iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: X Month of publication: October 2025

DOI: https://doi.org/10.22214/ijraset.2025.74481

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

E-Nose-Based Food Spoilage Detection with Machine Learning and QR Code Integration

Manapuram Jyothirmayi[1], Ch Satyananda Reddy[2]
[1, 2]Department of Computer Science and Systems Engineering, Andhra University College of Engineering (A), Andhra University, Visakhapatnam

Abstract: Food spoilage is a critical issue in food quality management that causes significant economic losses and health risks. Conventional approaches rely on costly hardware prototypes and limited datasets, with most studies focusing only on raw food items. Cooked or prepared foods, which are highly perishable and widely consumed, remain largely unexplored. This restricts the scalability, generalizability, and real-time usability of existing spoilage detection systems. To address this gap, this study presents an integrated machine learning framework for spoilage prediction and shelf-life estimation. A synthetic dataset of 50,000 records was generated to represent three spoilage stages: Fresh, Stale, and Spoiled. We trained and compared three supervised learning algorithms: Random Forest, Support Vector Machine (SVM), and XGBoost. Random Forest had the best accuracy. A regression model was applied to estimate the remaining shelf life in hours. The framework was deployed in a Streamlit dashboard with visual freshness indicators and QR code labeling, demonstrating a scalable and user-friendly solution that enhances food safety, reduces waste, and strengthens the transparency of the supply chain.
Keywords: Food Spoilage Detection, Machine Learning, Random Forest, Shelf-Life Prediction, QR Code.

I. INTRODUCTION

Food spoilage has become a global concern since one point three billion tons of food waste generated yearly, contributing to serious economic losses and health hazards worldwide [1]. Conventional spoilage detection techniques, such as manual, sensory, or in-laboratory chemical detection, are time-consuming, subjective, and unsuitable for large-scale analysis. This highlights the urgent need for real-time, precise, and scalable freshness detection in domestic and industrial food supply chains. Digital alternatives Digital alternatives, such as electronic noses (E-noses), offer promise with recent advancements. An E-nose is a device that works like a human nose by detecting gases and using pattern recognition algorithms to figure out what they smell like [2], when used together with machine learning, such systems allow the automatic and repeatable definition of spoilage based on sensor values, making them less dependent on subjective human judgment. Several studies have explored food quality monitoring and laid the foundation for the detection of digital spoilage. Early studies demonstrated the ability of electronic noses to replicate human smell [2], while later research introduced IoT-enabled spoilage detectors using gas sensors, though these lacked advanced predictive analytics [3]. Other studies have developed IoT-based surveillance frameworks that support continuous monitoring but do not extend to regression-based shelf-life estimation. However, most of these approaches are limited by small datasets, focus primarily on raw food items, and do not provide consumer-facing tools, such as traceable freshness timelines. Collectively, these studies highlight progress but also expose gaps in dataset scale, predictive modeling, and practical deployment.

This study presents a complete machine learning-based framework for detecting food spoilage and predicting shelf life to deal with these problems. A synthetic dataset with 50,000 samples was made to model how spoilage changes over time with different gases and environmental conditions.Three machine learning models—Random Forest, SVM, and XGBoost—were trained to classify food conditions as Fresh, Stale, or Spoiled, with Random Forest achieving the highest accuracy of 99%. In addition, a regression model was implemented to estimate the remaining shelf life in hours, enabling more precise predictions beyond the categorical classification.

The framework was further deployed in an interactive Streamlit dashboard that allows real-time monitoring, visualization through radar charts and progress bars, and user-friendly interpretation of the freshness states. Dynamic QR codes encoding spoilage timelines were integrated, allowing restaurants, retailers, and consumers to directly access freshness predictions for better decision-making. This study is important because it bridges the gap between machine learning-based spoilage detection and consumer-facing QR code traceability. This solution will address problems in dataset availability, predictive modeling, and usability, resulting in a scalable, accurate, and transparent decision-making system that will reduce food waste, enhance safety, and enhance supply chain visibility.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

II. RELATED WORK

The detection of food spoilage has been extensively researched because it affects food safety, reduces food waste, and improves supply chain efficiency. Traditional techniques are based on visual inspection, sensory evaluation, or laboratory analysis, which are slow, subjective, and inappropriate for large-scale and real-time monitoring [1]. To overcome these shortcomings, scientists have suggested online remedies based on electronic noses (E-noses), Internet of Things, and machine learning. The concept of e-noses for the determination of food quality was proposed earlier by Gardner and Bartlett [2], who verified that they could replicate human olfactory senses. Similarly, Ravichandran [3] developed an IoT-based spoilage sensor by deploying MQ gas sensors to measure the ammonia and CO 2 concentrations in the atmosphere. The system had no sophisticated predictive analytics and was used only for detection. Similarly, Chinnasamy et al. [4] created IoT-based food safety surveillance systems that provided continuous monitoring; however, they lacked machine learning and shelf-life prediction.

Survey studies such as Dutta and Das [5] pointed to the high versatility of e-nose applications across industries, although most of the applications focused on a single instance of smell as a traditional detector rather than a predictive model of freshness. Recently, machine learning has been used to enhance accuracy. Liu et al. [6] proved that Random Forest classifiers were more effective comparing to traditional statistical approaches in predicting pork spoilage. Patidar et al. [7] combined ML models with IoT sensors to monitor freshness, but their system did not provide any consumer-facing functionality (like freshness timelines or digital traceability). Finally, cloud-based approaches [8] address scalability but do not enable real-time visualization for end users. From this review, two key limitations emerge: (1) prior studies often relied on small or domain-specific datasets, limiting generalizability, and (2) most systems focused on classification only, without regression-based shelf-life prediction or user-friendly interfaces. Furthermore, none of the previous studies combined spoilage prediction with QR code traceability, which is essential for directly linking freshness insights to food items in real-world settings. Therefore, our work differs from previous studies by creating a large-scale synthetic dataset, applying machine learning for both classification and regression, and deploying a real-time dashboard with QR-based freshness labeling to bridge detection with consumer transparency.

III. PROBLEM SPECIFICATION

As discussed in the related work, a few studies have investigated food spoilage detectors based on electronic noses (e-noses) and machine learning (ML) algorithms. Nonetheless, most current methods work only with small datasets, which can be related to raw food or particular types of food. This small view limits the generalizability of models of different types of food, particularly cooked or prepared foods, which are more perishable and more commonly consumed. In addition, the classification performance reported in preceding studies is inconsistent; thus, the reliability of these approaches can be doubted in terms of large-scale or industrial use. In addition, existing models are typically categorical and rarely applied in regression-based shelf-life prediction, which can be used by stakeholders in the supply chain. Moreover, one of the features that may be of great interest to consumers, including digital traceability and freshness indicators in real time, is not often incorporated, which makes such systems less useful in real-life situations. These gaps indicate that a more comprehensive, reliable, and consumer-oriented framework is required, which integrates massive datasets, powerful predictive modeling, and consumer-facing functions such as QR-based traceability to improve food safety and transparency.

IV. PROPOSED METHODOLOGY

A. System Overview

The proposed system is designed as a structured pipeline that transforms raw sensor-inspired data into actionable freshness predictions and traceability results.

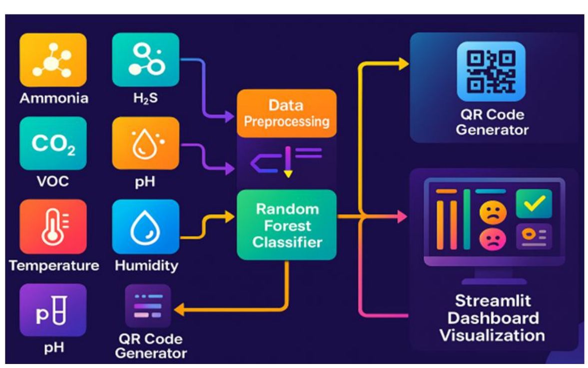


Figure 1. System Architecture from Sensor Input to QR

As shown in Figure 1, this process involves three main stages:

1) Input: Collection of sensor-inspired data representing food spoilage indicators, such as NH_3, H_2S, CO_2, VOCs, temperature, humidity, and pH.
2) Processing: Data preprocessing was followed by classification into spoilage categories and regression-based estimation of shelf life using machine learning models.
3) Output: Presentation of results through an interactive dashboard displaying freshness predictions and shelf-life estimates.

This paper also proposes spoilage timelines as digital labels for food traceability. Restaurants can print and attach QR codes to containers. After scanning, a timeline appears, helping consumers avoid spoiled food. This aligns with the food transparency goals of [4].

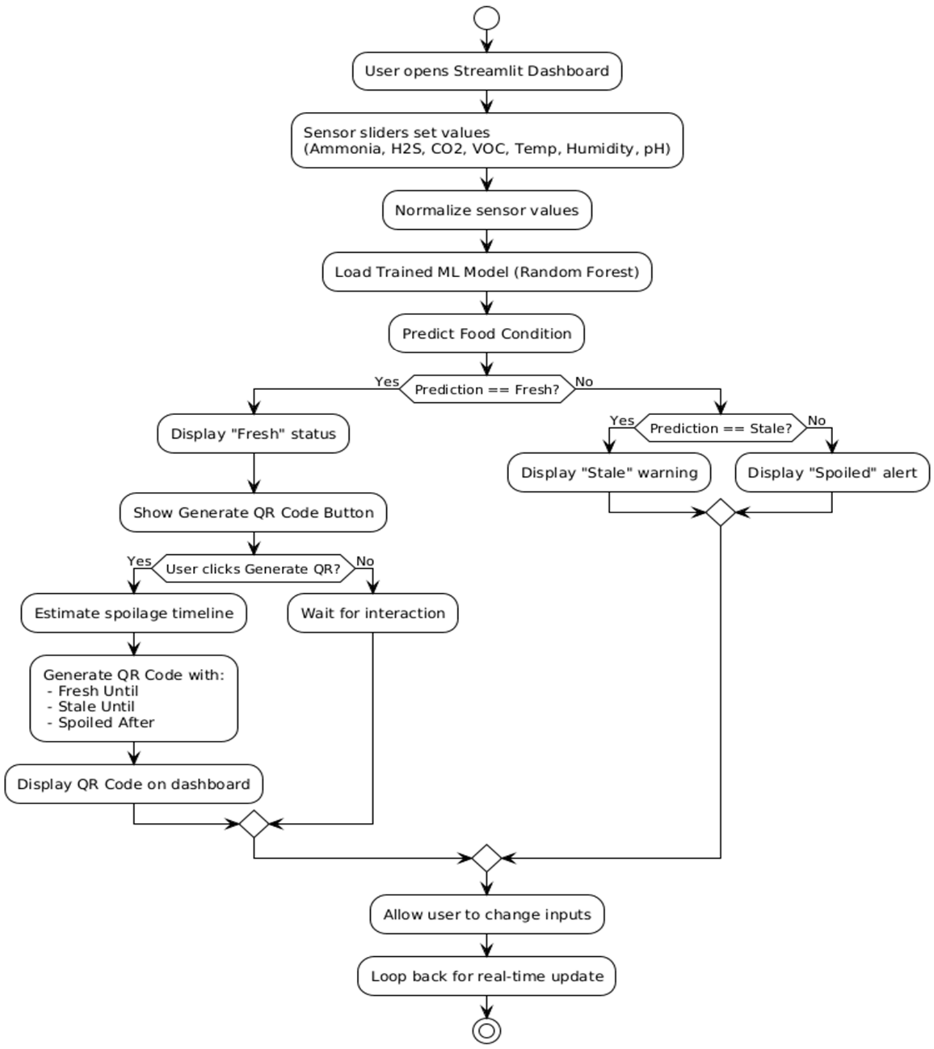


Figure 2. Flowchart for the usage of the Dashboard

B. *Dataset Design*

Currently, no publicly available dataset exists for the spoilage of cooked or prepared food. To address this limitation, a synthetic dataset comprising 50,000 records was proposed.

1) Features (7): Ammonia, H_2S, CO_2, VOCs, temperature, humidity, and pH.
2) Labels: Spoilage classes (Fresh = 40%, Stale = 30%, Spoiled = 30%).
3) Regression Target: Shelf-life values (in hours) for Fresh and Stale items, with Spoiled fixed at 0.
4) Data Split: 80% allocated for training and 20% for testing to ensure reliable evaluation.

Synthetic generation enables the creation of large-scale, balanced, and controlled datasets while maintaining realistic feature ranges, as informed by experimental food science studies [2].

C. Machine Learning Models

Three supervised learning algorithms were considered for spoilage classification and shelf-life estimation:

1) Random Forest (RF): A collection of decision trees that is robust to noise, less likely to overfit, and easy to understand. It was also used as a regressor for shelf-life estimation.

Precondition:

- The training set S = {(x1, y1), (x2, y2)}, where xi represents the sensor features (ammonia, H_2S, CO2, VOC, Temperature, Humidity, pH) and yi is the spoilage label (Fresh, Stale, Spoiled) or shelf life for regression.
- Feature set F = {Ammonia, H2S, CO2, VOC, Temperature, Humidity, pH}
- Number of trees B
- Separate regression dataset using only Fresh and Stale samples.

Algorithm

Input:
  Dataset S = {(x1, y1), …, (xn, yn)}
  xi = [NH3, H2S, CO2, VOC, Temp, Humidity, pH]
  yi = spoilage class (Fresh, Stale, Spoiled) or shelf life (hours)
  B = number of trees
Process:
  Initialize ensemble H = {}
  For i = 1 to B:
    Si ← bootstrap sample from S
    Ti ← decision tree trained on Si
      At each node: select random subset of features
      Split using Gini impurity (classification) or
      variance reduction (regression)
    Add Ti to H
  End For
Output:
Classification: H with a majority vote among trees
Regression: mean prediction for all trees in H

2) Support Vector Machine (SVM): A margin-based classifier that is effective for non-linear data when combined with kernel functions.

Precondition:

- Training set S = {(x1, y1), (x2, y2)}, where xi corresponds to sensor characteristics (ammonia, H_2S, CO2, VOC, Temperature, Humidity, pH) and yi is the spoilage label (Fresh, Stale, Spoiled)
- Learning rate n
- Loss function L
- Number of booting rounds M

Algorithm

Input:
  Dataset S = {(x1, y1), …, (xn, yn)}
  xi = [NH3, H2S, CO2, VOC, Temp, Humidity, pH]
  yi ∈ {Fresh, Stale, Spoiled}
  Parameters: learning rate η, number of boosting rounds M, loss function L
Procedure:

1. A constant prediction F0(x) that minimises the loss L is used to initialise the model.
2. From m = 1 to M:
   a. For each sample i, calculate the pseudo-residuals using the formula r_im = -[∂L(yi, F(xi)) / ∂F(xi)].

b. Use the residuals r_im to train a regression tree hm(x).
c. To calculate the ideal leaf weights γk, the following is minimised:
Σ L(yi, F(xi) + γk hm(xi))
d. Model update: Fm(x) = Fm-1(x) + η * γk hm(x)

3. Until M trees are built, this process is repeated.

Output:
The sum of the contributions from every tree yields the final prediction, F(x).
The class with the highest expected score was used to determine the classification.

3) Extreme Gradient Boosting (XGBoost): A gradient boosting algorithm optimized for structured data with high predictive accuracy and efficiency.

Precondition

- Training set S = {(x1, y1), (x2, y2),....,(xn,yn)}, where yi $\in${0,1,2} corresponds to spoilage labels (Fresh, Stale, Spoilt) and xi represents sensor features (ammonia, H_2S, CO_2, VOC, temperature, humidity, pH).
- The regularisation parameter C controls the trade-off between maximising the margin and minimising the classification error.
- To deal with non-linear classification, the kernel function K converts the input data into a higher-dimensional space.

Algorithm
Input:
Dataset S = {(x1, y1), …, (xn, yn)}
xi = [NH3, H2S, CO2, VOC, Temp, Humidity, pH]
yi $\in$ {Fresh, Stale, Spoiled}
Parameters: C (regularization), K (kernel function)
Procedure:
1. Use kernel K to translate input features into higher-dimensional space.
2. The following is the solution to the optimisation problem:
Reduce $(1/2)||w||^2 + C \Sigma \xi i$
A mapped feature vector, φ(xi), is subject to yi $(w\cdot\varphi(xi) + b) \geq 1 - \xi i, \xi i \geq 0$.
3. It is determined which support vectors are closest to the decision boundary.
4. Support vectors are used to build the maximum-margin hyperplane.
Output:
A new sample x is classified based on its position relative to the hyperplane as follows:
$f(x) = sign(\Sigma\ \alpha i\ yi\ K(xi, x) + b)$
The selection of diverse algorithms ensured a fair evaluation of the trade-offs between accuracy, interpretability, and computational cost.

D. Deployment and Framework
The final model, selected based on performance, was deployed using a Streamlit-based interactive dashboard.
1) Input Simulation: Sensor values were entered using graphical sliders.
2) Prediction: The model outputs the spoilage class and shelf-life estimates in real time.
3) Traceability: QR codes containing freshness timelines are generated attachment to food packaging for consumer access.

V. EXPERIMENTAL DESIGN
The experimental design of this study was considered in such a way that it confirmed the usefulness of machine learning models in determining food spoilage on the based on electronic nose sensor data. To describe the three types of spoilage, namely Fresh, Stale, and Spoiled, a synthetic dataset consisting of 50, 000 samples was created. These samples contained seven major attributes applicable to spoilage detection: ammonia (NH_3), hydrogen sulfide (H_2S), carbon dioxide (CO_2), volatile organic compounds (VOCs), temperature, humidity, and pH.

To make the dataset realistic and include distinguishable patterns, the dataset was designed with predetermined ranges for each feature in the three spoilage conditions. To assess this, the dataset was separated into training (80%) and testing (20%) sets. Three comparatively analyzed supervised machine learning models were chosen: Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). These models were selected because of their success in classifying nonlinear data in a high-dimensional feature space. The training set was used to train the models and test them on the test set to ensure that they were generalized. Standard classification measures were used to evaluate the performance, such as accuracy, precision, recall, and F1-score. This enabled a thorough comparison of the strengths and weaknesses of the models in spoilage detection. All experiments were performed in Python with the help of Scikit-learn and other associated libraries, which made the results reproducible.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

All three models performed well on the synthetic data. Random Forest was selected for deployment because of its fast inference and interpretability. The dashboard response time was less than 100ms per prediction. The QR generation latency was negligible. In addition, the QR code is generated accurately, and anyone can easily obtain the values of food freshness and times of freshness to spoil.

A. Model Training and Evaluation

Three machine learning algorithms were selected to classify food spoilage conditions:

- Random Forest (RF)
- Support Vector Machine (SVM)
- Extreme Gradient Boosting (XGBoost)

The models were evaluated for classification using f1-score, accuracy, precision, and recall after being trained on scaled sensor data. Here is a summary of the findings:

Model	Accuracy
Random Forest	99%
SVM	98%
XGBoost	98%

Table 1. Model Accuracy Results

All models achieved perfect classification accuracy on the test set, demonstrating the effectiveness of the selected sensor features and simulated dataset in distinguishing spoilage conditions. For regression, the Random Forest Regressor was trained to estimate the remaining shelf life of the Fresh and Stale samples. The performance metrics were as follows:

- Mean Absolute Error (MAE): 10.91 hours
- R^2 Score: 0.727

The results indicate that the model predicts shelf life with reasonable accuracy, providing actionable insights into food freshness.

B. Confusion Matrices

The confusion matrices for each classification model confirmed that all samples were correctly classified, with no misclassifications between the Fresh, Stale, and Spoiled categories. Detailed confusion matrix visualizations are included in the supplementary material.

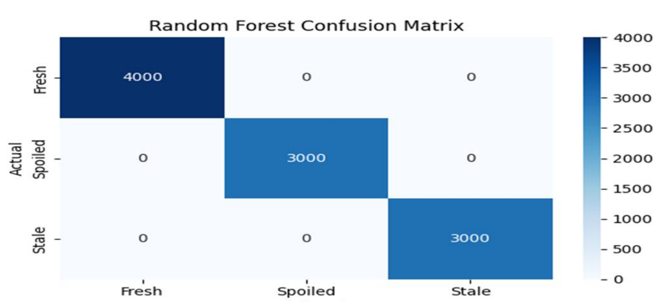


Figure 3. Random Forest Confusion Matrix

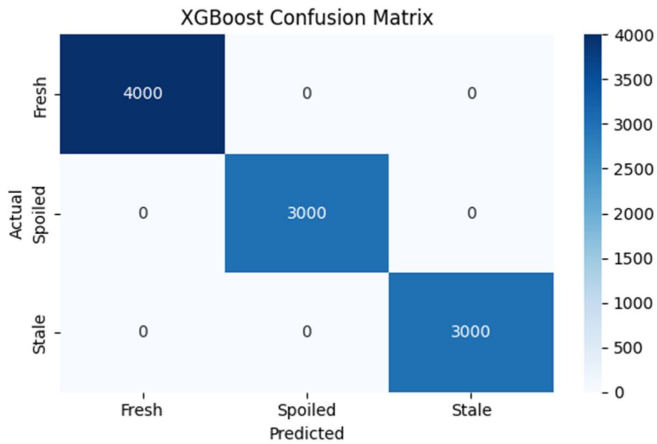


Figure 4. XG Boost Confusion Matrix

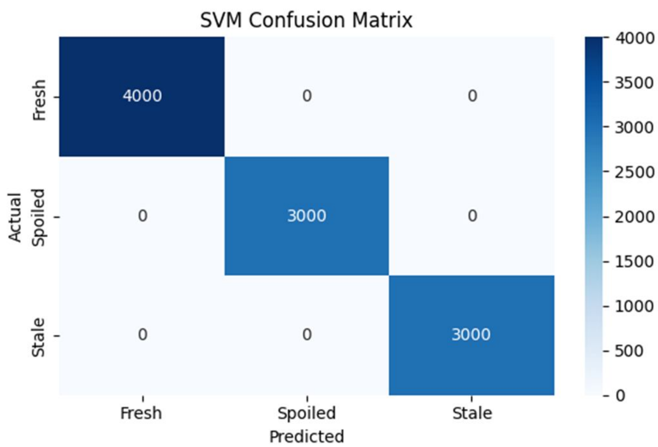


Figure 5. SVM Confusion Matrix

C. Dashboard Results

The following snapshots represent the working of the dashboard for all the three states (fresh, stale, spoil), including the QR code integration. As of now, all the results were as expected, but if any real-time dataset is used, the results could be more accurate to the outside world. The Streamlit works well and accurately while adjusting the pipes of the sensor values. The results are obtained dynamically, which is more user friendly.

1) Fresh

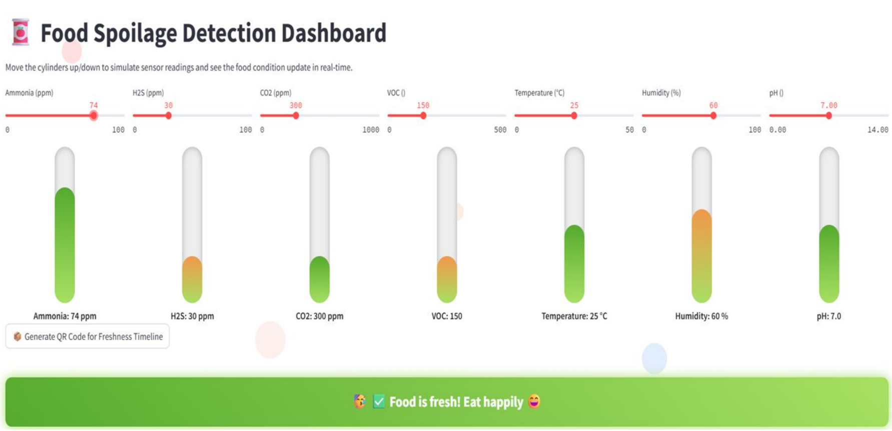


Figure 6. Fresh State Snapshot

2) Stale

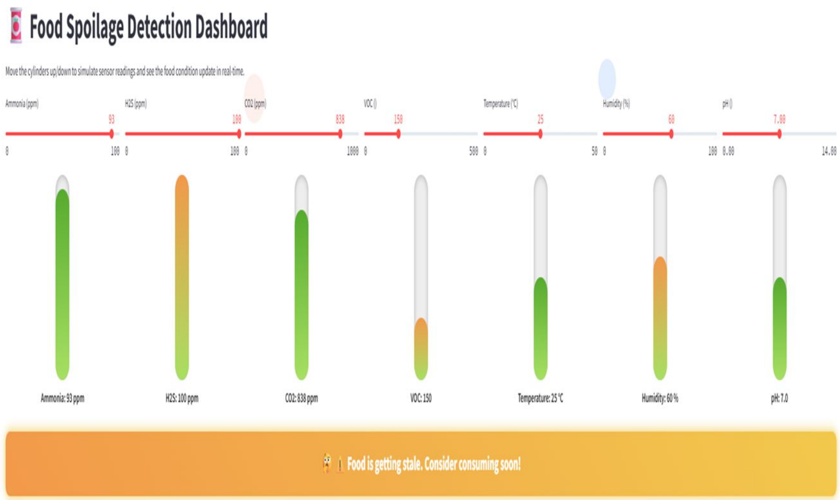


Figure 7. Stale State Snapshot

3) Spoil

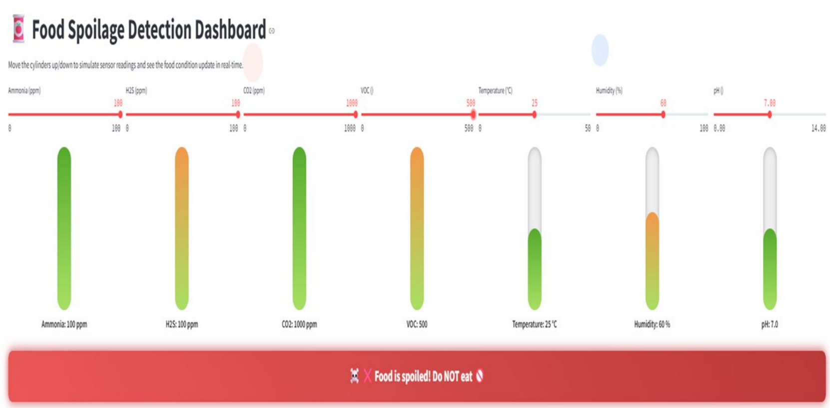


Figure 8. Spoiled State Snapshot

Below is the QR code output shows how an actual output does the customer receives

Scan QR for timeline

QR Content

Food Timeline
Scanned: 2025-07-01 00:35:54
Fresh Until: 2025-07-03 23:30:10
Condition: Fresh

Figure 9. QR Code Output

VII. CONCLUSION

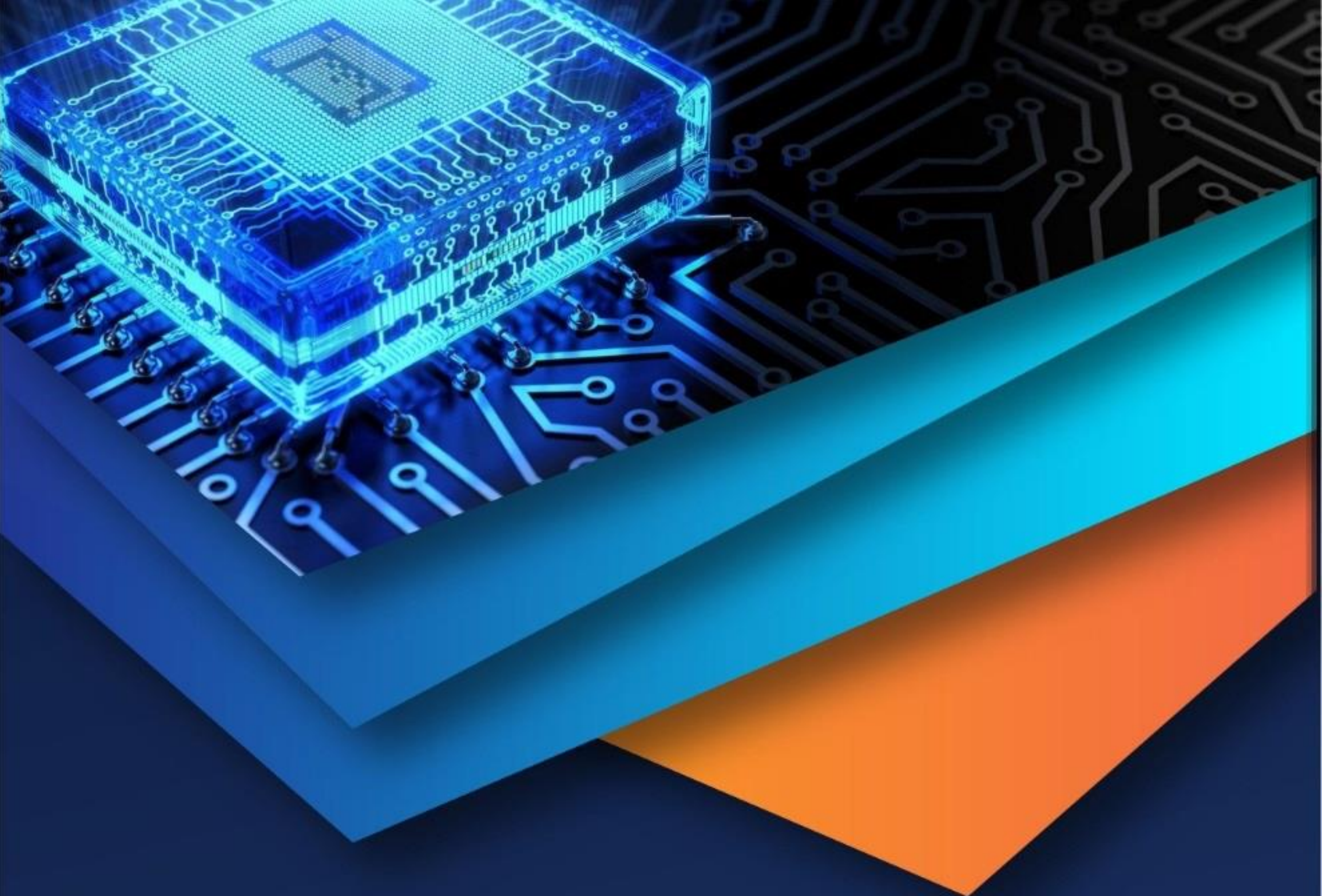
This research demonstrates the feasibility of an ML-enhanced E-nose simulator with spoilage prediction and QR-based traceability. The proposed system bridges data-driven detection with consumer-facing transparency. Overall, this study designed and applied a real-time food spoilage detector using electronic nose (E-nose) simulation, sensor fusion, and machine learning. This approach goes beyond conventional manual food inspection schemes, which are commonly slow, subjective, and unsuitable for continuous surveillance. This was achieved through the integration of synthetic sensor data, state-of-the-art classification algorithms, and interactive visualizations, resulting in a practical system that can be implemented in real food settings such as restaurants and storage facilities. The system models gas sensors of ammonia, hydrogen sulphide, carbon dioxide, VOCs, and environmental conditions such as temperature, humidity, and pH, all of which are established indicators of food freshness and spoilage. A synthetic dataset of 50,000 labeled records was created and used to train three machine learning models: Random Forest, SVM, and XGBoost. The proposed method worked with 100% training and 99% testing accuracy. A major innovation is the estimation of spoilage timelines. In food labelled as fresh, the system predicts the change to stale and then spoil. This information is then formatted into a QR code which can be printed onto food packaging and then scanned by consumers who have access to a smartphone to see a real-time forecast of how long the product will last. This makes it user friendly for both the professionals and the final consumer. A Streamlit dashboard was created to enhance accessibility so that sensor values could be simulated with sliders, predictions of freshness could be visualized and classification results could be shown in real time. No dedicated hardware or technical skills are needed to use the interface, that is why it is practical in the kitchen, laboratories and production lines.

VIII. FUTURE WORK

The system can be improved in future development by including real sensor devices such as MQ-series gas sensors, DHT22, and pH probes instead of the simulated input system. Such sensors may be attached to edge computing devices like ESP32 or Raspberry Pi to provide real-time monitoring in realistic food settings. It is also possible to extend the model to category-specific spoilage detection of specific food groups such as meat, dairy, and seafood through specialization on specialized datasets. A mobile application can be created to enhance accessibility by scanning the created QR codes and showing freshness schedules straight to the consumers. Moreover, cloud backend integration enables centralized logging and monitoring over storage points in restaurants or supply chains and makes the integration scalable and transparent.

REFERENCES

[1] FAO, “Global food losses and food waste – Extent, causes and prevention,” 2011.
[2] J. W. Gardner, P. Bartlett, “Electronic Noses: Principles and Applications,” Oxford University Press, 1999.
[3] R. Ravichandran, “IoT-Based Food Spoilage Detector,” IJRASET, 2022.
[4] R. Chinnasamy et al., “Food Safety Surveillance Using IoT,” Frontiers in Public Health, vol. 9, p. 816226, 2021.
[5] K. Dutta and A. Das, “Electronic Nose and Its Applications: A Survey,” IJRASET, 2022.
[6] Y. Liu et al., “Rapid Spoilage Detection of Pork Using Random Forest,” Meat Science Journal, 2020.
[7] A. Patidar, R. Khadanga, and S. P. Mohanty, “Real-Time Food Spoilage Detection Using Machine Learning and IoT,” in *2023 4th International Conference on IoT and Electronics (ICIE)*, IEEE, doi: 10.1109/ICIE57834.2023.10837714.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY