao, and Y. He, "Combining Fourier transform mid-infrared spectroscopy with chemometric methods to detect adulterations in milk powder," *Sensors*, vol. 19, no. 13, p. 2934, Jul. 2019.
[40] G. Bilge, B. Sezer, K. E. Eseller, H. Berberoglu, A. Topcu, and I. H. Boyaci, "Determination of whey adulteration in milk powder by using laser induced breakdown spectroscopy," *Food Chem.*, vol. 212, pp. 183–188, Dec. 2016.
[41] *SparkFun Triad Spectroscopy Sensor–AS7265x (Qwiic)*. Accessed: Mar. 1, 2019. [Online]. Available: https://www.sparkfun.com/products/15050
[42] (May 23, 2018). *AS7265x Smart 18-Channel VIS to NIR Spectral_ID 3-Sensor Chipset With Electronic Shutter [v1-5]*. Accessed: Mar. 1, 2020. [Online]. Available: https://cdn.sparkfun.com/assets/c/2/9/0/a/AS7265x_Datasheet.pdf

N. SOWMYA received the B.E. degree in ECE from Anna University, in 2008, and the master's degree in embedded systems from the SRM Institute of Science and Technology, Chennai, in 2014, where she is currently pursuing the Ph.D. degree with the Department of ECE.

VIJAYAKUMAR PONNUSAMY (Senior Member, IEEE) received the B.E. degree in ECE from Madras University, in 2000, the master's degree in applied electronic from the College of Engineering, Guindy, in 2006, and the Ph.D. degree in applied machine learning in wireless communication (cognitive radio) from SRM IST, Chennai, India, in 2018. He is currently working as an Associate Professor with the ECE Department, SRM IST. He is also a Certified IoT Specialist and the Data Scientist. His current research interests include machine and deep learning, the IoT based intelligent system design, blockchain technology, and cognitive radio networks. He was a recipient of the NI India Academic Award for excellence in research, in 2015.