Journal of Autonomous Intelligence (2024) Volume 7 Issue 3
doi: 10.32629/jai.v7i3.1285

ORIGINAL RESEARCH ARTICLE

Efficient sensor anomaly detection using Markov-LSTM architecture for methane sensing

S. Vishnu Kumar[1], G. Aloy Anuja Mary[1], Jasgurpreet Singh Chohan[2], Kanak Kalita[3,*]

[1] *Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi 600062, India*

[2] *Department of Mechanical Engineering, University Centre for Research & Development, Chandigarh University, Mohali 140413, India*

[3] *Department of Mechanical Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi 600062, India*

*** Corresponding author:** Kanak Kalita, drkanakkalita@veltech.edu.in, kanakkalita02@gmail.com

ABSTRACT

The integration of the Internet of Things (IoT) into industrial activities has unlocked myriad possibilities, particularly in applications like environmental monitoring, which facilitates effective landfill management. Nevertheless, IoT environments present challenges, including resource constraints, heterogeneity and potential hardware/software failures. These issues often lead to sensor anomalies, triggering false alarms and stalling data-driven systems. Existing models for edge devices frequently overlook the sensor life cycle, leading to extensive training times and significant computational demands. In this paper, a collaborative approach is proposed wherein a Markovian architecture gauges the operational state of a sensor, assisting the Long Short-Term Memory (LSTM) model in outlier detection within real-world data. Commercially available MQ-4 sensor alongside a microwave RADAR-based Methane (CH4) sensor in a tandem setup is employed to evaluate our methodology. The Bathtub curve and the Pearson Correlation Coefficient (PCC) function as the switching mechanisms for the Markov chain. Real-time data validation yielded an impressive 92.57% accuracy and 94.86% efficiency in anomaly detection. When benchmarked against the Autoregressive Integrated Moving Average (ARIMA) and the Prophet algorithm, our method demonstrated superior anomaly rejection rates of 9.63% and 3.01%, respectively. Implementing the Markov-LSTM model in methane sensing significantly enhances the accuracy of recorded sensor values compared to standard methane sensors.

Keywords: sensor anomaly detection; edge computing; bathtub curve; methane monitoring; industrial IoT; LSTM Network; Markov model

ARTICLE INFO

Received: 24 September 2023
Accepted: 2 November 2023
Available online: 25 December 2023

1. Introduction

The Internet of Things (IoT) has revolutionized our communication with physical devices, catalyzing a paradigm shift and paving the way for innovative applications, especially in real-time monitoring realms such as gas surveillance over a Solid Waste Landfill Site (SWLS). While IoT offers the capability to harness automation for capturing real-time data changes and environmental conditions, its reliability can be jeopardized by factors like sensor anomalies.

From the perspective of a SWLS, the emission of hazardous methane stands out as a paramount concern. methane, a potent greenhouse gas, can trap heat in the atmosphere at magnitudes far greater than carbon dioxide. Beyond its environmental implications, methane emissions from landfills also pose substantial health and

safety hazards. Short-term exposure to humans can induce symptoms like headaches, dizziness, nausea and fatigue. Long-term exposure has been associated with chronic respiratory ailments, neurological damage and even an augmented risk of cancer[1]. Given these stakes, the need for precise and dependable methane sensing technologies becomes paramount.

One contemporary solution involves deploying drones outfitted with sensors to survey landfills, presenting an economical method to monitor methane emissions. Various sensor technologies cater to methane concentration measurements, including Metal Oxide sensing[2], Electrochemical sensing[3], Infrared sensing[4] and Laser-Based sensing[5]. However, these methods grapple with challenges like sensor poisoning and external interferences, including light, dust and fog[6]. Recent advancements in using microwave RADAR to sense methane, which factors in microclimatic conditions, are showing promise. Yet, in aerial survey setups, sensor accuracy is vulnerable to an array of elements, such as dynamic moving environments, UAV battery life, sensor placement and overarching atmospheric conditions[7]. These variables can induce sensor anomalies, causing discrepancies between sensor readings and actual values.

Given the potential for false readings in SWLS management, the implications are grave—leading to suboptimal waste management and the unwarranted release of pollutants. To address this, our study employs two distinct methane sensors in a tandem arrangement, as depicted in **Figure 1**, bolstering the UAV-IoT system's ability to execute accurate in-situ measurements.

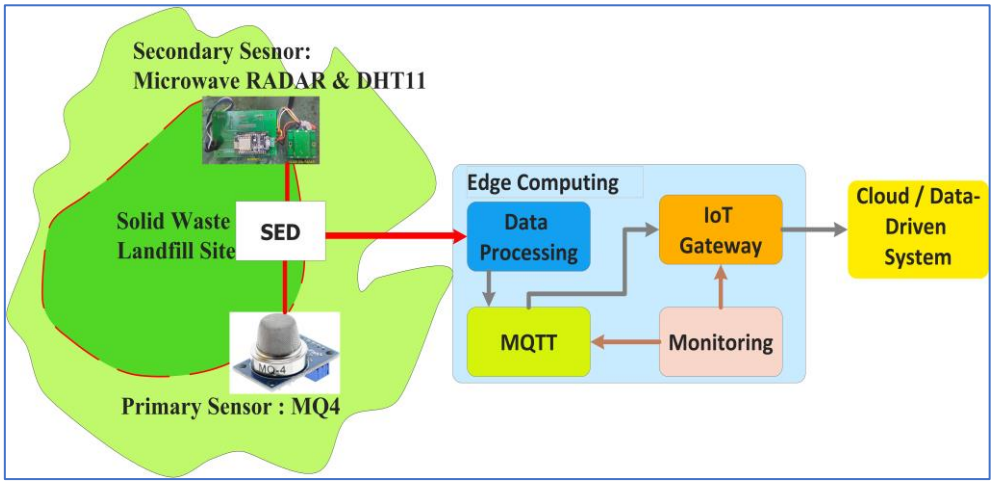


Figure 1. Architecture of methane monitoring expert system.

Both internal and external factors often contribute to sensor anomalies. Internal elements encompass issues in transducers, signal conditioner circuits and power supplies. In contrast, external factors include flight turbulence, vibrations, dynamic environments and cross-gas interference. Moreover, cyber-attacks and weak communication network links can also be counted as external factors affecting sensor reliability[8]. Given that IoT sensor systems are susceptible to either or both types of anomalies[9], pinpointing and managing these discrepancies to maintain optimal system functionality can be daunting and time-intensive[10,11]. In scenarios where malfunction arises, an IoT expert system should possess the capability to detect and mitigate the fallout—like isolating a node transmitting erroneous data. Thus, robust monitoring systems that can detect anomalies and security breaches in real-time, issue alerts promptly and initiate corrective actions are pivotal to uphold a system's reliability and precision[12,13]. Conventional methods that rely on threshold-based anomaly detection for edge devices grapple with the challenges of requiring skilled operators to set appropriate threshold values and an inability to auto-adjust these values based on changing situations.

There's been a growing interest in devising anomaly detection techniques for sensors that can autonomously identify issues and notify users. These techniques employ diverse methods, ranging from

Statistical Analysis and Machine Learning (ML) to Signal Processing, to discern and categorize sensor anomalies. An effective approach involves leveraging ML-based anomaly detection in edge computing, which lightens the load on the cloud side. This strategy promises zero-latency in data processing, improved response times, reduced network burdens and decreased cloud costs by relocating data processing tasks closer to data sources.

Methane sensing is paramount in landfill operations. Disruptions due to sensor malfunction or failure can severely impact these operations, leading to public health risks and environmental deterioration. Consequently, systems should be structured to autonomously detect and rectify technical glitches. To preemptively identify faults and outliers in lightweight sensor data, an intelligent framework that synergizes Markov and LSTM models for deployment on edge devices is designed. This paper elaborates on an IoT-driven methane sensing architecture proficient in real-time anomaly detection.

The rest of this paper is structured as follows: Section II provides a succinct overview of related works. Section III elucidates the proposed methodology. Section IV delves into the results procured from laboratory experiments and insights from an in-situ study at a landfill site. Section V concludes the paper, encapsulating the primary findings and discussing prospective avenues.

2. Related works

Landfill emissions, deemed hazardous, have garnered attention due to their contribution to climate change and public health concerns. For surveillance applications like SWLS monitoring of lower atmospheric gas dispersion, multi-rotor UAVs have emerged as the optimal solution[14–16], addressing both horizontal and vertical dimensions. Parameters such as Received Signal Strength and battery voltage can serve as indicators of communication-link quality and energy levels. These could potentially pinpoint sensor malfunctions and ensure dependable service in power-restrained UAV-based gas monitoring setups. However, sensors affixed to drones come with inherent challenges, such as data processing, wireless communication and particularly power consumption. Given that IoT modules draw from the primary power source, power-intensive algorithms can drastically reduce flight endurance. Semiconductor and Electrochemical gas sensors, often chosen for gas monitoring, are prone to issues like non-linear responses, unstable baselines and sensitivities to temperature and humidity[17]. Catalytic bead sensors, extensively used in the gas industry, have a vulnerability to drift and cross-responses in mixed gas environments[18]. With the advent of Industry 4.0, research has intensified around communication protocols, emphasizing energy efficiency, link stability and delay minimization[8], all critical to averting sensor data deviations alongside transducer malfunctions.

In the realm of IoT, Contreras-Castillo et al.[19] championed a routing method using the Markov model, accentuating the enhancement of network lifespan and decision-making efficiency—pertinent for UAV-based wireless gas sensing. Martín et al.[20] provided an in-depth comparison of three binary classifiers: ARIMA, Generalized Additive Model (GAM) and LOcal RegrESSion (LOESS). Factoring in IoT's inherent resource constraints, ARIMA emerged as the superior outlier detection tool when parameters were aptly calibrated. ML holds promise for anomaly detection. Chen et al.[21] introduced Autoencoder-based deep learning algorithms for anomalies, tailored for image and dense sensor values. These algorithms, focused on dimensionality reduction for outlier identification, are computationally intensive. Another proposal by Jiang et al.[22] centered on a dynamic collaborative structure for continuous IoT-generated data, with impressive classification accuracy but lacking in predictive foresight. This limitation can be rectified by adopting the techniques of Siami-Namini et al.[23] who championed time series data prediction. Kim et al.[24] suggested a "correlation coefficients and clustering" approach for industrial IoT, which, although promising, lacks real-time data validation. Meanwhile, the Prophet algorithm found favor with Yang et al.[25] for its stellar prediction of time-series solar irradiance, showcasing a remarkable 108.30% RMSE reduction against the widely-used ARIMA.

Deep learning approaches to audio signal outlier detection via Recurrent Neural Network (RNN) were explored by Arronte Alvarez and Gómez[26]. Further, Gökdemr and Çalhan[27] experimented with MQTT message security, probing for duplications, intrusions and alterations using labeled datasets and algorithms such as SVM, Naive Bayes and LSTM. LSTM emerged superior in terms of accuracy and reduced loss.

A review of existing literature reveals ARIMA, Convolutional Autoencoder (CAE), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Prophet and LSTM as predominant anomaly prediction algorithms for univariate time series data. However, some of these are incapable of long-term predictions. A summarized evaluation (**Table 1**) of these algorithms underscores their limitations: heavy dependence on historical data, inability to accurately detect sensor malfunctions, over-reliance on training datasets and neglect of sensor aging factors.

Table 1. Summary of literature on anomaly prediction.

Algorithm	Methodology	Inference
CAE[28]	Improves accuracy by focusing on data out layers.	Restricted to time intervals.
LSTM[29]	Three Long Short-Term Memory configurations to capture future states.	Deviations in bivariate data sequence.
LSTM[30]	LSTM-model with additional gates for lengthy time series data.	Detection of outliers is deviated.
Prophet[31]	Temperature forecasting using Facebook's Prophet Forecasting Model.	Lack of flexibility and limited support for multiple seasonality.
KNN[32]	A nonparametric, lazy method for predicting future target.	Distant evaluation yields mistaken predictions.
SVM[33]	Estimating spare part consumption utilizing univariate sales data.	Only works with univariate time series data.
Prophet[34]	Climate observations through outlier detection	Not tested in real-time data

3. Sensor anomaly detection using Markov-LSTM architecture

IoT devices have garnered considerable attention in monitoring applications due to their advantages, including cost-effectiveness and real-time data provision. When integrating IoT-enabled UAVs in landfill gas monitoring, these UAVs are equipped with sensors designed to detect and measure gas concentrations at varying locations across the landfill. This offers a comprehensive view of gas distribution, highlighting potential hazards. Additionally, the proposed study includes provisions to monitor parameters like temperature and humidity—factors that can notably influence gas production and dispersion within the landfill. Such data analytics enable operators to refine the landfill's design and operational strategies, aiming to curtail gas emissions and avert environmental contamination. However, the success of these actions heavily hinges on the accuracy and reliability of the sensor data relayed to the processing end. Any discrepancies in the received data can jeopardize the entire data-driven system, leading to potential malfunctions.

For data-driven environmental systems, sensor anomalies can be adeptly identified by juxtaposing current data with historical records, a strategy rooted in anomaly detection techniques[35]. This methodology encompasses real-time data acquisition from sensors, including but not limited to microwave RADAR and MQ-4 sensors. Subsequently, a processor equipped with an anomaly detection algorithm interprets this data. To enhance anomaly detection in the current study, the sensor's historical data and data from neighboring sensors is factored in. The conceptual framework of this architecture is depicted in **Figure 2**.

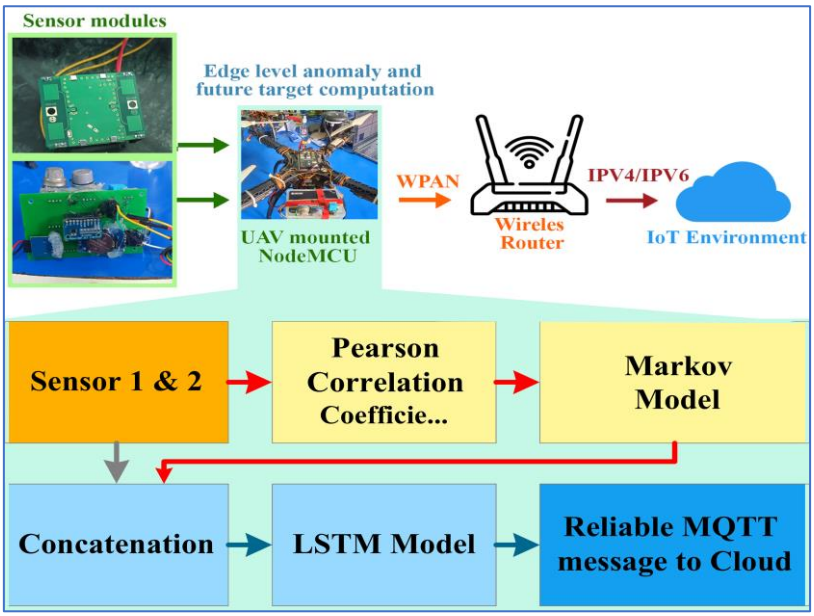


Figure 1. Overview of the proposed architecture.

The procured, timestamped real-time sensor data undergoes refinement via a data averaging algorithm, which purges extraneous noise. This streamlined data then feeds into a probabilistic model. The Markov model assesses the sensor's operational status—whether it is fully functional, intermittently faulty, or completely malfunctioning—as transitions hinge exclusively on the present state[36]. This Markovian state, alongside the sensor values, serves as input for the LSTM model to pinpoint anomalies[37]. A comprehensive overview of this model's functionality, coupled with the relevant algorithm, is presented in subsection 3.5.

3.1. IoT sensor module

This research employs two distinct sensors, leveraging metal oxide and microwave technologies, arranged in tandem to gauge methane concentrations. The MQ-4 sensor, renowned for its proficiency in open-environment methane detection, is selected as the primary sensor. Its operational mechanism involves gauging alterations in the electrical conductivity of its electrodes upon methane exposure. According to its specifications, this sensor mandates a 24-h pre-heating phase and exhibits a sensing range spanning from 200 ppm to 10,000 ppm. However, external factors like humidity and temperature can compromise its accuracy and sensitivity[38,39]. To circumvent these limitations, a secondary sensor is positioned adjacent to the MQ-4. This secondary sensor, an X-band 10.525 GHz bistatic microwave RADAR, boasts a coverage range of 2–16 meters. The RADAR's output undergoes amplification via an instrumentation amplifier (AD-620). Post amplification, any noise is eliminated by the processing unit and the resultant signal is translated into an equivalent methane concentration, factoring in the ambient temperature and humidity[40]. These state-of-the-art sensors are meticulously integrated into a custom-engineered Printed Circuit Board (PCB), as illustrated in **Figure 3**.

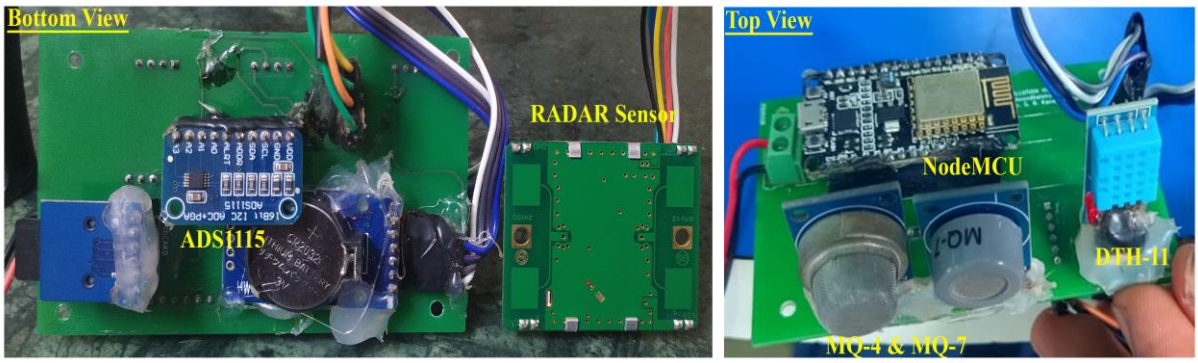


Figure 2. Top and bottom view of the custom-PCB.

3.2. Finite state machine (FSM) based on Bathtub curve

The life cycle of a sensor can be succinctly divided into three stages: the infant stage, the useful stage and the obsolete stage. As illustrated in **Figure 4**, the time interval $0 \leq t < T_i$ represents the infant stage. During this phase, premature failures commonly occur due to undetected defects and design oversights. The interval $T_i \leq t < T_u$ delineates the useful stage, where the sensor operates with minimal failures. Finally, the time frame $t \geq T_u$ indicates the obsolete stage, wherein the sensor experiences heightened failure rates due to aging and related wear[39]. The sensor's infant, useful and degradation rates are represented by the parameters $\lambda_t, \mu_t \; and \; \phi_t$ respectively, as defined in Equation (1):

$$\lambda(t) = \begin{cases} \lambda_t; & 0 \leq t < T_i \\ \mu_t; & T_i \leq t < T_u \\ \phi_t; & t \geq T_u \end{cases} \tag{1}$$

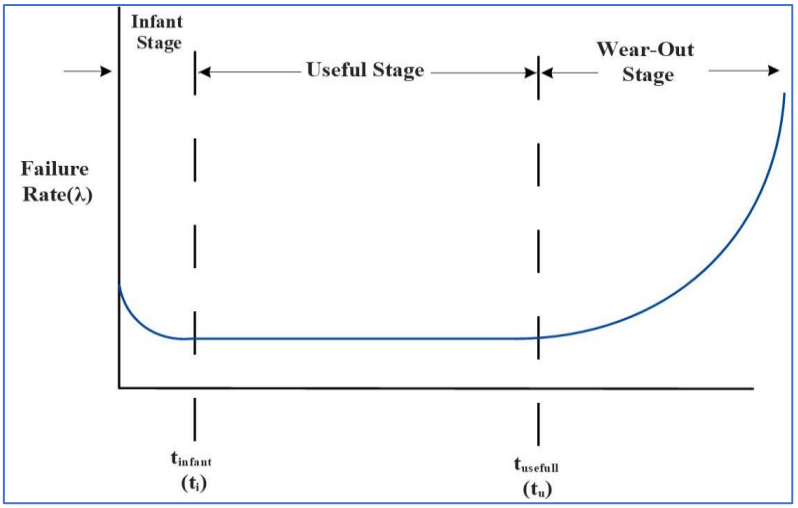


Figure 3. Bathtub curve of a sensor.

IoT sensors, especially those employed in data-driven environmental monitoring, can succumb to sensor-poisoning, a result of contaminants such as dust and cross-reactive chemicals. As the transition between states is dictated solely by the current state and is agnostic to historical data, the sensor's operational state is evaluated using a Markovian probabilistic model. The FSM, visualized in **Figure 5**, outlines the various functional states of the methane sensing system within the IoT framework, factoring in both bathtub curve and Markovian state transition conditions.

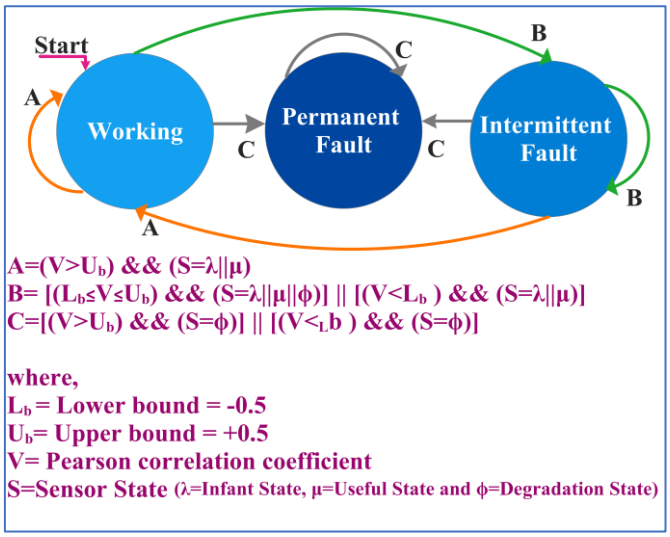


Figure 4. Markovian transition architecture.

3.3. Markovian architecture

Markovian models serve as robust tools for probabilistic modeling. Their foundational tenet is that the future state solely depends on the current state, regardless of the system's historical trajectory. This inherent memoryless property renders Markovian models especially apt for IoT systems, as sensor values primarily hinge on current data without past influence[41]. In this research, the Pearson correlation coefficient (PCC), combined with the bathtub curve value, functions as a switching mechanism between Markovian states. The bathtub curve value is manually designated, while the PCC value is deduced from the relationship between the MQ-4 and RADAR-derived methane concentrations, as expressed in Equation (2):

$$PCC = \frac{(n\sum xy) - \sum(x)\sum(y)}{\sqrt{(n\sum(x^2) - (\sum x)^2)(n\sum(y^2) - (\sum y)^2}} \tag{2}$$

Three operational states have been identified: the working state ("W"), intermittently faulty state ("I") and the faulty state ("F"). The potential Markovian state transitions are described in Equation (3), with the corresponding state transition probabilities captured in Equation (4):

$$S = \{W, I, F\} \tag{3}$$

$$P = \begin{matrix} \\ W \\ I \\ F \end{matrix} \begin{matrix} \begin{matrix} W & I & F \end{matrix} \\ \begin{Bmatrix} P_{WI} & P_{WI} & P_{WF} \\ P_{IW} & P_{II} & P_{IF} \\ P_{FW} & P_{FI} & P_{FF} \end{Bmatrix} \end{matrix} \tag{4}$$

Within a Markovian chain, the probability of selecting a transition condition after n steps is determined by Equation (5). The likelihood of moving from one state (point-a to point-b) after a single step is conveyed in Equation (6):

$$P_{ab} = P(P_n = b | P_0 = a) \tag{5}$$

$$P_{ab} = P(P_1 = b | P_0 = a) \tag{6}$$

Equations (5) and (6) delineate the time-homogeneous probability of node transition within the system, guiding the selection of the subsequent state. For a Markov chain characterized by time-homogeneous properties and an "s" saturation state, the probability is ascertained by Equation (7):

$$P_r(P_n = b) = \sum(P_{rb}P_r(P_{n-1} = r) \tag{7}$$

Equation (8) represents the "r" steps generalized probability,

$$P_r(P_n = b) = \sum(P_{rb}P_r(P_0 = r) \tag{8}$$

Upon determining the transition, the subsequent sensor state is computed. This value is then amalgamated

with the original sensor values to form the LSTM network's input data.

3.4. LSTM architecture

Long Short-Term Memory (LSTM) networks, a variety of recurrent neural networks (RNNs), are uniquely equipped to mitigate the gradient vanishing and exploding issues that plague traditional RNNs. For the task of univariate sensor sequence prediction, this research leverages a many-to-one LSTM architecture, which is essentially an RNN fortified with memory units. The memory capability of the LSTM architecture is endowed by its gate-controlled mechanism, which judiciously governs information retention. Two primary activation functions, "tanh" and "sig", are utilized. The tanh function yields values in the range of −1 to 1, while "sig" bounds values between 0 and 1 based on the sigmoid function, thereby enhancing the model's propensity to remember or forget information[42]. **Figure 6** visualizes these present and previous states as h_t and h_{t-1} respectively.

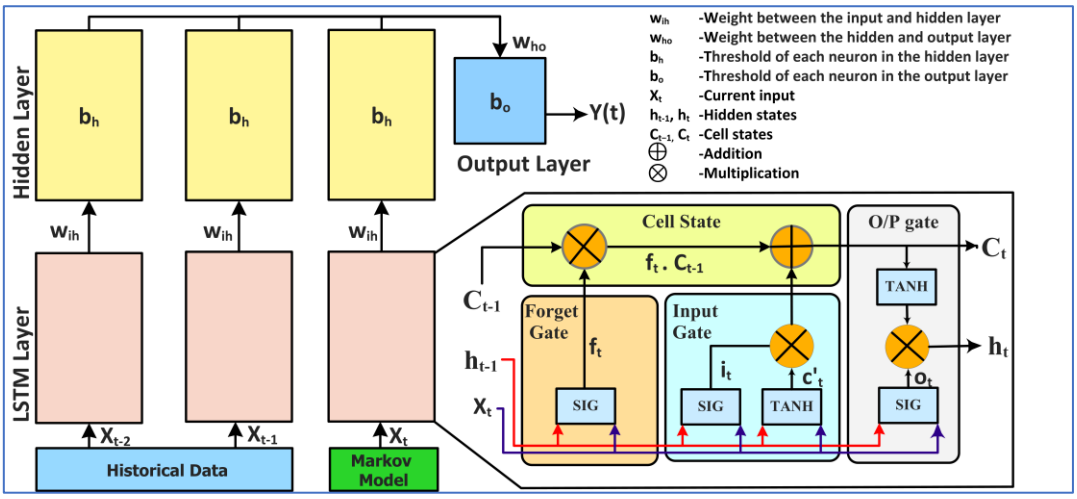


Figure 5. Proposed LSTM Architecture to predict outlier and future target values.

The LSTM unit cell's working can be categorized into three stages. In the first stage, the forget gate decides to retain or discard the input value. The hidden state ht-1 and the input state xt are both subjected to the sigmoid function, which results value, between 0 and 1. The calculation steps of the forget gate are described in Equation (9), where, f_t represents the forget gate, σ represents the sigmoid function, W_f and b_f represents the weight and bias vector of forget gate, h_{t-1} represents the output at time $t-1$ and x_t is the input vector at time *t*.

$$f_t = \sigma\left(W_f[h_{t-1}, x_t'] + b_f\right) \tag{9}$$

In stage two, the input gate applies the second sigmoid function to current state x_t and the previous hidden state h_{t-1} to produce values between 0 and 1. Contrary, the "tanh" function produce result between −1 and 1 for the same set of inputs[43]. The calculated sigmoid and tanh function outputs are multiplied[44]. The calculation steps involved in stage two are described by Equations (10) and (11), where, it represents input gate, c_t' represents updated current state, W_i and b_i represents the weight and bias vector of input gate and W_c and b_c represents the updated weights and biases respectively.

$$i_t = \sigma(W_i[h_{t-1}, x_t'] + b_i) \tag{10}$$

$$c_t' = tanh\,(W_c[h_{t-1}, x_t'] + b_c) \tag{11}$$

In stage three, the f_t is multiplied with previous cell-state C_{t-1} and C_t' is multiplied with it to produce new cell state C_t, which is described in Equation (12).

$$C_t = (f_t \odot C_{t-1}) + (i_t \odot C_t') \tag{12}$$

The output gate utilizes the updated cell state C_t to identify output h_t. The necessary values for this are acquired by applying the sigmoid function to the current state x_t and the previous hidden state h_{t-1}. The "tanh" function generated values are then used with the sigmoid function values to choose the output. The calculations involved in third stage are described by Equations (13) and (14), where, o_t represents the output gate, h_t represents the updated output and W_o and b_o represent the weight bias vector of the output gate.

$$o_t = \sigma(W_o[h_{t-1}, x_t'] + b_o) \quad (13)$$

$$h_t = o_t \odot tanh\,(c_t) \quad (14)$$

Equation (15) describes the mapping from x_t to $h_t$at time t. To improve prediction accuracy, the Markov approach is used as input to the LSTM along with the sensor values.

$$h_t = f(h_t - 1) \quad (15)$$

In scenarios where Unmanned Aerial Vehicles (UAVs) employ microwave RADAR or MQ-4 sensors to detect methane inherent sensor limitations come into play. The microwave RADAR, being analog and relying on reflected signal strength, is prone to false readings. Conversely, MQ-4 sensors can be compromised by dust and temperature fluctuations and might also respond to gases other than the target. Therefore, amalgamating readings from both sensors, enhanced by an anomaly detection algorithm, yields superior results compared to relying on individual sensors or a combined setup without anomaly detection. The harvested sensor data is then relayed to a cloud-based data system utilizing the IEEE 802.11.4 protocol and MQTT methodology.

3.5. Markov-LSTM integration methodology

The objective of the proposed method is to formulate a ML based model that efficiently operates within a resource-limited environment to eliminate anomalous sensor values. The model leverages the capabilities of the LSTM network to discern patterns within time-series sensor data across various operational states. This ability of the LSTM network to grasp context enables it to differentiate between standard variations in sensor data due to transitional states and genuine anomalies warranting intervention.

To categorize the operational states of the sensor, a Markov model is employed. Notably, factors like the aging effect of the sensors and deviations within sensor readings are harnessed as transition functions for the Markovian chain.

The operation of the proposed model can be elucidated in the subsequent steps, with Algorithm 1 providing the model's pseudocode:

Step #1: Initial Data Collection and Cleaning

- Accumulate real-time sensor data.
- Use data averaging algorithms to cleanse the data.
- Register environmental factors such as temperature, humidity and pressure that can influence sensor measurements.

Step #2: State Prediction Using Markov Model

- Ascertain the initial state of the sensor (e.g., Working, Intermittently Faulty, or Faulty) via the Markov model, considering current and historical data. Designate this as "Sensor_State".
- Employ Equations (1) through (8) to derive a probability matrix "P", as denoted in Equation (4) and infer the potential subsequent states of the sensor.

Step #3: Feature Vector Preparation

- Create a feature vector for the LSTM. This vector amalgamates genuine sensor readings (termed "Sensor_Values") with the prevailing sensor state ("Sensor_State"). Feature Vector = [*Sensor_Values*, *Sensor_State*].

Step #4: Data Transformation for LSTM Input

- Repackage the feature vectors into sequences which serve as LSTM input.
- The span of each sequence hinges on the configuration of the LSTM and the contemplated time steps.

Step #5: Anomaly Detection using LSTM

- Employ the LSTM model to prognosticate anomalies, informed by the sequences forged in the antecedent step.
- Referencing Equation (15), update h_t based on h_{t-1} and x_t, where x_t stands for the feature vector $[Sensor\_Values,\ Sensor\_State]$.

Step #6: Anomaly Handling

- In situations where the LSTM model spots an anomaly, a notification is triggered, or rectifying measures are initiated.

Step #7: Continuous Update Loop

- The Markov model refreshes the "$Sensor\_State$", gauging fresh readings against historical data.
- Concurrently, the LSTM model is updated with new feature vectors, ensuring the continuous detection of anomalies.

Algorithm 1 Proposed methane anomaly detection algorithm

```
# Initialize
Sensor_Data = []
Sensor_State = None
Feature_Vectors = []
LSTM_Model = initialize_LSTM()
    while True:
        # Data Collection
        new_data = collect_data()
        Sensor_Data.append(new_data)
        clean_data = clean(Sensor_Data)
        # State Prediction using Markov Model
        Sensor_State = Markov_Model(clean_data)
        # Feature Vector Preparation
        Feature_Vector = [clean_data, Sensor_State]
        Feature_Vectors.append(Feature_Vector)
        # Prepare LSTM Sequence
        LSTM_Sequence = prepare_sequence(Feature_Vectors)
        # Anomaly Detection using LSTM
        anomaly_flag = LSTM_Model.predict(LSTM_Sequence)
        if anomaly_flag:
            handle_anomaly()
        # Update Loop
Sensor_Data = Sensor_Data[-50:]   # Keep last 50 readings for efficiency
Feature_Vectors = Feature_Vectors[-50:]   # Keep last 50 feature vectors
```

4. Results and discussion

ML algorithms are gaining traction in the realm of anomaly detection due to their enhanced accuracy. While initially deployed to detect malicious customer behavior, anomaly detection now finds utility in industrial scenarios as well. The Markov-LSTM-based anomaly detection system presented in this study underwent rigorous field testing with real-time data.

4.1. Performance metrics of the proposed system

The distinct architecture of RNNs incorporates a hidden state pivotal in shaping the output for each timestamp. Typically initialized to zero at $t = 0$, this state is instrumental in prepping the RNN for training and testing. A specific function unfurls the network, yielding a sequence of three preceding data inputs relative

to the current input. **Figure 7** delineates this mechanism, revealing a notable uptick in algorithmic accuracy within the initial 200 iterations.

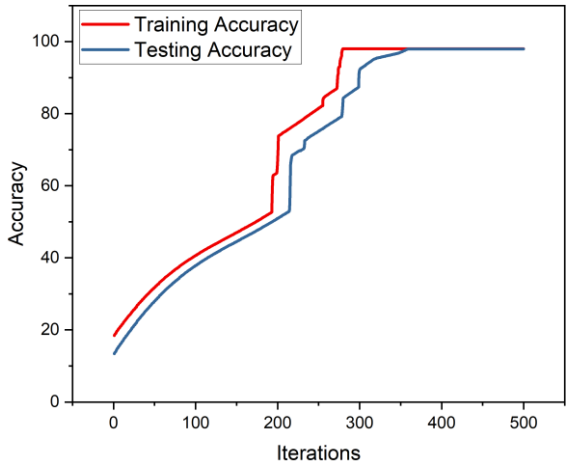


Figure 6. Real time sensor value-accuracy.

The training loss juxtaposed against validation loss for live-sensor data is depicted in **Figure 8**. Further, **Figure 9** contrasts the real-time methane values with predicted methane values, showcasing anomalies detected in real-time data. Impressively, the LSTM approach adeptly identifies and sidesteps anomalies stemming from external disturbances.

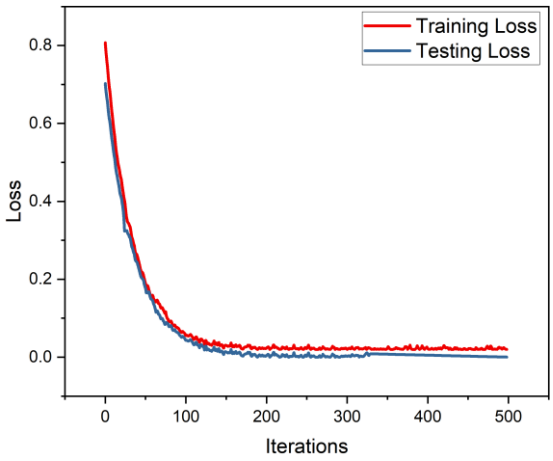


Figure 7. Real time sensor value-training and validation loss.

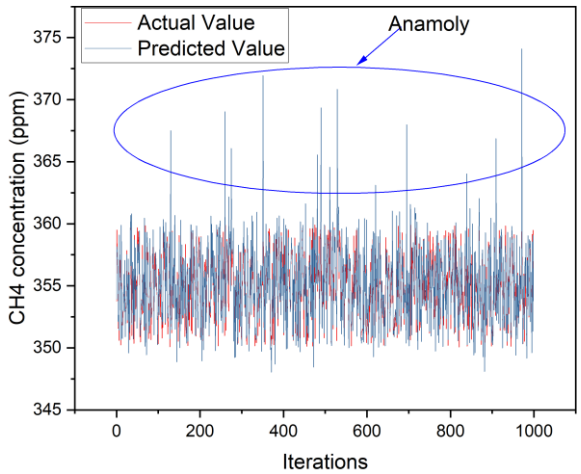


Figure 8. Real time sensor value-Anomaly detection.

For benchmarking purposes, the proposed model's training accuracy and efficiency were set against renowned time-series prediction algorithms, namely ARIMA and Prophet. The comparative insights are graphically represented in **Figure 10**.

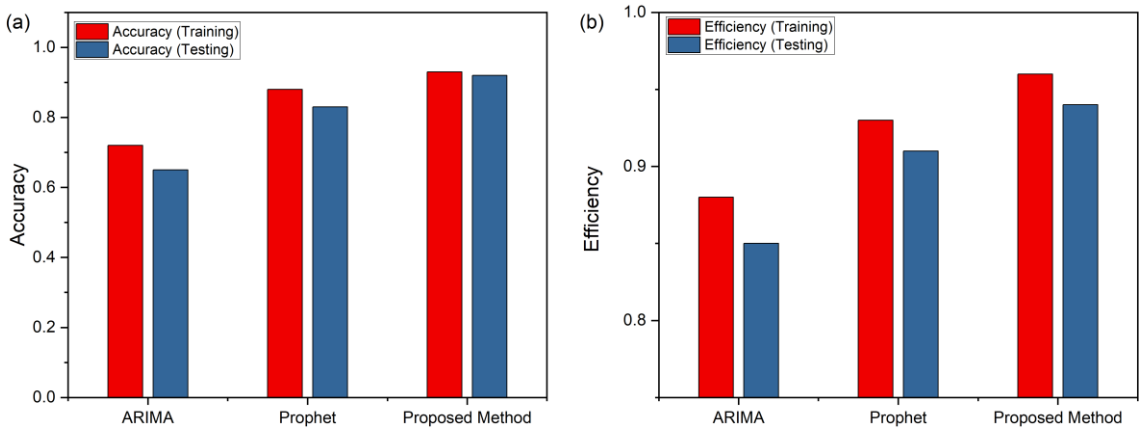


Figure 9. Comparison between various methods for **(a)** Accuracy and **(b)** Efficiency.

To minimize loss, the Root Mean-Square Logarithmic Error (RMSLE) was employed. This metric, formulated as in Equation (16), contrasts actual y and forecasted y' values. By leveraging the logarithmic function, the metric effectively nullifies minor disparities between the predicted and actual values.

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(log\left(y_i' + 1\right) - log\left(y_i + 1\right)\right)} \tag{16}$$

A comparative RMSLE analysis spanning ARIMA, Prophet and the proposed model underscores the latter's supremacy, as corroborated by **Figure 11**.

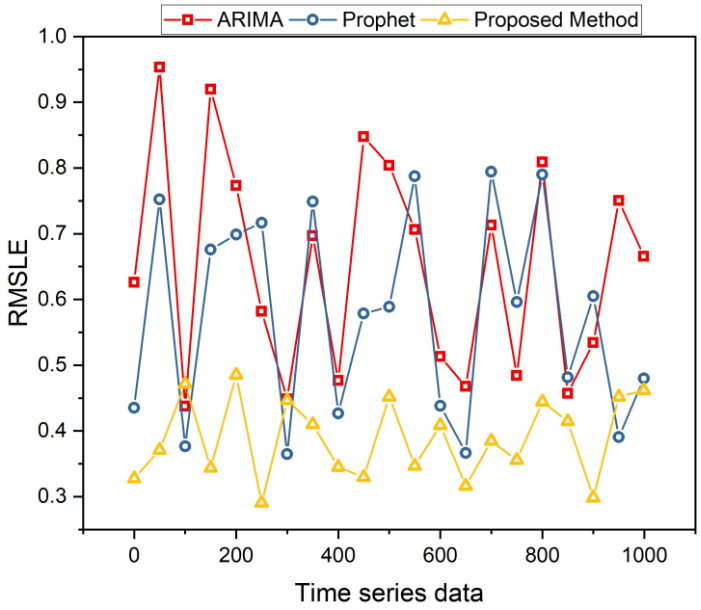


Figure 10. Comparison of RMSLE values between various methods.

The Mean Absolute Error (MAE) for the time-series data, expressed in Equation (17), provides another lens of comparison. As **Figure 12** suggests, the introduced algorithm eclipses the Prophet technique in terms of MAE.

$$MAE = \frac{1}{n} - \sum_{i=1}^{n} |y_i' - y_i| \tag{17}$$

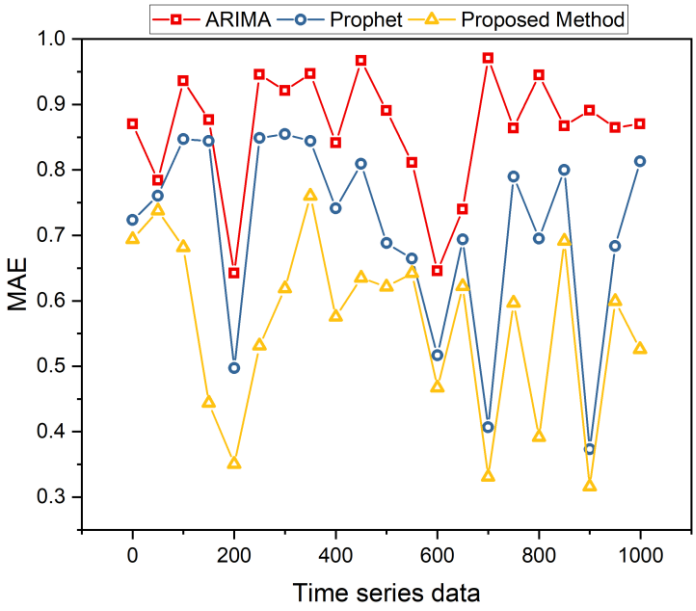


Figure 11. Comparison of MAE values between various methods.

In order to determine the determination coefficient (R^2), proportionate squares of the residuals to the total data are used and **Figure 13** projects the R^2 comparison between ARIMA, Prophet and proposed model. A comprehensive comparison between the ARIMA, Prophet and the proposed model is described in **Table 2** and it is evident that the proposed approach outperforms the other time-series techniques. The proposed model has the highest accuracy at 92.57%, followed by Prophet at 83.81% and ARIMA at 65.79%. Again, the proposed model leads with 94.86%, followed by Prophet at 91.85% and ARIMA at 85.23%. The proposed model has the lowest RMSLE value, indicating better performance in terms of logarithmic error. The proposed model also has the lowest MAE, suggesting that its predictions are closest to the actual values. The proposed model has the highest R^2 value, indicating that it explains the variance in the dependent variable most effectively.

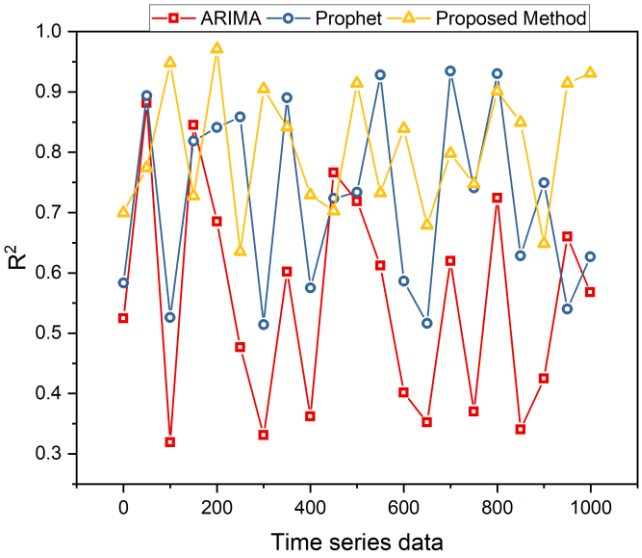


Figure 12. Comparison of R^2 values between various methods.

Table 2. Testing-Performance evaluation between the proposed and Prophet algorithm.

Algorithm	Accuracy (%)	Efficiency (%)	RMSLE (%)	MAE (%)	R^2 (%)
ARIMA	65.79	85.23	0.43	0.64	0.88
Prophet	83.81	91.85	0.36	0.37	0.93
Proposed Model	92.57	94.86	0.29	0.31	0.97

4.2. Case study: SWLS-methane monitoring

The proposed methodology for anomaly detection was employed in an Unmanned Aerial Vehicle (UAV)-based environmental monitoring project over a Solid Waste Landfill Site (SWLS) to measure methane emissions. **Figure 14** showcases the real-time deployment of sensors via UAV, which was flown over a municipal dump yard situated in Tamil Nadu, India. The primary objective of this mission was to monitor methane concentrations above the SWLS region.

Figure 13. Real-time deployment of sensors over a SWLS.

Before embarking on the experiment, a thorough survey of the region was undertaken using industrial-grade equipment. The measurements recorded methane values ranging from 300 to 400 ppm. Both the microwave sensor and the MQ-4 sensor were securely positioned adjacent to each other on the UAV, as visualized in **Figure 15**. Real-time data acquisition via Google Sheets is portrayed in **Figure 16**. This data, which includes methane concentrations as well as temperature and humidity readings, is continuously updated through the NodeMCU system.

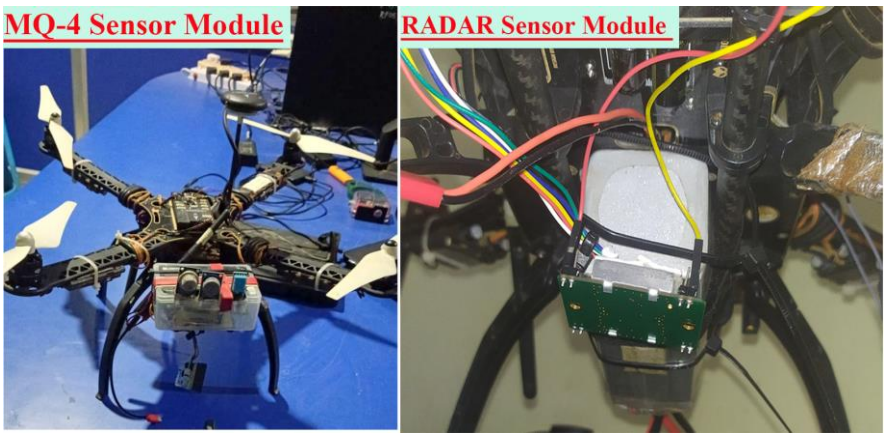
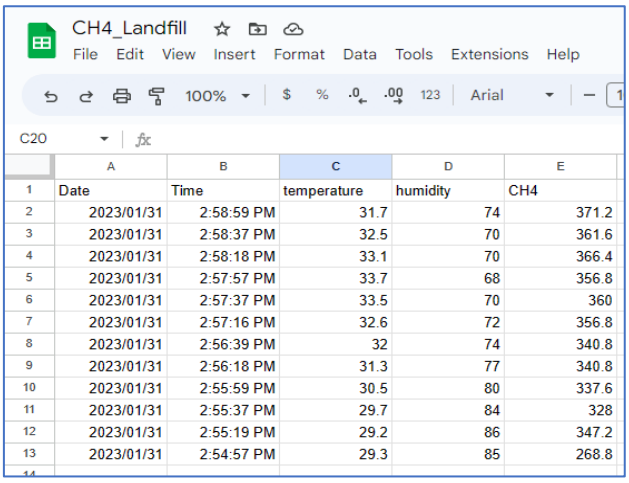


Figure 14. Views of mounted sensor modules in an UAV.



	A	B	C	D	E
1	Date	Time	temperature	humidity	CH4
2	2023/01/31	2:58:59 PM	31.7	74	371.2
3	2023/01/31	2:58:37 PM	32.5	70	361.6
4	2023/01/31	2:58:18 PM	33.1	70	366.4
5	2023/01/31	2:57:57 PM	33.7	68	356.8
6	2023/01/31	2:57:37 PM	33.5	70	360
7	2023/01/31	2:57:16 PM	32.6	72	356.8
8	2023/01/31	2:56:39 PM	32	74	340.8
9	2023/01/31	2:56:18 PM	31.3	77	340.8
10	2023/01/31	2:55:59 PM	30.5	80	337.6
11	2023/01/31	2:55:37 PM	29.7	84	328
12	2023/01/31	2:55:19 PM	29.2	86	347.2
13	2023/01/31	2:54:57 PM	29.3	85	268.8

Figure 15. Screenshot of Real-Time data collection in Google Sheet.

5. Conclusions

This research presents a proficient sensor anomaly detection methodology leveraging a Markov-LSTM architecture tailored for methane sensing. It underscores the burgeoning relevance of ML in anomaly detection, expanding its horizons from conventional uses, such as identifying malicious customer activities, to encompass industrial contexts. The Markov-LSTM model underwent rigorous testing with real-time data and was benchmarked against prominent time-series prediction methods like ARIMA and Prophet. The outcomes distinctly showcase the model's superior performance across diverse evaluation criteria. Specifically, with an impressive accuracy of 92.57%, an efficiency rate of 94.86% and the lowest recorded values for RMSLE and MAE, the model's supremacy is evident. Furthermore, its elevated R^2 value signifies its commendable capability to elucidate the variance within the dependent variable.

The practical utility of the model was further endorsed through a case study centered on monitoring METHANE emissions at a Solid Waste Landfill Site (SWLS) in Tamil Nadu, India. The real-time data collection, facilitated by UAV-based sensor deployment, accentuated the model's tangible benefits in environmental monitoring scenarios.

Thus, the Markov-LSTM anomaly detection framework not only epitomizes precision and efficiency but also stands as a steadfast, reliable solution for industrial endeavors, especially within the ambit of methane emission monitoring. This research, therefore, furnishes a meaningful augmentation to the domain, setting the stage for ensuing explorations and pragmatic deployments of machine learning techniques in sensor anomaly detection.

Author contributions

Conceptualization, SVK; methodology, SVK, GAAM, JSC and KK; software, SVK and KK; validation, SVK; formal analysis, SVK; investigation, SVK; resources, SVK; data curation, SVK; writing—original draft preparation, SVK; writing—review and editing, SVK, GAAM, JSC and KK; visualization, SVK and KK; supervision, GAAM; project administration, SVK; funding acquisition, SVK and KK. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no conflict of interest.

References

1. Jo JY, Kwon YS, Lee JW, et al. Acute Respiratory Distress Due to Methane Inhalation. Tuberculosis and Respiratory Diseases. 2013, 74(3): 120. doi: 10.4046/trd.2013.74.3.120
2. Wu R, Tian L, Li H, et al. A selective methane gas sensor based on metal oxide semiconductor equipped with an on-chip microfilter. Sensors and Actuators B: Chemical. 2022, 359: 131557. doi: 10.1016/j.snb.2022.131557
3. Halley S, Tsui L kun, Garzon F. Combined Mixed Potential Electrochemical Sensors and Artificial Neural Networks for the Quantification and Identification of Methane in Natural Gas Emissions Monitoring. Journal of The Electrochemical Society. 2021, 168(9): 097506. doi: 10.1149/1945-7111/ac2465
4. Sun J, Tian L, Chang J, et al. Adaptively Optimized Gas Analysis Model with Deep Learning for Near-Infrared Methane Sensors. Analytical Chemistry. 2022, 94(4): 2321-2332. doi: 10.1021/acs.analchem.1c05059
5. Iwaszenko S, Kalisz P, Słota M, et al. Detection of Natural Gas Leakages Using a Laser-Based Methane Sensor and UAV. Remote Sensing. 2021, 13(3): 510. doi: 10.3390/rs13030510
6. Chai H, Zheng Z, Liu K, et al. Stability of Metal Oxide Semiconductor Gas Sensors: A Review. IEEE Sensors Journal. 2022, 22(6): 5470-5481. doi: 10.1109/jsen.2022.3148264
7. Balestrieri E, Daponte P, De Vito L, et al. Sensors and Measurements for UAV Safety: An Overview. Sensors. 2021, 21(24): 8253. doi: 10.3390/s21248253
8. Himeur Y, Ghanem K, Alsalemi A, et al. Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. Applied Energy. 2021, 287: 116601. doi: 10.1016/j.apenergy.2021.116601
9. Shagari NM, Idris MYI, Salleh RB, et al. Heterogeneous Energy and Traffic Aware Sleep-Awake Cluster-Based Routing Protocol for Wireless Sensor Network. IEEE Access. 2020, 8: 12232-12252. doi: 10.1109/access.2020.2965206
10. Ahmed G, Zou J, Fareed MMS, et al. Sleep-awake energy efficient distributed clustering algorithm for wireless sensor networks. Computers & Electrical Engineering. 2016, 56: 385-398. doi: 10.1016/j.compeleceng.2015.11.011
11. Sharma D, Bhondekar AP. Traffic and Energy Aware Routing for Heterogeneous Wireless Sensor Networks. IEEE Communications Letters. 2018, 22(8): 1608-1611. doi: 10.1109/lcomm.2018.2841911
12. Mahima V, Chitra A. Battery Recovery Based Lifetime Enhancement (BRLE) Algorithm for Wireless Sensor Network. Wireless Personal Communications. 2017, 97(4): 6541-6557. doi: 10.1007/s11277-017-4854-3
13. M T, Thangaraj P. Fuzzy Ontology for Distributed Document Clustering based on Genetic Algorithm. Applied Mathematics & Information Sciences. 2013, 7(4): 1563-1574. doi: 10.12785/amis/070442
14. Duangsuwan S, Prapruetdee P, Subongkod M, et al. 3D AQI Mapping Data Assessment of Low-Altitude Drone Real-Time Air Pollution Monitoring. Drones. 2022, 6(8): 191. doi: 10.3390/drones6080191
15. Cozma A, Firculescu AC, Tudose D, et al. Autonomous Multi-Rotor Aerial Platform for Air Pollution Monitoring. Sensors. 2022, 22(3): 860. doi: 10.3390/s22030860
16. Sonkar SK, Kumar P, George RC, et al. Detection and Estimation of Natural Gas Leakage Using UAV by Machine Learning Algorithms. IEEE Sensors Journal. 2022, 22(8): 8041-8049. doi: 10.1109/jsen.2022.3157872
17. Wang D, Pan J, Huang X, et al. Virtual Alternating Current Measurements Advance Semiconductor Gas Sensors' Performance in the Internet of Things. IEEE Internet of Things Journal. 2022, 9(7): 5502-5510. doi: 10.1109/jiot.2021.3108799
18. Luo P, Harrist J, Menduni G, et al. Simultaneous Detection of Methane, Ethane, and Propane by QEPAS Sensors for On-Site Hydrocarbon Characterization and Production Monitoring. ACS Omega. 2022, 7(4): 3395-3406. doi: 10.1021/acsomega.1c05645
19. Contreras-Castillo J, Zeadally S, Guerrero Ibáñez JA. A seven-layered model architecture for Internet of Vehicles. Journal of Information and Telecommunication. 2017, 1(1): 4-22. doi: 10.1080/24751839.2017.1295601
20. Martín D, Fuentes-Lorenzo D, Bordel B, et al. Towards Outlier Sensor Detection in Ambient Intelligent Platforms—A Low-Complexity Statistical Approach. Sensors. 2020, 20(15): 4217. doi: 10.3390/s20154217
21. Chen Z, Yeo CK, Lee BS, et al. Autoencoder-based network anomaly detection. 2018 Wireless Telecommunications Symposium (WTS). Published online April 2018. doi: 10.1109/wts.2018.8363930
22. Jiang J, Liu F, Liu Y, et al. A dynamic ensemble algorithm for anomaly detection in IoT imbalanced data streams. Computer Communications. 2022, 194: 250-257. doi: 10.1016/j.comcom.2022.07.034
23. Siami-Namini S, Tavakoli N, Namin AS. The Performance of LSTM and BiLSTM in Forecasting Time Series. 2019 IEEE International Conference on Big Data (Big Data). Published online December 2019. doi: 10.1109/bigdata47090.2019.9005997
24. Kim J, Shin J, Park KW, et al. Improving Method of Anomaly Detection Performance for Industrial IoT Environment. Computers, Materials & Continua. 2022, 72(3): 5377-5394. doi: 10.32604/cmc.2022.026619
25. Yang X, Li Y, Shen J. Forecasting Research on Long-term Solar Irradiance with An Improved Prophet Algorithm. IFAC-PapersOnLine. 2022, 55(9): 491-494. doi: 10.1016/j.ifacol.2022.07.085

26. Arronte Alvarez A, Gómez F. Motivic Pattern Classification of Music Audio Signals Combining Residual and LSTM Networks. International Journal of Interactive Multimedia and Artificial Intelligence. 2021, 6(6): 208. doi: 10.9781/ijimai.2021.01.003
27. Gökdemr A, Çalhan A. Deep learning and machine learning based anomaly detection in internet of things environments (Turkish). Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi. 2022, 37(4): 1945-1956. doi: 10.17341/gazimmfd.962375
28. Zhou C, Paffenroth RC. Anomaly Detection with Robust Deep Autoencoders. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Published online August 4, 2017. doi: 10.1145/3097983.3098052
29. Xue H, Huynh DQ, Reynolds M. SS-LSTM: A Hierarchical LSTM Model for Pedestrian Trajectory Prediction. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Published online March 2018. doi: 10.1109/wacv.2018.00135
30. Siami-Namini S, Tavakoli N, Siami Namin A. The performance of LSTM and BiLSTM in forecasting time series. In: Proceedings of 2019 IEEE International Conference on Big Data (Big Data); 9–12 December 2019; Los Angeles, CA, USA. doi:10.1109/BigData47090.2019.9005997
31. Zar Zar OO, Sabai P. Time Series Prediction Based on Facebook Prophet: A Case Study, Temperature Forecasting in Myintkyina. International Journal of Applied Mathematics Electronics and Computers. 2020, 8(4): 263-267. doi: 10.18100/ijamec.816894
32. Almanjahie IM, Kaid Z, Laksaci A, et al. Predicting temperature curve based on fastkNN local linear estimation of the conditional distribution function. PeerJ. 2021, 9: e11719. doi: 10.7717/peerj.11719
33. Sareminia S. A Support Vector Based Hybrid Forecasting Model for Chaotic Time Series: Spare Part Consumption Prediction. Neural Processing Letters. 2022, 55(3): 2825-2841. doi: 10.1007/s11063-022-10986-4
34. Khansa HE, Gervet C, Brouillet A. On the Ranking of Variable Length Discords Through a Hybrid Outlier Detection Approach. Lecture Notes in Computer Science. Published online 2022: 329-344. doi: 10.1007/978-3-031-18840-4_24
35. Garg S, Kaur K, Kumar N, et al. A Hybrid Deep Learning-Based Model for Anomaly Detection in Cloud Datacenter Networks. IEEE Transactions on Network and Service Management. 2019, 16(3): 924-935. doi: 10.1109/tnsm.2019.2927886
36. Garg S, Kaur K, Batra S, et al. A multi-stage anomaly detection scheme for augmenting the security in IoT-enabled applications. Future Generation Computer Systems. 2020, 104: 105-118. doi: 10.1016/j.future.2019.09.038
37. Moustafa N, Hu J, Slay J. A holistic review of Network Anomaly Detection Systems: A comprehensive survey. Journal of Network and Computer Applications. 2019, 128: 33-55. doi: 10.1016/j.jnca.2018.12.006
38. Carrillo-Amado YR, Califa-Urquiza MA, Ramón-Valencia JA. Calibration and standardization of air quality measurements using MQ sensors. Respuestas. 2020, 25(1): 70-77. doi: 10.22463/0122820x.2408
39. Alvarez-Alvarado MS, Jayaweera D. Bathtub curve as a Markovian process to describe the reliability of repairable components. IET Generation, Transmission & Distribution. 2018, 12(21): 5683-5689. doi: 10.1049/iet-gtd.2018.5505
40. Kumar VS, Mary AAG, Esakki B. Analyzing the Applicability of ML Powered Microwave Sensor for UAV Based CH4 Sensing. International Journal of Electrical and Electronic Engineering & Telecommunications. Published online 2023: 161-170. doi: 10.18178/ijeetc.12.3.161-170
41. Wang B, Chen Y, Liu D, et al. An embedded intelligent system for on-line anomaly detection of unmanned aerial vehicle. Journal of Intelligent & Fuzzy Systems. 2018, 34(6): 3535-3545. doi: 10.3233/jifs-169532
42. Cauteruccio F, Cinelli L, Corradini E, et al. A framework for anomaly detection and classification in Multiple IoT scenarios. Future Generation Computer Systems. 2021, 114: 322-335. doi: 10.1016/j.future.2020.08.010
43. Aljuhani A. Machine Learning Approaches for Combating Distributed Denial of Service Attacks in Modern Networking Environments. IEEE Access. 2021, 9: 42236-42264. doi: 10.1109/access.2021.3062909
44. Pourhabibi T, Ong KL, Kam BH, et al. Fraud detection: A systematic literature review of graph-based anomaly detection approaches. Decision Support Systems. 2020, 133: 113303. doi: 10.1016/j.dss.2020.113303