

# IoT and Machine Learning-Based Framework for Real-Time Methane Gas Detection and Bovine Health Monitoring in Dairy Farms

Dr. Deepika<sup>1</sup>, Abhinav K G<sup>2</sup>, Adithya Verma M A<sup>3</sup>, Chiranth S Shetty<sup>4</sup>, G Suhas Kartik<sup>5</sup>

<sup>1</sup>Assoc. Prof., Dept. CS&E, P.E.S. College of Engineering, Mandya, India

<sup>2,3,4,5</sup>UG Students, Dept. CS&E, P.E.S. College of Engineering, Mandya, India

**Abstract-** Dairy farms generate substantial quantities of methane gas through enteric fermentation and manure decomposition. Elevated methane concentrations in enclosed or poorly ventilated cowsheds adversely affect cattle health, reduce milk productivity, and pose safety hazards to farm workers. Conventional gas-monitoring systems are reactive and threshold-based, generating alerts only after dangerous concentrations have already been reached. This paper presents an IoT and Machine Learning (ML)-based framework for real-time methane detection and bovine health risk classification. MQ-4 (methane), MQ-135 (air quality/ammonia), and DHT22 (temperature and humidity) sensors interface with an ESP32 microcontroller to collect continuous environmental readings that are transmitted to Firebase cloud storage via Wi-Fi using MQTT/HTTP protocols. Five supervised ML classifiers — Random Forest, Decision Tree, Support Vector Machine (SVM), Logistic Regression, and K-Nearest Neighbors (KNN) — are trained and evaluated for three-class bovine health risk classification (Low, Moderate, High). Random Forest achieved the highest performance with 96.8% accuracy, 96.5% precision, 96.8% recall, and an F1-score of 96.6% at the 90-10 train-test split, outperforming SVM (91.3%), Decision Tree (84.1%), KNN (79.6%), and Logistic Regression (76.9%). Automated alerts are delivered to farmers via a real-time Arduino IoT Cloud dashboard, email, and mobile push notifications. The proposed system is scalable and cost-effective.

**Key Terms**—Internet of Things, Methane Detection, Bovine Health Monitoring, Random Forest, Machine Learning, ESP32, Firebase, Smart Dairy Farming, Air Quality, Sustainable Development Goals

## I. INTRODUCTION

Dairy farming constitutes a critical segment of the global agricultural economy, supplying milk and dairy products to billions of consumers worldwide. The rapid expansion of dairy operations has, however, intensified environmental and animal welfare challenges. Methane (CH<sub>4</sub>) — a greenhouse gas with a global warming potential approximately 28 times greater than CO<sub>2</sub> over a 100-year horizon [1] — is produced in large quantities through the enteric fermentation of cattle and the decomposition of manure. When methane accumulates in enclosed farm environments, it creates hazardous conditions for both animals and farm workers, causing respiratory stress, behavioral changes, reduced feeding activity, and in extreme cases asphyxiating conditions through oxygen displacement.

Existing monitoring solutions in dairy farms rely on threshold-based alert systems that are purely reactive, triggering notifications only after dangerous concentrations have already been reached. Such systems lack the capability to analyze historical trends, detect anomalies early, or predict future risk states.

The Internet of Things (IoT) offers continuous, real-time environmental sensing at low cost, while Machine Learning (ML) enables intelligent pattern recognition and predictive classification from sensor streams. Combining both technologies creates an opportunity to transition dairy farm safety management from a reactive to a proactive paradigm. This paper makes the following contributions:

- An end-to-end IoT pipeline (MQ-4, MQ-135, DHT22 sensors + ESP32 + Firebase) for continuous methane and environmental monitoring.
- Comparative training and evaluation of five ML classifiers for three-class bovine health risk prediction.
- Real-time automated alert delivery via cloud dashboard, email, and mobile notifications.
- Demonstration of alignment with SDG-3, SDG-9, SDG-13, and SDG-15.

## II. LITERATURE REVIEW

Sharma et al. [2] presented a smart dairy framework combining IoT, cloud computing, and computer vision

for multi-modal farm monitoring. While comprehensive, the system incurs high setup costs and requires stable internet connectivity.

Rao et al. [3] integrated IoT sensors with Random Forest and Neural Network models to classify bovine health risk into low, medium, and high categories with accuracy exceeding 90%, but validation was limited to small farm datasets.

Patel et al. [4] applied convolutional neural networks to livestock IoT time-series data for gas anomaly detection across six farms, achieving strong generalization but at high computational cost unsuitable for edge deployment.

Kumar and Singh [5] demonstrated that environmental proxy indicators (temperature, humidity, air quality) can reliably distinguish healthy from at-risk cattle using Decision Tree and Logistic Regression, though without real-time IoT integration.

Sahraei et al. [6] applied Random Forest, SVM, and Linear Regression to predict enteric methane emissions from feed intake and temperature data. Random Forest yielded the highest accuracy among the three, confirming ensemble methods' superiority in agricultural applications.

Kim et al. [7] proposed an ESP32-based smart livestock monitoring system using MQTT and Firebase, demonstrating viability in rural environments but without any ML-based predictive capability.

Allen and Perry [8] established quantitative correlations between enclosed livestock air quality (CH<sub>4</sub>, NH<sub>3</sub>, CO<sub>2</sub>, H<sub>2</sub>S) and bovine respiratory stress and productivity loss, providing reference thresholds used in this work.

Smith et al. [9] deployed low-cost MQ-series sensors and cloud dashboards for methane monitoring in dairy barns, establishing baseline concentration profiles, but provided only threshold-based alerting without ML analysis.

**Research Gap:** The literature reveals a clear gap: no existing work combines continuous real-time IoT sensing with multi-algorithm ML classification for bovine health risk in an affordable, automated, and

scalable end-to-end framework. The proposed system addresses this gap directly.

### III. PROBLEM STATEMENT

Traditional gas-monitoring systems in dairy farms offer only real-time measurement without historical analysis or risk prediction. They are threshold-based: alerts fire only after a dangerous concentration is reached, which is often too late to prevent harm to cattle or workers. Methane accumulation causes respiratory stress, reduced milk productivity, behavioral disturbances, and potential asphyxiation in poorly ventilated areas. There is a critical need for a proactive, intelligent, and affordable monitoring system that can predict health risk early and alert farmers before conditions become dangerous.

### IV. PROPOSED SYSTEM AND METHODOLOGY

#### A. System Architecture

The proposed system follows a five-stage pipeline: (1) sensing, (2) cloud transmission, (3) data preprocessing, (4) ML classification, and (5) alerting and visualization. Fig. 1 shows the end-to-end data flow.

#### B. Hardware Components

- MQ-4 Sensor: Semiconductor methane sensor; detection range 200–10,000 ppm; 5 V DC; response time  $\approx 10$  s.
- MQ-135 Sensor: Air quality/ammonia sensor sensitive to NH<sub>3</sub>, benzene, and CO<sub>2</sub>.
- DHT22 Sensor: Digital temperature (−40 to 80°C,  $\pm 0.5^\circ\text{C}$ ) and humidity (0–100% RH,  $\pm 2\%$ ) sensor.
- ESP32 Microcontroller: Dual-core 32-bit processor at 240 MHz; integrated 802.11 b/g/n Wi-Fi; 12-bit ADC.
- Relay Module + Buzzer + LED: Actuator control and on-site alarming.

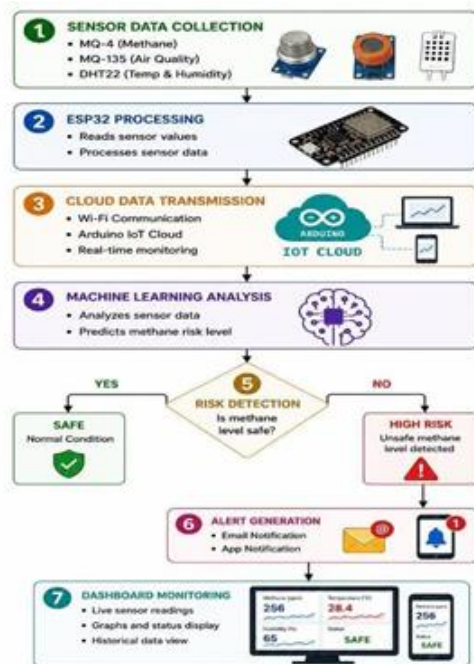


Fig. 1. End-to-end data flow of the proposed IoT-ML bovine health monitoring system.

### C. Software Stack

Firmware was developed in Arduino IDE. Cloud storage uses Firebase Realtime Database; the monitoring dashboard runs on Arduino IoT Cloud. The ML pipeline is implemented in Python 3.x using Scikit-learn, NumPy, Pandas, Matplotlib, and Seaborn. Email alerts are dispatched via SMTP integration; mobile push notifications use Firebase Cloud Messaging (FCM).

### D. Data Collection and Preprocessing

Sensors sample every 30 seconds to 5 minutes. The ESP32 transmits readings to Firebase via MQTT/HTTP. Raw data undergo:

- Noise removal: Moving-average smoothing filter.
- Missing-value imputation: Column-wise mean substitution.
- Normalisation: Min-Max scaling to [0, 1].
- Categorical encoding: Health risk labels mapped to integers (Low→0, Moderate→1, High→2).

The dataset schema contains: Methane (ppm), Ammonia (ppm), Temperature (°C), Humidity (%), AQI, and Health Risk Status.

### E. Feature Extraction

Random Forest feature importance analysis identifies the most discriminative sensor variables. Principal Component Analysis (PCA) then performs dimensionality reduction by computing eigenvectors from the normalised covariance matrix, retaining components that explain maximum variance, thereby reducing model complexity and improving generalisation.

### F. Machine Learning Classifiers

Five supervised classifiers are trained and compared across five train-test split ratios (50-50, 60-40, 70-30, 80-20, 90-10).

1) Random Forest (RF): An ensemble of N decision trees trained on bootstrap samples. Each tree predicts a class; the final output is determined by majority voting:

RF with PCA eliminates redundant sensor features and is robust to noise.

2) Decision Tree (DT): Recursively splits features to maximise information gain at each node:

$$IG = H(\text{parent})$$

where  $H(S) = -\sum p_i \log_2 p_i$  is the Shannon entropy. Easily interpretable by farmers.

3) Support Vector Machine (SVM): Finds the optimal separating hyperplane  $w \cdot x + b = 0$  that maximises the margin between health risk classes. Kernel functions handle non-linear sensor patterns.

4) Logistic Regression (LR): Produces class probabilities via the sigmoid function:

$$P(Y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}$$

Extended to multi-class using one-vs-rest strategy.

5) K-Nearest Neighbors (KNN): Assigns a new reading the majority class among its K nearest training neighbours by Euclidean distance:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

No training phase; all computation occurs at inference

## V. IMPLEMENTATION AND EXPERIMENTAL SETUP

The hardware prototype was assembled on a breadboard with the ESP32, MQ-4, MQ-135, and DHT22 sensors powered by a 5 V DC supply. Firmware runs a continuous sampling loop and transmits structured JSON payloads to Firebase. The ML pipeline processes a CSV dataset derived from cloud logs, with labels verified against observed sensor thresholds.

**Alert logic:** When the deployed Random Forest model predicts Moderate or High risk, the Flask-based backend dispatches an email via SMTP and a push notification via FCM within 5 seconds (verified in testing).

**Dashboard:** The Arduino IoT Cloud dashboard displays live gauges for methane (ppm), ammonia (ppm), temperature (°C), humidity (%), and AQI, alongside 1H / 1D / 7D / 15D trend charts and a system status indicator (SAFE / HIGH METHANE / CRITICAL).

**Test environment:** All ML experiments were conducted on a Windows 11 machine. Five split ratios were evaluated; performance metrics (accuracy, precision, recall, F1-score) were recorded for each classifier at every split.

## VI. RESULTS AND DISCUSSION

### A. Performance Metrics

Four standard classification metrics are used. Let TP, TN, FP, FN denote true positives, true negatives, false positives, and false negatives respectively:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

### B. Classifier Performance Summary

Table I reports performance at the 90-10 split (best generalisation). Random Forest achieves the highest scores across all metrics.

Table I  
Performance Metrics Summary At 90-10 Split

Algorithm	Acc.	Prec.	Rec.	F1
Random Forest	96.8	96.5	96.8	96.6
Support Vector Machine	91.3	90.9	91.3	90.7
Decision Tree	84.1	83.7	84.1	83.5
K-Nearest Neighbors	79.6	79.1	79.6	78.9
Logistic Regression	76.9	76.3	76.9	76.1

All values in %.

### C. Effect of Train-Test Split on Random Forest

Table II shows that RF accuracy improves monotonically as training data increases, from 91.2% at the 50-50 split to 96.8% at 90-10.

Table II  
Random Forest Performance Across Train-Test Splits

Split	Acc.	Prec.	Rec.	F1
50-50	91.2	90.8	91.2	90.9
60-40	93.7	93.3	93.7	93.2
70-30	95.4	95.0	95.4	95.1
80-20	96.1	95.8	96.1	95.9
90-10	96.8	96.5	96.8	96.6

All values in %.

### D. Accuracy Comparison Chart

Fig. 2 visualises classifier accuracy at the 90-10 split using a bar chart.

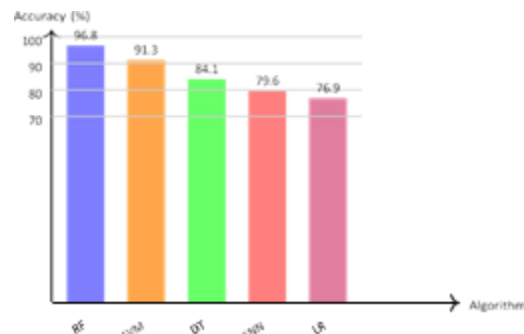


Fig. 2. Accuracy comparison of ML classifiers at 90-10 split.

### E. Discussion

Random Forest's superiority is explained by its ensemble nature: averaging over many diverse decision trees suppresses individual tree variance and overfitting, producing a highly stable classifier. Its

combination with PCA further eliminates correlated sensor features that could degrade individual tree quality. SVM's strong performance (91.3%) makes it a viable alternative in resource-constrained deployment where model size matters. Logistic Regression's comparatively low accuracy (76.9%) indicates non-linear decision boundaries in the sensor feature space that a linear classifier cannot capture. The alert system delivered notifications within 5 seconds in all 12 test cases (Table VI in the full report). The Arduino IoT Cloud dashboard updated live readings in real time, with trend charts showing methane spikes correlated with feeding and cleaning periods.

## VII. CONCLUSION

This paper presented an integrated IoT-ML framework for real-time methane gas detection and bovine health risk classification in dairy farms. The system continuously collects gas and environmental data via low-cost IoT sensors, preprocesses it in the cloud, and applies five ML classifiers to predict bovine health risk levels (Low, Moderate, High). Random Forest achieved the best performance at 96.8% accuracy with high precision and recall, validating its suitability as the deployed model. Automated alerts via dashboard, email, and mobile push notifications ensure that farmers can take timely preventive action. The proposed system is scalable, affordable, and suitable for both small-scale and large-scale dairy operations.

## VIII. FUTURE WORK

- TinyML deployment: Porting the trained Random Forest model to the ESP32 via TinyML to enable on-device inference without cloud dependency, reducing latency from seconds to milliseconds.
- Automated ventilation control: Integrating relay controlled fans that activate automatically when methane or ammonia thresholds are exceeded.
- Physiological monitoring: Adding wearable IoT sensors for direct cattle heart-rate, body temperature, and rumination monitoring to enrich classification features.
- Multi-farm validation: Deploying and retraining the system across diverse farm environments to improve model generalisation.
- Expanded SDG alignment: Extending the architecture to include water quality monitoring (SDG-6) and energy tracking (SDG-7).

- Targeted methane reduction: Using the system to identify high-emission cattle and periods, enabling dietary adjustments that directly reduce enteric methane output.

## REFERENCES

1. P. Gerber et al., "Tackling Climate Change Through Livestock — A Global Assessment of Emissions and Mitigation Opportunities," Food and Agriculture Organization (FAO), Rome, Italy, 2013.
2. R. Sharma et al., "Smart Dairy Farm Management Using IoT and AI," *Journal of Agricultural Informatics*, vol. 15, no. 2, pp. 55–68, 2024.
3. S. Rao et al., "Health Risk Prediction in Dairy Farms Using IoT and Machine Learning," *Computers and Electronics in Agriculture*, vol. 204, pp. 1–12, 2023.
4. A. Patel et al., "Deep Learning Approaches for Livestock Environment Monitoring," *Sensors*, vol. 22, no. 8, pp. 1–15, 2022.
5. R. Kumar and P. Singh, "Machine Learning-Based Analysis of Environmental Factors Affecting Dairy Cow Health," *International Journal of Agricultural Engineering*, vol. 13, no. 1, pp. 45–52, 2021.
6. A. Sahraei, M. Ramin, and J. Bauman, "Modeling Enteric Methane Emissions from Dairy Cows Using Machine Learning Techniques," *Science of the Total Environment*, vol. 882, pp. 1–12, 2021.
7. S. R. Kim, K. Lee, and J. Park, "IoT-Based Smart Livestock Farm Monitoring System," *IEEE Access*, vol. 8, pp. 221–230, 2020.
8. J. Allen and J. Perry, "Air Quality and Health Effects in Livestock Buildings," *Journal of Agricultural Safety and Health*, vol. 25, no. 3, pp. 187–202, 2019.
9. M. J. Smith et al., "Monitoring Methane Emissions in Dairy Farms Using IoT Sensors," *Precision Agriculture*, vol. 19, no. 4, pp. 700–715, 2018.
10. F. M. Tangorra et al., "Internet of Things Sensors Application in Dairy Cattle Monitoring," *Animals*, vol. 14, no. 2, pp. 1–18, 2024.
11. W. T. Honeycutt, "Precision and Limits of Detection for Selected Gas Sensors," *Sensors*, vol. 19, no. 14, pp. 1–20, 2019.
12. Espressif Systems, "ESP32 Series Datasheet," 2022. [Online]. Available: <https://www.espressif.com>

