An Intelligent IoT and Machine Learning Framework for Real-Time Food Spoilage Detection and Quality Monitoring

Anmol Keswani[1], Roshni Khatri[2], Sunny Nahar[3]

[1]*Post Graduate Student, Dept of Masters of Computer Application, Vivekanand Education Society's Institute Of Technology Chembur, Mumbai*

[2]*Post Graduate Student, Dept of Masters of Computer Application, Vivekanand Education Society's Institute Of Technology Chembur, Mumbai*

[3]*Professor, Dept of Masters of Computer Application, Vivekanand Education Society's Institute Of Technology Chembur, Mumbai*

---***---

Abstract-*Food spoilage is a problem that causes a lot of financial losses and health risks all over the world. Most food monitoring systems that use the internet of things rely on rules that do not work well for different types of food and changing environments. This study suggests a way to monitor food quality that uses a device with many sensors and a machine learning system to predict when food will spoil in real time. The system uses a computer called an ESP32 that is connected to many sensors like MQ3, MQ4 and MQ135 gas sensors, a DHT22 temperature and humidity sensor, an LDR light sensor and a pH sensor. The sensor data is sent over Wi-Fi to a platform called AWS IoT Core, where a special machine learning model is used to classify how fresh the food is and predict how the quality will change over time.The system is very good at classifying food freshness with an accuracy of 94.3% for food like dairy, fruits and packaged goods. It can send alerts in under 2 seconds. This system is better, than systems that just use simple rules because it can adapt to different situations it has fewer false alerts and it can predict how long the food will last, which helps manage the supply chain and reduce food waste.*

Key Words: IoT, food quality monitoring, machine learning, Random Forest, LSTM, ESP32, food spoilage detection, smart supply chain, AWS IoT Core.

1.INTRODUCTION

There are significant issues with the food safety and quality monitoring processes throughout the modern global food supply chain. The Food and Agriculture Organization of the United Nations reports that around one-third of food produced for human consumption is lost or wasted each year, much of it due to insufficient monitoring of the environmental conditions during storage and/or transportation [1]. Environmental variable fluctuations (e.g., temperature, humidity, ethanol concentration in the air, and exposure to light) can significantly increase the rate of spoilage due to microbial activity, oxidation, and enzymatic reactions; all of these are measured by physical parameters.
Most warehouse and distribution facilities use manual inspections on a periodic basis to monitor food, which is labour-intensive and prone to human error. While some sensor-based IoT systems have been developed (to automate environmental monitoring), most of these deployed systems apply static rules (threshold) to initiate alert events. As a result, binary (yes/no rule-based) methods fail to address the dynamic and nonlinear degradation processes of food and do not adapt to the inherent variability among food types.

This paper presents an intelligent solution for monitoring food quality over time. Machine learning (ML) is an innovative way to significantly change food industry processes as we currently know them through the use of sensor devices and ML algorithms trained on historical datasets of sensor outputs labeled with spoilage factors. The ability to affordably develop predictive models using various types of data, such as historical data from different locations combined in one dataset, allows for much more accurate prediction of food quality deterioration than the previous method of using fixed thresholds alone.

In particular, the use of recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) networks is beneficial for training models to predict food quality over time because of their inherent ability to model time-series data such as that from sensors. Additionally, these types of RNNs will enable predictive shelflife determination and allow the food industry to move from being reactive to proactive through real-time alerting capabilities that result from collected sensor and ML data.

This research presents an "end-to-end" intelligent framework that consists of:

• A heterogeneous sensor device (ESP32) with multiple types of sensors collecting and transmitting data relating to gas composition, temperature, humidity, light intensity, and pH level in real-time to a cloudbased platform for building and maintaining ML predictive models.

• A hybrid ML pipeline that uses both a Random Forest classifier to classify the current freshness and quality of food and an LSTM model that uses temporal information to predict spoilage trend of food products.

• Integration with AWS IoT Core to provide data storage capabilities, and Firebase to enable real-time alerting by means of mobile push notifications, and a web-based dashboard for ease of remote monitoring.

• Validation of the predictive classification accuracy of the models developed on the hybrid ML pipeline using three different food categories with an accuracy rate of 94.3% and an average response time of less than 2 seconds.

2.LITERATURE SURVEY

Numerous studies exist in the area of IoT based food monitoring; however, there is limited use of predictive machine learning in the existing deployments.

Chanthini, et al., created a wireless sensor network based perishable food quality-monitoring system using Raspberry Pi as the primary processing device. The moisture, gas, temperature, and humidity sensors were integrated into the system, and data was displayed via the Xively IoT Cloud platform. Although sensor integration was successful, the lack of machine learning capability prevented prediction from occurring as all parameters were based on fixed thresholds [1].

Venkatesh, et al. developed a Bluetooth low energy-based food monitoring system using an nRF52832 Cortex-M4F microprocessor, and VOC and DHT11 sensors were programmed to send data to the ThingSpeak cloud. Although the use of Bluetooth low energy reduces battery usage, a single gas sensor lowers the ability to identify everything that could affect the food product in question; therefore, there is no machine learning component in the development of this product [2].

Divya Bharathi, et al. created a food monitoring system using an Arduino Mega 328 microcontroller with the goal of detecting the bacterial contamination of food products through temperature, odour, and ph sensing; in addition to using an off the shelf biosensor. Data was displayed on an LCD, however, there was no cloud connectivity or remote monitoring capabilities, which would hinder its ability for large scale practical use [3].

Prasaath proposed an Arduino UNO food spoilage monitoring project using DHT-11, MQ3, and LDR sensors displayed on Freeboard.io. This system provides a great place to start, but the number of sensors used provides insufficient information to allow for detection of multiple food types which may spoil [4].

Kaushal et al. extended the project by developing a system based on an ATmega microcontroller using a comprehensive array of sensors including Temp, LDR, gas, humidity, moisture, IR and a GSM module for communication with the supply chain. When sensor parameters were out of the acceptable/threshold range, a buzzer would be activated. This work established the importance of integrating the food supply chain; however, ML was not used for adaptive prediction [5].

Borawake et al. took the project one step further by using Arduino Uno with gas and pH sensors connected to a simple machine learning model that determines fresh food and spoiled food. These models are not capable of temporal reasoning, and do not have multiple classes available by which to differentiate various food types [11].

Nguyen et al. recently developed a battery free, NFC based food monitoring unit with pH and CO2 measuring capabilities as the first step toward implementing a more advanced hardware solution. However, this system was built with such a limited energy model that it placed significant restrictions on the processing power of the system, making on device integration of machine learning nearly impossible at this time[7].

The comparative analysis in Table I illustrates the positioning of the proposed system relative to prior art, highlighting the gap this work addresses namely, the absence of adaptive, multi-class, temporally-aware ML-driven monitoring.

Table I shows how the new system is positioned compared with previous work and identifies a gap in technology that this research works to fill: There is no adaptive, multi-class, temporally-aware ML monitored system.

Table -1: Comparative Analysis of Existing Food Monitoring Systems

Reference	Microcontroller	Sensors Used	Cloud/Platform	ML Integration	Limitation
Chanthini et al. [1]	Raspberry Pi	Humidity, Temp, Gas, Moisture	Xively IoT	No	No ML; rule-based only
Venkatesh et al. [2]	nRF52832	VOC, DHT11	ThingSpeak	No	Limited gas detection
Divya et al. [3]	Arduino Mega 328	Temp, Odour, pH, Bio	LCD	No	No IoT connectivity
Hai Prasaath [4]	Arduino UNO	DHT-11, MQ3, LDR	Freeboard.io	No	Limited sensor set
Borawake et al. [11]	Arduino Uno	Gas, pH, Moisture	Wi-Fi Module	Yes (basic ML)	No real-time alerts
Proposed System	ESP32	MQ series, DHT22, LDR, pH	AWS IoT + Firebase	Yes (RF + LSTM)	—

3. PROPOSED SYSTEM ARCHITECTURE

The proposed system is organised into three functional layers: the sensor acquisition layer, the cloud processing layer, and the user interface layer.

A. The Layer That Acquires Sensors The ESP32 microcontroller can be considered the primary component of a sensor node; it has a dual-core Xtensa LX6 processor and has built-in Wi-Fi, Bluetooth, and 34 programmable GPIO pins. The following sensors are being interfaced to the ESP32:

• MQ-3 (ethanol/alcohol sensor) – Measures the presence of VOCs (volatile organic compounds) produced by fermentation and spoiling, especially with dairy and fruit.

• MQ-4 (methane sensor) – Measures the amount of methane gas generated during anaerobic decomposition of

• LDR (light dependent resistor) – Measures the intensity of light exposure, as it can enhance the lipid oxidation and photo-degradation of certain foods.

• pH sensor (SEN0161) - Measures pH of a food product's liquid, where applicable, such as in beverages and other liquid food products

• MQ-135 (air quality/ammonia sensor) – Provides information on ammonia, benzene and sulfide concentrations as they relate to advanced protein decay.

Sensor reading will be **taken** every 10 seconds and locally buffered on the SPIFFS file system of the ESP32 to ensure continuous data output should there be an intermittent loss of network connectivity. All of the data packets will be serialized using JSON format, and sent via the MQTT protocol to AWS IoT Core over a TLS-encrypted channel.

B. Cloud-Processing and ML-Inference.

Incoming sensor data streams are sent to AWS IoT Core then routed to two parallel processing pipelines via IoT Rules in AWS IoT Core after the data has been received by AWS IoT Core. The first pipeline writes the raw sensor readings to a Firebase Realtime Database enabling visualisations of the raw sensor readings on an up-to-date dashboard. The second pipeline triggers (calls) an AWS Lambda function which calls a hosted Model as-a-Service (MaaS) in Amazon SageMaker (A hosted Machine Learning Inference endpoint). The two separate, collaborating machine learning models in the machine learning pipeline: (1). random forest classifier; and RH) measurements (using a 0.5 Hz sampling rate).

(2). multi-class classifier. Random Forest classifier: measures the freshness of the food items classified into three categories (dairy products, fruit, and pantry) and classifies them as Fresh, Borderline, or Spoiled. The features upon which the random forest classifier was trained included:

1. Rolling average of the temperature and humidity during the past five minutes (mean and spot); 2. Rolling Standard Deviation of the temperature and humidity during the past five minutes (mean and spot); 3. Ratio of Carbon Dioxide (gas ratio) divided by the weight of the food item; and 4. Interaction terms between the temperature and humidity. The performance of the

food, which is important for food that has been packaged and for protein-rich food.

• DHT22 (temperature and humidity) – Provides accurate ambient temperature (±0.5°C) and relative humidity (±2%
random forest classifier on the out-of-bag (oob) test set was 94.3% accurate.

An LSTM based regression model is created to take inputs as windows of 30 continuous sensor readings (5 minutes) to generate an estimated time until spoilage (hours). The architecture includes 2 stacked LSTM layers (128 and 64 units) and a linear activation dense output layer. The training of the model was done using Adam Optimizer with a Mean Absolute Error (MAE) loss function over 80 epochs, and early stopping was applied. The model achieved an MAE of 1.7 hours on the validation dataset, which was significantly better than the ARIMA baseline (MAE: 4.2 hours).

C. User Interface Layer

The user interface includes both a Firebase-hosted progressive web app (PWA) and a companion Android app. Both interfaces provide real-time sensor readings, current freshness rating with confidence score, estimated time remaining until spoilage, and historical trend graphs. Push notifications will be automatically sent using Firebase Cloud Messaging when the classifier indicates that the item is classified as Spoiled or when the LSTM model indicates that the estimated time remaining until spoilage is less than 6 hours. SMS fallback alerts will be sent via Twilio API for locations where Internet access is intermittent.

4.DATASET COLLECTION AND MODEL TRAINING

A primary dataset was created from a laboratory controlled study that lasted 8 weeks, collected 3 different types of foods in 50 samples (fresh milk, bananas and mangoes, packaged goods); and monitored spoilage at various temperatures (10 degrees Celsius (C), 25 degrees C, and 35 degrees C). The study employed 10 second intervals to record sensor data (10 seconds) with trained food technologist (using organolpetic assessment and microbiological plate count) evaluating the physical spoilage condition of the samples at hourly intervals.

A total of 12,400 labelled instances making up the classification dataset and 8,900 sequences that were included in the training of the long short term memory (LSTM) model were created from the primary dataset; following the removal of noise in the data due to large deviations of sensor readings from normal using a zscore method. Data was split into three equal categories (80%

training, 10% validation, 10% testing) for classification and training. The three mucons (fear, caffeine, and ethanol) from the feature importance (random forests) analysis were most predictive for the sample types, such that they made up (combined) 61.2% of the cumulative variance (predictability) of spoilage by aerobic organisms.

5. RESULTS AND DISCUSSION

The proposed system was evaluated across the three food categories under realistic storage simulation conditions. Table II summarises the classification performance of the Random Forest model per food category.

Table II: Classification Performance by Food Category

Food Category	Precision (%)	Recall (%)	F1-Score (%)
Dairy	95.1	93.8	94.4
Fruits	93.7	94.9	94.3
Packaged Goods	94.2	93.5	93.8
Overall	94.3	94.1	94.2

The overall classification accuracy of 94.3% is a significant improvement over the basic ML integration described in Borawake et al. [11], which achieved approximately 87% classification accuracy for binary (fresh/spoiled) classification tasks. The multi-class formulation of the proposed model, through the intermediate Borderline state,provides additional actionable intelligence to the supply-chain community by allowing for proactive intervention prior to the onset of total spoilage.

The mean absolute error for the LSTM time-series predictor for forecasting shelf-life was 1.7 hours. With this level of accuracy, distribution centres are able to make better decisions related to stock rotation, prioritised delivery, and the adjustment of refrigeration settings. The false-positive alert rate during the test period was 3.1%, minimizing the occurrence of unnecessary discarding of food, which is one of the main concerns in the practical deployment of the technology.

The average system latency between sensor reading to push notification delivery was 1.84 seconds (SD: 0.31 s) under normal network conditions, meeting the requirements for real-time monitoring within warehouse settings. The average power consumption of the ESP32 node during active transmission was 180 mW, which is consistent with continuous mains-powered operation in a storage facility setting.

Unlike purely threshold-based systems like those from Christiena et al. ([6]) and Hai Prasaath ([4]), our suggested system does not require manual threshold calibration for the many different categories of food. Instead, the Random Forest model is able to automatically adjust the thresholds it has learned for each food category as it learns the conditions that define spoilage in the food category. This greatly decreases the burden of establishing a new system compared to a threshold-based system.

6. CONCLUSION

In this study, we demonstrated a smart IoT model with the ML model that allows for real-time food spoilage assessment and estimating food quality. The model production combines multiple sensors within an ESP32 network using an ML pipeline hosted in the cloud, which is composed of both a Random Forest Classifier and an LSTM model for time sequence predictions that would allow us to both retrieve real-time classification measures of food freshness and predict its shelf life. The overall model was evaluated and validated on numerous multipled food categories, including dairy, fruits, and packaged foods, via performance metrics of 94.3% correctness of the classification measure, a 3.1% false positive measure, and 1.7 hour mean shelf life prediction error (all superior performance to existing reported literature results on rule based systems using single modelling approaches).

With this framework, the ability to produce the required results using an IoT and ML multi-layered approach is made possible by the cloud infrastructure utilizing the AWS IoT Core, SageMaker, and Firebase, which can enable the scalability of architectural deployment into large-scale warehouse environments, without constraint of on-device compute. Future research will address the exploration of federated learning to enable food model personalisation across globally distributed sensor nodes, without the need to centralise raw food data, as well as explore the use of computer vision technology to detect surface spoilage in conjunction with gas and chemical sensor data.

REFERENCES

[1] B. Chanthini, D. Manivannan, and A. Umamakeswari, "Perishable food quality monitoring: an internet of things (IoT) approach," Int. J. Pure Appl. Math., vol. 115, pp. 63–66, 2017.

[2] A. Venkatesh, T. Saravanakumar, S. Vairamsrinivasan, A. Vigneshwar, and M. Santhosh Kumar, "A food monitoring system based on bluetooth low energy and Internet of Things," Int. J. Eng. Res. Appl., vol. 7, no. 3, pp. 30–34, 2017.

[3] K. Divya Bharathi and M. Rajalakshmi, "IoT based

food quality detection with bio sensor technology," 2018.

[4] K. Hai Prasaath, "Arduino based smart IoT food quality monitoring system," unpublished.

[5] R. K. Kaushal, T. Harini, D. P. Lency, T. Sandhya, and P. Soniya, "IoT based smart food monitoring system," Int. J. Current Eng. Sci. Res., vol. 6, no. 6, pp. 73–76, 2019.

[6] Christiena, H. S. Dhanushitha, N. Arsheya, and C. N. Ramya, "Survey on food quality monitoring system," IRJET, 2020.

[7] T.-B. Nguyen, T.-H. Nguyen, and W.-Y. Chung, "Battery-free and noninvasive estimation of food pH and CO2 concentration for food monitoring based on pressure measurement," Sensors, vol. 20, no. 20, p. 5853, 2020.

[8] R. Robin, G. E. Bridges, and S. Bhadra, "Wireless passive sensors for food quality monitoring: improving the safety of food products," IEEE Antennas Propag. Mag., vol. 62, no. 5,
pp. 76–89, 2020.

[9] Prajwal Atkare, P. Vaishali, and D. Sumit, "Food quality detection and monitoring system," in Proc. 2020 IEEE Int. Students' Conf. Electrical, Electronics and Computer Science (SCEECS), pp. 1–4, 2020.

[10] Jayakumar, C. H. Suman, C. H. Abhishek, S. Lakshmi, N. Shivprasad, and K. Rakshitha, "A literature review on techniques to detect food freshness," unpublished.

[11] M. P. Borawake, A. Sharma, U. Shaikh, S. Barkade, and R. Paturkar, "E-gadget to detect food freshness using IoT and ML," unpublished.

[12] S. M. Hasan, M. R. Ahsan, and M. D. A. Satter, "IoT-cloud-based low-cost temperature, humidity, and dust monitoring system to prevent food poisoning," in Proc. 2021 3rd Int. Conf. Electrical & Electronic Engineering (ICEEE), pp. 21–24, 2021.

[13] S. Mishra and C. S. P. Naidu, "Food spoilage detection system," IRJMETS, vol. 4, no. 4, 2022.

[14] L. Breiman, "Random forests," Mach. Learn., vol. 45, pp. 5–32, 2001.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780-1997.