Grenze International Journal of Engineering and Technology, June Issue

Estimation of Freshness in Seafood based on Machine Learning Algorithms

Sowmya S[1], Vijaya Praneetha A[2], S. Tephillah[3] and M.V. Karthikeyan[4]
[1-4]Department of ECE, St. Joseph's Institute of Technology, Chennai, India.
Email: ssowmya2910@gmail.com, vijayapraneetha14@gmail.com, tephillahs@stjosephstechnology.ac.in, karthik.me09@gmail.com

***Abstract*— This work introduces a real-time system for monitoring seafood freshness that integrates a multi-gas sensor hardware module with predictions based on machine learning. A set of MQ-series sensors (NH, HS, CH, CO, Alcohol, H, LPG), combined with temperature and humidity sensing, detects volatile compounds emitted during seafood deterioration. The gathered data is analyzed via a machine learning pipeline where XGBoost, Random Forest, and KNN models execute TVC prediction (regression) and classify freshness into Fresh, Moderate, and Spoiled. XGBoost reached peak performance with a classification accuracy of 99.95% and an R^2 score of 0.9997 in estimating TVC. The suggested system offers a quick, affordable, and non-invasive method for real-time evaluation of seafood quality, appropriate for markets, cold storage, and supply-chain oversight.**

***Index Terms*— Gas sensor array, Total Viable Count (TVC), MQ series sensor, XGBoost, Random Forest, KNN, Freshness Classification.**

I. Introduction

Seafood spoils quickly because of its high level of moisture content and rich content of nutrients. Spoilage of seafood leads to release of VOCs (Volatile Organic Compounds) such as Ammonia (NH3), Hydrogen Sulfide(H2S), Methane (CH4) and Alcohol vapors Not only that the level of humidity and temperature can also change indicating that the seafood is no longer fresh. Pre-existing methodologies such as Total Viable Count (TVC), Total Volatile Base Nitrogen (TVB-N) and Chromatic analysis are complex and impractical in real- time. Advancements in sensors and embedded systems made it possible to develop compact, low-cost devices which can sense efficiently in real-time data. MQ gas sensor series can be integrated with microcontrollers like Arduino and ESP32. The sensors provide outputs proportional to the detected gas. Whereas Machine Learning helps efficiently in analyzing sensor patterns, Algorithms like XGBoost, Random Forest and KNN provides nonlinear relationship between gas con- centration and spoilage indicators. Thus integrating sensors with Machine learning algorithm results in a smart hardware-software system for predicting the freshness in seafood.

II. Objective

Designing a sensor module with a series of sensors such as MQ135, MQ137, MQ9, MQ2, MQ136, MQ8 MQ5, temperature and humidity sensor to measure real-time gas concentrations from the seafood. Then interfacing the sensors with the micro- controller should be done for data acquisition from the sensors.

Grenze ID: 01.GIJET.12.2.177_22

Developing and training ML models to perform regression and classification and categorize seafood in three major categories – Fresh, Moderate or Spoiled as result. Compare the performance of various algorithms used (XGBoost, Random Forest and KNN) in predicting freshness of the seafood.

III. Literature Review

Ever since the rise of demand for food safety, spoilage detection in various food items, especially seafood, got people's attention. Several methods exist for the same which include TVB-N measurement, pH testing and microbial plate counting. these methods provide accurate results, but time consumption is high, requires manual calculation and also they are not portable. Gas sensing of VOCs is a reliable alternative to estimate spoilage levels. Olafsdottir et al. [1] studied key factors in determining seafood quality and found out microbial activity, temperature and handling condition plays important role in it. Similarly [2] discussed spoilage patterns and identified volatile compounds such as ammonia and Hydrogen sulfide as major contributors for decomposition of seafood. With growth of sensor technology Electronic-nose systems gained attention for detection of food spoilage. [7] proved importance of gas sensors in monitoring fish freshness by detecting VOCs. Negi et al. [9] proposed a spoilage detection system with semiconductor gas sensors. these systems rely on limited number of sensors thus prediction efficiency is very low. Recent studies show the effectiveness of ensemble learning methods in prediction of food quality. [5] proposed XGBoost outperforms conventional ML models in terms of accuracy and efficiency. [10] applied ML techniques in IoT based food quality monitoring system. However, the majority of the studies solely focused on simulation of sensor datasets. While few focused on sensor hardware and ML prediction. Pre-existing research lacks multi species scalability and regression-classification framework. This works aims fulfilling the gap by integrating hardware sensor array with advanced ML prediction pipeline.

IV. Methodology

The methodology consists of designing Hardware Module and development, data acquisition, preprocessing data, training model and evaluation.

A. Hardware Design

The hardware system consists of the following sensors:
MQ135 – NH_3, CO_2, MQ137 – NH_3, MQ136 – H_2S, MQ9 – CH_4, MQ2 – Alcohol, MQ8 –Hydrogen, MQ5 – LPG and DHT22/DS18B20 – Temperature and Humidity

B. Sensor Calibration and Noise Handling

Before collecting the data, MQ series sensors are pre heated and calibrated in pure air conditions to improve stability. Calibration of sensors is performed by exposing them to clean air and reference gases to minimize the offset errors.
To handle noises from surrounding and sensor drift, measurements were taken repeatedly for every sample and averaged before storage. During the preprocessing of data feature scaling is carried out to normalize the sensor output and reduce effects of sensor drift over the time.

C. Working

Sensors are placed in a sealed chamber with seafood samples. Sensors detect VOC, temperature and humidity of samples. Microcontroller reads values from the sensors. Sensor values are transmitted for machine learning processing

D. Data Preprocessing

All sensor outputs are merged together from various seafood species (COD, Tuna, Shrimp, Crab, Salmon). The dataset used in the study was collected using multi sensor array.
Data Acquisition is carried for multiple freshness stages such as fresh, moderate and spoiled conditions. A total of 3,36,000 sensor samples were collected with each Seafood species. For development of the model the data was randomly divided by 80:20 train-test split methodology which ensures balance in representation of all the three freshness classes in both subsets. Training data was used for learning regression and classification patterns whereas test data was used to measure performance efficiency. Species are encoded into numerical values and processed. TVC is used to classify among freshness classes

- Freshness=0(Fresh), when $TVC < 4$
- Freshness=1(Moderate), when $3 < TVC < 7$

- Freshness=2 (Spoiled), when TVC>6

Total Viable Count (TVC) as ground truth were referenced based on established microbiological standards from food safety literature.

E. Model Training

Algorithms:
- XGBoost
- Random Forest
- KNN

Metrics:
- MAE
- RMSE
- R^2
- F1-score
- Confusion matrix

F. Model Validation

Model performance is evaluated using standard metrics of unseen test data. Regression models were evaluated by Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and R^2 score(accuracy). Whereas the classification models were evaluated using accuracy, precision, recall and F1-score. Additionally, confusion matrices were also considered for evaluating efficiency between models. These validation approaches provided reliable comparison of different algorithms. XGBoost achieved best value of 0.9997 in regression and produced the highest accuracy of 99.95% in classification.

G. Feature Importance Analysis

- NH3 (MQ135, MQ137)
- H2S (MQ136)
- Co2
- Methane (MQ9)
- Temperature and humidity

As dominant features for spoilage prediction.

The Feature Importance graph from XGBoost shows, in Regression highest importance is given to Time. Followed by Features like NH3 received from MQ135 sensor .H2S, Alcohol and Hydrogen are given almost equal importance. other features such as Methane, LPG, humidity and temperature are also considered as important features in Regression (TVC Prediction) by XGBoost Algorithm.

Similar to Regression, Classification also uses different features in determining freshness class of the seafood sample. Ammonia from both sensors (MQ135 and MQ137) plays an important role in Classification of freshness. Followed by other features such as LPG, Methane, Time, Hydrogen, Alcohol, H2S, Humidity and Temperature.

V. Results and Discussions

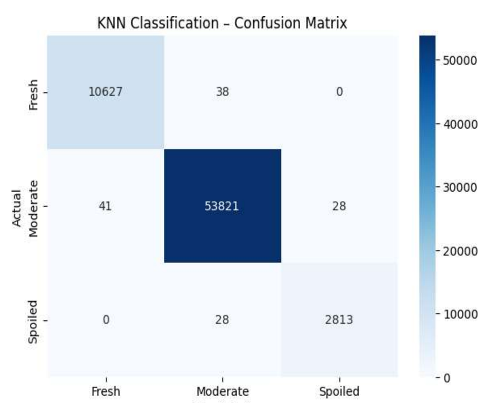

Figure 1. Confusion Matrix-KNN

The Performance of proposed system was evaluated by both regression and classification models which are trained on multi sensor data. The system successfully integrates gas sensor output with Machine Learning algorithms to predict the freshness of seafood. Out of various models trained XGBoost showed great predictive capability. while Random Forest achieved lowest MAE which shows its strong error stability. Whereas in classification all three models accurately predicted Fresh, Moderate and Spoiled classes. Confusion Matrix of different ML algorithms used is shown above.

The confusion matrix of KNN algorithm Figure.1, uses 2 factors (actual and predicted). The model which is trained out of 10668 Fresh samples - 10627 are correctly predicted as fresh with 41 fresh samples predicted as moderate and 0 fresh sample predicted as spoiled. while moderate is sometimes wrongly classified as fresh and spoiled. Spoiled is correctly classified as spoiled with very few samples (28) wrongly predicted as moderate.

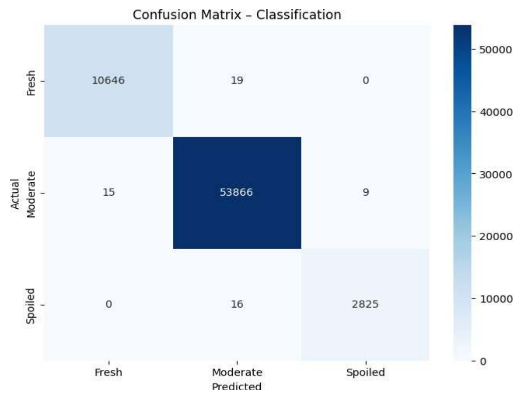

Figure 2. Confusion Matrix-Random Forest

Confusion matrix of Random Forest Figure.2., shows even less wrong prediction than KNN algorithm. Out of 10661 fresh samples 15 samples are wrongly predicted as moderate.19 moderate samples are wrongly predicted as Fresh and 16 as Spoiled while 53866 samples are correctly predicted as moderate. only 9 spoiled samples are wrongly predicted using the Random Forest Algorithm for training the model.

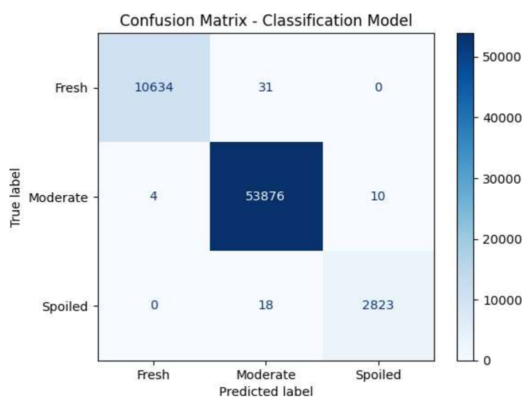

Figure 3. Confusion Matrix-XGBoost

The comparison of various Models used for seafood freshness prediction is given in Table. 1. with its metrics.

TABLE I. PERFORMANCE COMPARISON OF DIFFERENT MODELS

Model	MAE	RMSE	R^2 Score
XGBoost	0.00366	0.01226	0.99970
Random Forest	0.00262	0.01344	0.99963
KNN	0.00468	0.01923	0.9926

As shown in the table the Comparison of all three models are done by comparing its MAE value, RMSE value and R2 score. R2 score of XGBoost is the highest which indicates that the highest accuracy achieved model among the three is XGBoost algorithm.

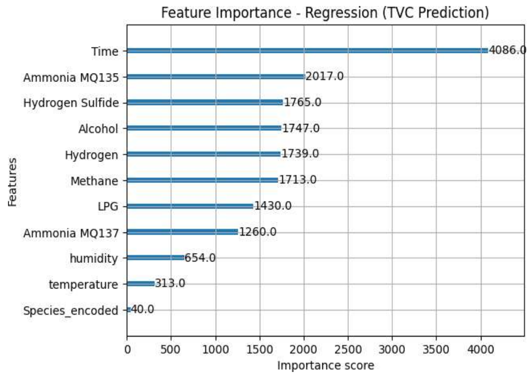

Figure 4. Feature Importance- Regression-XGBoost

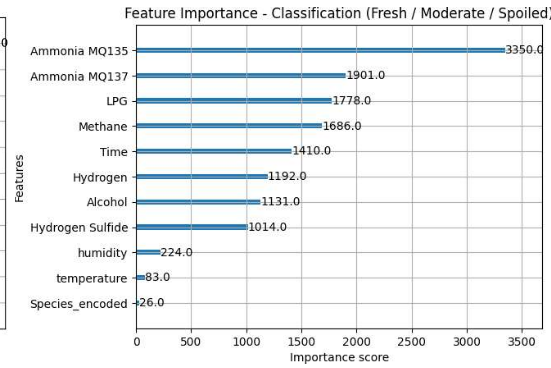

Figure 5. Feature Importance -Classification-XGBoost

VI. FUTURE ENHANCEMENT

Though the system is highly effective, few improvements can be made which includes
Integration of Cloud – sending real time sensor data to an IoT dashboard which enables remote monitoring of seafood freshness across markets and storage chains.
Mobile App Development – developing a mobile app to notify freshness alert, provide QR based batch identification and recommendations for storage condition.
Deep Learning Models – CNN or LSTM models can be used to analyze patterns in sensor data.
Usage in cold-chain Transportation – this system could be used by embedding it in transportation containers for continuous tracking of freshness during shipment of the seafood.
Development of e-nose integrated with sensors and ML model as shown in Figure.6 provides accurate results in prediction of freshness in seafood.

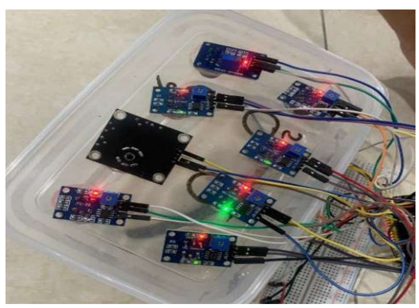

Figure 6. E-Nose Model

VII. CONCLUSION

This work introduces a integrated (Hardware-Software) system for real-time freshness prediction in seafood using multi- sensor gas module and ML algorithms. The series of sensors detect VOCs (volatile organic compound) which relates to spoilage and the Machine learning model especially XG Boost provided high accuracy for regression (TVC estimation) and classification. By evaluating the model, classification accuracy reached 99.95% and regression scores exceeded 0.999, thus the proposed approach proves to be efficient and reliable alternative to laboratory-based freshness prediction techniques. The integration of gas sensors and ML algorithms shows high potential for detecting freshness of food and is ideal for applications in seafood market, retail environment and supply chain monitoring. Future improvisations include connectivity with Cloud, usage of DL algorithms and IoT usage.

REFERENCES

[1] M. Olafsdottir et al., “Factors influencing fish quality,” Journal of Food Science and Technology, vol. 65, pp. 5–9, 2016.
[2] F. Shahidi and Y. Hong, “Seafood spoilage and safety,” Journal of Food Chemistry, vol. 161, no. 3, pp. 17–26, 2014.
[3] A. Boziaris, Seafood Processing: Technology, Quality, and Safety. Wiley, 2014.

[4] A. Alpaydin, Introduction to Machine Learning, MIT Press, 2020.
[5] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD, 2016, pp. 785–794.
[6] L. Breiman, “Random forests,” Machine Learning, vol. 45, pp. 5–32, 2001.
[7] D. Zhang, M. Li, and Q. Wang, “Electronic nose systems for monitoring fish freshness,” Sensors and Actuators B, vol. 345, 2021.
[8] Y. Li, X. Sun, and L. Pan, “Application of gas sensors in food spoilage detection,” IEEE Sensors Journal, vol. 22, no. 5, pp. 4782–4791, 2022.
[9] S. Negi et al., “Real-time spoilage detection using semi- conductor gas sensors,” International Journal of Food Engineering, vol. 13, no. 7, 2019.
[10] P. Kumar and R. Singh, “IoT-based food quality monitoring using machine learning,” IEEE Access, vol. 9, pp.79640–79652,2021.