Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

DOI: 10.37943/APTT7109

Anuar Kussainov
PhD Student, School of Digital Technologies and Artificial Intelligence
ak@kab.kz, orcid.org/0009-0005-9962-7237
D. Serikbayev East Kazakhstan Technical University, Kazakhstan
Gulnaz Zhomartkyzy
PhD, Associate Professor, Department of Information Systems
gzhomartkyzy@edu.ektu.kz, orcid.org/0000-0003-1465-3451
D. Serikbayev East Kazakhstan Technical University, Kazakhstan
Rajermani Thinakaran
PhD, Associate Professor, Faculty of Information Technology
rajermani@yahoo.com, orcid.org/0009-0007-2677-3406
INTI International University, Malaysia

INTELLIGENT MODEL FOR PREDICTING METHANE CONCENTRATION IN INDUSTRIAL ENVIRONMENTS BASED ON INTERNET OF THINGS GAS ANALYZERS AND HYBRID MACHINE LEARNING ALGORITHMS

Abstract: Ensuring industrial safety at enterprises that use combustible and toxic gases requires continuous monitoring of the gas environment and timely identification of potentially hazardous situations. One of the most serious threats in the mining and industrial sectors is the accumulation of methane, which has a high explosion hazard and can lead to technogenic accidents. Traditional monitoring systems based on threshold alarm algorithms allow detection of exceedances of permissible gas concentrations; however, they do not provide the capability to forecast the development of hazardous situations.

This paper proposes an intelligent model for predicting methane concentration based on data obtained from IoT gas analyzers of the SENSOR - Mine 4GN series. The monitoring system integrates a distributed network of gas sensors, wireless data transmission, and machine learning algorithms for time-series analysis of gas concentration measurements.

Experimental data were obtained under laboratory conditions over a ten-day period using certified calibration gas mixtures simulating methane release events under different temperature conditions. For time-series analysis, a hybrid machine learning model was applied, combining a Long Short-Term Memory (LSTM) recurrent neural network and the ensemble algorithm XGBoost.

The obtained results demonstrate that the proposed model improves the accuracy of methane concentration dynamics prediction and enables early detection of hazardous trends in the gas environment. The use of predictive analytics methods combined with Internet of Things (IoT) gas analyzers significantly increases the effectiveness of industrial safety systems and reduces the risk of accidents at industrial facilities. The proposed methodology can be applied as a foundation for the development of intelligent predictive monitoring systems for mining, energy, and other hazardous industrial environments.

Keywords: Methane monitoring, Internet of Things, sensors, gas analyzer, predictive analytics, machine learning, Long Short-Term Memory (LSTM), XGBoost, industrial safety, gas environment monitoring.

Introduction

Ensuring safe working conditions at industrial enterprises remains one of the key challenges of modern industry. According to the International Labour Organization, more than 2.3 million fatal incidents related to occupational risks are recorded worldwide each year, along with over 300 million cases of work-related injuries and occupational diseases [1]. A significant

Received: 09.06.2026 Accepted: 17.07.2026 Published: 30.06.2026

proportion of these incidents is associated with exposure to hazardous and toxic gases in the industrial atmosphere.

Particularly high risks of emergency situations are observed in the mining industry, where technological processes are accompanied by the release of various gases, including methane (CH_4), carbon monoxide (CO), hydrogen sulfide (H_2S), and nitrogen oxides [2]. In underground workings and environments with limited ventilation, the concentrations of such gases can rapidly increase, posing a serious threat to workers' health and the stability of technological processes.

Methane is one of the most hazardous gases in the mining sector. When its concentration in air reaches the range of 5% to 15%, an explosive mixture is formed, capable of causing destructive explosions and fires [3]. Even relatively small methane leaks can lead to the formation of localized zones with elevated gas concentrations, significantly increasing the risk of accidents.

Traditional gas monitoring systems used at industrial facilities are typically based on stationary sensors and threshold alarm algorithms. Such systems are capable of detecting exceedances of maximum permissible concentrations (MPC); however, they do not provide the capability to forecast the development of hazardous situations [4]. In most cases, alarm systems are triggered only after the gas concentration reaches a dangerous level, which significantly limits the ability to prevent accidents.

The development of Internet of Things (IoT) technologies has significantly expanded the capabilities of gas environment monitoring. Internet of Things sensors can generate continuous time series of gas concentration measurements and transmit data in real time to centralized data processing systems [5]. The use of wireless communication technologies, such as LoRaWAN, enables monitoring of large industrial areas while maintaining low power consumption of sensing devices [6].

In recent years, considerable attention has been paid to the application of machine learning methods for analyzing data obtained from sensor systems. Machine learning algorithms enable the identification of hidden patterns in data, detection of anomalies, and prediction of the development of potentially hazardous situations [7]. Of particular interest are recurrent neural networks, especially the Long Short-Term Memory (LSTM) architecture, which is widely used for time-series analysis and forecasting the dynamics of system parameters [8].

In addition to neural networks, ensemble machine learning methods are widely applied in the analysis of industrial data. One of the most effective algorithms in this class is XGBoost, which is based on gradient boosting of decision trees [9]. This algorithm demonstrates high accuracy when dealing with nonlinear relationships and noisy data typical of sensor measurements.

Despite the rapid development of intelligent monitoring technologies, most existing studies focus either on the development of sensor systems or on data analysis separately. Only a limited number of works address the integration of monitoring hardware with predictive analytics algorithms within a unified industrial safety system.

In this study, an intelligent model for predicting methane concentration based on data obtained from «SENSOR - Mine 4GN» gas analyzers is proposed. The system integrates IoT gas analyzers, a centralized data storage infrastructure, and machine learning algorithms for the analysis of gas concentration time series.

The main objective of this research is to develop and experimentally validate a methane concentration prediction model based on the hybrid application of LSTM and XGBoost algorithms, enabling the detection of hazardous trends in the gas environment and providing early warning of potential industrial accidents.

Literature Review

The development of digital technologies and data analysis methods has led to the widespread adoption of intelligent monitoring systems in industrial environments. One of the most promising directions is the use of distributed Internet of Things (IoT) sensor networks for monitoring environmental parameters at industrial facilities. Such systems enable real-time

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

acquisition of gas concentration data and support rapid response to hazardous changes in air quality parameters [5, 10].

Recent studies indicate that the use of IoT-based sensor networks significantly improves the efficiency of monitoring industrial sites. Several researchers have proposed air quality monitoring systems based on wireless sensors and cloud-based data processing platforms [5,10]. These systems enable the generation of continuous time-series data and allow detailed analysis of gas concentration dynamics.

Special attention has been given to wireless data transmission technologies. Several studies demonstrate that LoRaWAN technology is one of the most effective solutions for industrial sensor networks due to its long communication range and low energy consumption of devices [11]. The use of LoRaWAN enables the deployment of scalable monitoring systems capable of covering large industrial areas, including mines, quarries, and underground facilities.

In recent years, significant progress has been made in the field of intelligent analysis of data obtained from sensor systems. Machine learning techniques enable the identification of hidden patterns in datasets and allow prediction of potentially hazardous situations. According to several studies, the application of machine learning algorithms significantly improves the accuracy of gas environment analysis and reduces the probability of false alarms in monitoring systems [12].

One of the most common approaches to analyzing sensor data involves regression methods and ensemble machine learning models. Algorithms such as Random Forest and Gradient Boosting have demonstrated high efficiency in processing datasets characterized by nonlinear relationships and noise [13]. These approaches allow the development of predictive models for gas concentration levels based on historical measurements.

Particular attention has been given to the XGBoost algorithm, which represents one of the most efficient implementations of gradient boosting. Studies show that this algorithm provides high prediction accuracy in tasks involving complex multidimensional datasets [14]. By applying regularization techniques and optimized decision tree structures, XGBoost effectively handles large data volumes while reducing the risk of model overfitting.

In addition to ensemble learning methods, neural networks are widely applied in time-series analysis. In particular, Recurrent Neural Networks (RNN) are commonly used for analyzing sequential data and predicting the dynamics of system parameters [15]. One of the most effective architectures within this class is the Long Short-Term Memory (LSTM) network proposed by Hochreiter and Schmidhuber [16]. This architecture is capable of capturing long-term dependencies in datasets and has proven highly effective for time-series forecasting tasks.

In recent years, LSTM networks have been widely applied in the analysis of data generated by IoT sensors. Several studies demonstrate that the application of LSTM significantly improves the accuracy of predicting concentrations of various gases in industrial and environmental monitoring systems [17]. These neural networks can detect complex temporal patterns and forecast future values of system parameters based on historical data.

Some studies propose hybrid approaches that combine neural networks with ensemble machine learning methods. Such hybrid models allow the advantages of different algorithms to be leveraged simultaneously, resulting in improved prediction accuracy [18]. For example, the combination of LSTM networks with gradient boosting methods enables effective analysis of both temporal dependencies and complex nonlinear relationships between system parameters.

Despite the significant progress achieved in intelligent monitoring technologies, many existing studies focus either on the development of sensor hardware or on data analysis independently, without addressing the integration of hardware and software components within a unified industrial safety system. Furthermore, a considerable portion of research is oriented toward environmental monitoring and urban air quality control, whereas industrial safety applications require consideration of specific operational conditions and constraints.

Therefore, an important research challenge is the development of integrated gas environment monitoring systems that combine IoT gas analyzers, data storage infrastructures, and predictive analytics algorithms. Such systems enable not only the detection of gas concentration exceedances but also the forecasting of hazardous situations at early stages.

In this study, an approach for predicting methane concentration based on data obtained from «SENSOR - Mine 4GN» IoT gas analyzers is proposed. The approach employs a hybrid machine learning model combining the LSTM recurrent neural network and the XGBoost algorithm. The proposed method aims to improve the accuracy of predicting gas environment dynamics and reduce the risk of accidents at industrial facilities.

Methods and Materials

Architecture of the Gas Environment Monitoring System: Modern industrial safety systems require continuous monitoring of gas environment parameters and rapid detection of potentially hazardous situations. To address this challenge, distributed monitoring systems are used that integrate sensor devices, wireless data transmission channels, and software-based intelligent data analysis methods.

In this study, a gas environment monitoring system was developed based on IoT gas analyzers of the SENSOR - Mine 4GN series and predictive analytics algorithms. The main objective of the system is continuous monitoring of gas concentrations and forecasting hazardous changes in air environment parameters at industrial facilities.

The overall architecture of the proposed gas monitoring and predictive analytics system is presented in Figure 1. The system integrates IoT-based gas sensing devices, wireless communication technologies, centralized data storage, and machine learning modules for real-time methane concentration forecasting and industrial safety monitoring.

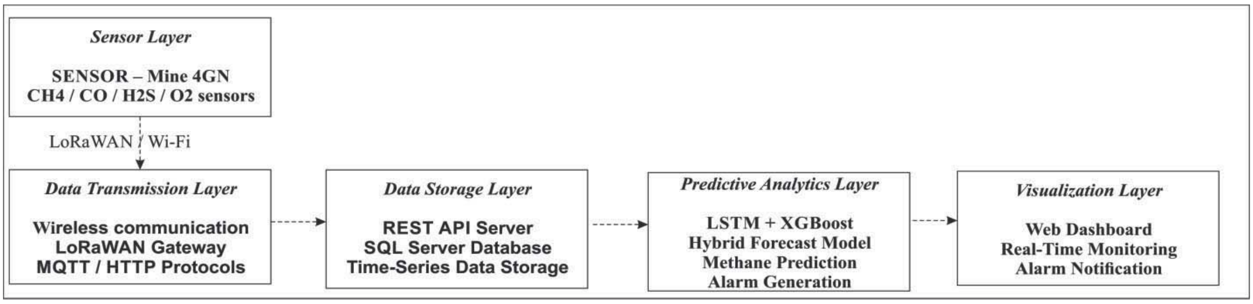

Figure 1. Architecture of the proposed IoT-based gas monitoring and predictive analytics system.

As shown in Figure 1, gas concentration measurements are collected by the SENSOR–Mine 4GN gas analyzer equipped with CH_4, CO, H_2S, O_2, temperature, and humidity sensors. The acquired measurements are processed by the embedded ESP32 controller and transmitted through LoRaWAN or Wi-Fi communication channels using MQTT or HTTP protocols. The collected data are received by a REST API server and stored in a centralized SQL Server database for long-term time-series data management.

The predictive analytics module applies machine learning algorithms, including LSTM and XGBoost, to analyze temporal patterns and forecast methane concentration dynamics. Forecasting results, alarm notifications, and real-time measurements are displayed through a web-based monitoring dashboard, supporting early warning and decision-making processes in industrial safety systems.

Sensor Layer of the System: The sensor layer of the system is represented by SENSOR - Mine 4GN gas analyzers, designed for monitoring gas concentrations in industrial atmospheres.

Each gas analyzer includes the following components: gas sensors, signal processing microcontroller, wireless data transmission module, alarm and notification system.

The sensing unit is implemented as a multi-component gas sensor module capable of simultaneously measuring the concentrations of several gases. The performance of gas sensor

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

modules depends on calibration quality, environmental conditions, and sensor drift; therefore, proper sensor selection and calibration are essential for reliable industrial gas monitoring [19]. The module contains four sensors designed to measure the concentrations of: methane (CH_4), carbon monoxide (CO), hydrogen sulfide (H_2S), oxygen (O_2).

The sensor module performs real-time gas concentration measurements and transmits the data through a digital serial interface. The main technical characteristics of the sensor module are presented in Table 1.

Table 1. Technical characteristics of the gas analyzer sensor module

Parameter	**CH_4**	**CO**	**H_2S**	**O_2**
Measurement range	0 - 100 % LEL	0 - 1000 ppm	0 - 100 ppm	0 - 30 %
Resolution	1 % LEL	1 ppm	1 ppm	0.1 %
Response time	<30 s	<30 s	<30 s	<30 s
Operating temperature	−20…50 °C			
Relative humidity	15 - 90 % RH			

The use of a multi-component sensor module allows simultaneous monitoring of several gas environment parameters, which is an important factor in the analysis of complex industrial processes.

Data Transmission Layer: The transmission of data from gas analyzers to the monitoring server is carried out using wireless communication channels.

The system employs two main communication technologies: LoRaWAN for long-distance data transmission and Wi-Fi for local device connection and configuration.

The use of LoRaWAN technology enables the creation of distributed monitoring networks capable of covering large industrial areas. The main advantages of this technology include: low power consumption of devices, long communication range, high resistance to electromagnetic interference.

When an exceedance of permissible gas concentration levels is detected, the device automatically transmits an alarm signal to the monitoring system server.

Data Storage Layer: All measured gas concentration values are transmitted to the monitoring system server and stored in a database. Each database record contains information about the measurement parameters and the state of the gas environment. The structure of the stored data is presented in Table 2.

Table 2. Structure of gas environment monitoring data

Parameter	**Description**
Sensor_ID	Unique identifier of the gas analyzer
Gas_type	Type of measured gas
Concentration	Gas concentration value
Timestamp	Measurement time
Temperature	Ambient temperature
Humidity	Relative humidity
Status	System status (normal, warning, alarm)

The records generated in the database form time series of gas concentration measurements. Formally, the measurement stream can be represented as a time series:

$$X_t = (CH_4, CO, H_2S, O_2, T, H) \tag{1}$$

where CH_4 - methane concentration, CO - carbon monoxide concentration, H_2S - hydrogen sulfide concentration, O_2 - oxygen concentration, T - ambient temperature, H - relative humidity.

The resulting time-series data are used for further analysis and prediction of the dynamics of gas environment parameters.

Intelligent Data Analysis Layer: To analyze time-series measurement data, a predictive analytics system based on machine learning algorithms is applied. In this study, a hybrid prediction model is used, combining the following approaches: Long Short-Term Memory (LSTM) recurrent neural network, XGBoost ensemble learning algorithm.

The use of LSTM enables the model to capture temporal dependencies in the data and predict future values of gas concentrations. This architecture is particularly effective for analyzing sequential sensor measurements and identifying patterns in time-series data.

The XGBoost algorithm is applied to detect complex nonlinear relationships between system parameters. Due to its gradient boosting framework and efficient handling of structured data, XGBoost demonstrates high predictive performance when processing sensor measurements that may contain noise and nonlinear dependencies.

The combined application of these algorithms improves the accuracy of methane concentration prediction and enables early detection of hazardous changes in the gas environment.

Mathematical Model for Methane Concentration Prediction. Sensor-based gas environment monitoring systems generate continuous data streams representing time series of gas concentration measurements. To analyze such data, time-series processing methods and machine learning algorithms are applied, enabling the identification of hidden patterns in the dynamics of system parameters.

In this study, the task of predicting methane concentration is considered based on data obtained from IoT gas analyzers.

Problem Formalization: Let the monitoring system generate a sequence of gas concentration measurements over time. The measurement vector at time ttt can be represented as:

$$X_t = (CH_4^t, CO^t, H_2S^t,\ O_2^t, T^t, H^t) \tag{2}$$

where CH_4^t - methane concentration, CO^t - carbon monoxide concentration, H_2S^t - hydrogen sulfide concentration, O_2^t - oxygen concentration, T^t - ambient temperature, H^t - relative humidity.

The prediction task consists of estimating the future value of methane concentration based on previous measurements:

$$CH_4^{t+\Delta t} = f(X_t, X_{t-1}, \dots, X_{t-n}) \tag{3}$$

where $CH_4^{t+\Delta t}$ - predicted methane concentration value at time.

Sensor Data Filtering. Sensor measurements often contain noise caused by temperature variations, turbulence of air flows, and electrical interference. To reduce the influence of such noise, a data smoothing method based on the moving average is applied. The moving average is calculated as follows:

$$MA_k(t) = \frac{1}{k}\sum_{i=0}^{k-1} x_{t-i} \tag{4}$$

where k - size of the averaging window. This method makes it possible to reduce short-term fluctuations in measurements and improve the stability of the resulting time series.

Temperature Compensation of Sensor Measurements. One of the characteristic features of gas sensors is the dependence of their sensitivity on ambient temperature. Temperature variations may cause a shift in the baseline signal level and alter the sensor sensitivity coefficient, which ultimately affects the accuracy of gas concentration measurements.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

To reduce the influence of temperature effects, the monitoring system applies a correction of measured gas concentration values based on the ambient temperature. The corrected methane concentration can be expressed as follows:

$$C_{corr} = C_{\mathrm{meas}} \times (1 + \alpha(\mathrm{T} - T_0)) \quad (5)$$

where C_{corr} - temperature-compensated gas concentration, C_{meas} - measured gas concentration value, T - current ambient temperature, T_0 - reference temperature of the sensor, α - temperature sensitivity coefficient of the sensor.

The application of temperature compensation reduces systematic measurement errors and improves the accuracy of gas concentration dynamics analysis. The corrected concentration values are subsequently used in the formation of time series and in the training of machine learning models.

LSTM-Based Prediction Model. To analyze time series of gas concentrations, a Long Short-Term Memory (LSTM) recurrent neural network is employed. An LSTM network consists of a set of memory cells, each containing several control mechanisms known as gates, which regulate the flow of information through the network.

The input gate is defined as:

$$i_t = \sigma(W_i x_t + U_i H_{t-1} + b_i) \quad (6)$$

where x_t - input data vector, h_{t-1} - hidden state from the previous time step, W_i, U_i - weight matrices.

The forget gate is computed as:

$$\mathrm{f_t} = \sigma(\mathrm{W_f X_t} + \mathrm{U_f H_{t-1}} + \mathrm{b_f}) \quad (7)$$

The cell state is updated according to the following expression:

$$\mathrm{C_t} = \mathrm{f_t C_{t-1}} + \mathrm{i_t \widetilde{C_t}} \quad (8)$$

The output of the network is determined by:

$$\mathrm{h_t} = \mathrm{o_t \tanh(C_t)} \quad (9)$$

where o_t - output gate. The use of the LSTM architecture enables the model to capture long-term dependencies in time-series data and effectively predict the dynamics of gas concentration changes.

XGBoost Model. To analyze nonlinear relationships between system parameters, the Extreme Gradient Boosting (XGBoost) algorithm is employed. The XGBoost model represents an ensemble of decision trees, where the final prediction is obtained as the sum of predictions from individual trees:

$$\widehat{y_i} = \sum_{k=1}^{K} f_k(x_i) \quad (10)$$

where f_k - decision tree in the ensemble.

The optimization objective of the model is defined as:

$$L = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (11)$$

where l - loss function, Ω - model regularization term.

The use of regularization reduces the risk of model overfitting and improves the stability and generalization ability of the predictions.

Hyperparameter Configuration. The predictive performance of machine learning models strongly depends on the selection of appropriate hyperparameters. In this study, the hyperparameters of the LSTM and XGBoost models were selected empirically based on preliminary experiments and validation performance.

For the LSTM model, two hidden layers with 64 units each were used to capture temporal dependencies in methane concentration dynamics. The Adam optimizer was applied with a learning rate of 0.001. A dropout rate of 0.2 was introduced to reduce overfitting. The batch size was set to 32, and the model was trained for 100 epochs using the mean squared error (MSE) loss function.

For the XGBoost model, the number of trees was set to 500, the maximum tree depth to 6, and the learning rate to 0.05. Subsample and colsample_bytree parameters were both set to 0.8 to improve model generalization and reduce overfitting.

The selected hyperparameters provided stable convergence during training and achieved the best predictive performance on the validation dataset.

The main hyperparameters used in the LSTM and XGBoost models are summarized in Table 3.

Table 3. Hyperparameters of the LSTM and XGBoost models

Model	Hyperparameter	Value
LSTM	Number of hidden layers	2
LSTM	Hidden units per layer	64
LSTM	Dropout rate	0.2
LSTM	Activation function	tanh
LSTM	Optimizer	Adam
LSTM	Learning rate	0.001
LSTM	Batch size	32
LSTM	Epochs	100
LSTM	Loss function	Mean Squared Error (MSE)
XGBoost	Number of trees (n_estimators)	500
XGBoost	Maximum tree depth (max_depth)	6
XGBoost	Learning rate (eta)	0.05
XGBoost	Subsample ratio	0.8
XGBoost	Column sample ratio (colsample_bytree)	0.8
XGBoost	Regularization parameter (lambda)	1.0
XGBoost	Objective function	reg:squarederror

The selected hyperparameters were determined based on validation performance and model stability during the training process. These settings provided the best balance between forecasting accuracy and generalization capability.

Hybrid Prediction Model. In this study, the hybrid prediction model was implemented as a two-stage forecasting approach. At the first stage, the LSTM neural network was used to model temporal dependencies in methane concentration time series. The LSTM model generated the preliminary prediction of methane concentration for the selected forecast horizon.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

At the second stage, the XGBoost algorithm was used to correct the residual prediction error of the LSTM model. The residual error was calculated as the difference between the actual methane concentration and the value predicted by the LSTM model:

$$e_t = y_t - \hat{y}_t^{LSTM} \quad (12)$$

where e_t is the residual error at time t, y_t is the actual methane concentration, and $\hat{y}_t^{LSTM}$ is the methane concentration predicted by the LSTM model.

The XGBoost model was trained using the input features of the gas environment and the residual errors obtained from the LSTM model. Therefore, the task of XGBoost was not to independently predict methane concentration, but to estimate the correction term for the LSTM forecast.

The final prediction of the hybrid model was calculated as follows:

$$\hat{y}_t^{Hybrid} = \hat{y}_t^{LSTM} + \hat{y}_t^{XGB} \quad (13)$$

where $\hat{y}_t^{Hybrid}$ is the final predicted methane concentration, $\hat{y}_t^{LSTM}$ is the preliminary LSTM prediction, and $\hat{y}_t^{XGB}$ is the correction value predicted by XGBoost.

This hybrid structure allows the model to combine the advantages of both approaches: LSTM captures temporal dependencies in sequential sensor data, while XGBoost corrects nonlinear residual errors associated with temperature, humidity, and multicomponent gas concentration changes. As shown in Algorithm 1, the proposed hybrid model first generates a preliminary methane concentration forecast using the LSTM network. The residual prediction error is then modeled using XGBoost and applied as a correction term to improve the final forecasting accuracy.

Algorithm 1. Hybrid LSTM–XGBoost Methane Concentration Prediction

Input: X = {CH_4, CO, H_2S, O_2, Temperature, Humidity}
Output: Predicted methane concentration
1. Data preprocessing
   - Remove outliers
   - Handle missing values
   - Apply moving-average filtering
2. Data normalization
3. Formation of time-series sequences
4. Train LSTM model
5. Generate preliminary forecast
6. Compute residual error
7. Train XGBoost model on residuals
8. Calculate residual correction
9. Generate final hybrid forecast
10. Evaluate MAE, RMSE and R^2

The workflow of the proposed hybrid prediction model is illustrated in Figure 2. The model combines the temporal sequence learning capability of LSTM networks with the residual error correction mechanism of XGBoost. First, the LSTM model generates a preliminary forecast of methane concentration. Then, the residual prediction error is estimated and used as the target variable for the XGBoost model. Finally, the residual correction is added to the LSTM forecast to obtain the final hybrid prediction.

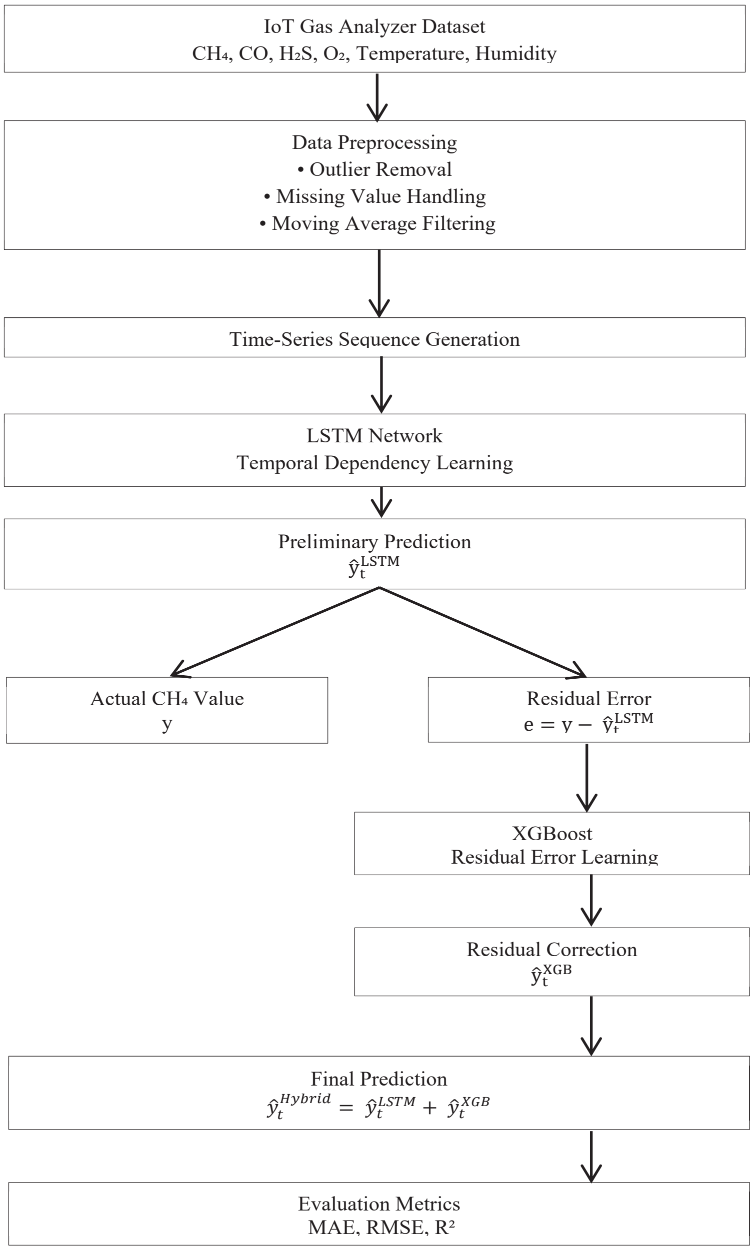

Figure 2. Architecture of the proposed Hybrid LSTM–XGBoost model for methane concentration prediction.

Experimental Study and Dataset Formation. To evaluate the effectiveness of the proposed methane concentration prediction model, experimental studies were conducted under laboratory conditions. The primary objective of the experiment was to generate time-series measurements of gas concentrations under various environmental conditions, including controlled methane releases and temperature variations.

The experiment was carried out using SENSOR - Mine 4GN gas analyzers, designed for monitoring gas environment parameters in industrial conditions. These gas analyzers provide continuous measurement of multiple gas concentrations and transmit data to the monitoring system server.

The devices were placed inside a transparent sealed module designed to simulate gas environment conditions in a confined space.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

The transparent module functioned as an experimental chamber, structurally similar to a laboratory gas reactor and constructed from transparent material. The geometry of the module allowed visual observation of gas mixture injection and the distribution of gases within the chamber. The transparent design also enabled monitoring of the gas mixing dynamics during the experiment.

This experimental chamber was used to simulate conditions typical of underground mining environments, where methane accumulation may occur in confined spaces with insufficient ventilation. Such an approach makes it possible to reproduce characteristic processes of gas concentration changes commonly observed in mines and underground structures.

Gas analyzers were installed at different locations within the experimental chamber, allowing gas concentration changes to be recorded at multiple measurement points. The use of multiple devices increased the reliability of measurements and enabled the formation of a more representative time-series dataset of gas concentrations.

Gas mixtures were introduced into the experimental chamber using certified calibration gas mixtures containing methane. The gas concentration was regulated to reproduce different scenarios of gas environment dynamics, including gradual increases in methane concentration as well as short-term methane release events.

This experimental setup provided controlled experimental conditions and ensured the reproducibility of the obtained results.

Data transmission from the gas analyzers was carried out via a wireless network with a 15-second interval, enabling the formation of time-series datasets of gas concentration measurements. Such a transmission interval is typical for industrial monitoring systems and provides sufficient measurement resolution while reducing the load on communication channels and data storage systems.

The experiment was conducted over a 10-day period, resulting in the formation of a multidimensional time series of gas environment parameters.

*Experimental Conditions.*During the experiment, the concentrations of several gases were monitored, including methane (CH_4), carbon monoxide (CO), hydrogen sulfide (H_2S), and oxygen (O_2). Additionally, environmental parameters such as temperature and relative humidity were recorded.

Data transmission to the server occurred every 15 seconds. Therefore, the number of measurements collected per day was:

$$N_{day} = \frac{24 \times 60 \times 60}{15} = 5760 \qquad (14)$$

The total number of measurements recorded during the entire experimental period was:

$$N_{total} = 5760 \times 10 = 57600 \qquad (15)$$

The main parameters of the experimental study are presented in Table 4.

Table 4. Parameters of the experimental study

Parameter	Value
Experiment duration	10 days
Measurement interval	15 seconds
Measurements per day	5,760
Total number of measurements	57,600
Measured gases	CH_4, CO, H_2S, O_2
Temperature range	10 - 40 °C
Relative humidity	20 - 90 %

Gas leak simulation method	Certified calibration gas mixtures
Gas analyzer used	SENSOR - Mine 4GN

Thus, the generated dataset represents a multidimensional time series of gas environment parameters, including gas concentrations and environmental conditions.

Experimental Gas Injection Scenarios. To increase the variability of methane concentration measurements, certified methane-air calibration gas mixtures were used throughout the experimental campaign. Gas mixtures with methane concentrations of 0.20%, 2.10%, and 4.20% by volume were introduced into the laboratory chamber according to predefined test scenarios.

Controlled gas injection events were performed during different stages of the 10-day monitoring period to reproduce realistic fluctuations in methane concentration. The experimental protocol included background monitoring, short-duration gas injections, gradual concentration increases, and subsequent ventilation periods. These scenarios enabled the collection of representative methane concentration dynamics under controlled laboratory conditions.

All experiments were conducted under safe operating conditions, and methane concentrations remained below explosive limits within the laboratory environment. The resulting dataset included both stable operating periods and transient concentration changes required for machine learning model development and evaluation.

Simulation of Methane Releases. To construct the experimental dataset, certified calibration gas mixtures containing methane were introduced into the experimental chamber at specific time intervals. This approach enabled the reproduction of various gas environment scenarios, including potentially hazardous gas concentration levels.

Methane is one of the most hazardous gases in the mining industry. The lower explosive limit (LEL) of methane is approximately 5% by volume in air, which corresponds to 100% LEL. In industrial practice, gas concentrations are commonly expressed as a percentage of the lower explosive limit (%LEL), allowing the degree of proximity to explosive conditions to be evaluated.

The relationship between volumetric gas concentration and concentration expressed as a percentage of the lower explosive limit can be represented as follows:

$$C_{vol} = \frac{C_{LEL}}{100} \times LEL \tag{16}$$

where C_{vol} - volumetric gas concentration, C_{LEL} - gas concentration expressed as a percentage of the lower explosive limit, LEL - lower explosive limit of the gas.

For methane, the lower explosive limit is 5% by volume. During the experiment, several methane concentration scenarios were simulated, including normal operating conditions, gradual increases in concentration, and emergency gas release events.

The simulated methane release scenarios are presented in Table 5.

Table 5. Scenarios of simulated methane releases

Day of experiment	**CH_4 concentration (%LEL)**	**CH_4 volume fraction (%)**	**Event type**
2	5 %LEL	0.25 %	Minor increase
4	10 %LEL	0.50 %	Warning level
6	15 %LEL	0.75 %	Hazardous concentration
8	25 %LEL	1.25 %	Emergency situation

The use of controlled gas release scenarios made it possible to construct a dataset containing both normal operating conditions and potentially hazardous changes in gas environment parameters.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

Data Preparation for Machine Learning. Before applying machine learning algorithms, preprocessing of the time-series measurement data was performed. This step was necessary to improve data quality and ensure the reliability of the prediction models.

The data preparation procedure included the following stages: removal of incorrect measurements and outliers, filtering of noise in sensor data, normalization of parameters, formation of training sequences from time-series data.

For model training, sequences of measurements with a fixed time window length were used. Each sequence served as input for predicting the future value of methane concentration.

A sliding-window approach was applied to transform the continuous time-series data into supervised learning samples. In this study, the input sequence length was set to 120 consecutive measurements. Since gas concentration measurements were collected every 15 seconds, each input sequence represented 30 minutes of monitoring history. The target variable was defined as the methane concentration 120 measurements ahead, which corresponds to a 30-minute forecasting horizon.

Thus, each training sample included historical values of methane concentration, carbon monoxide concentration, hydrogen sulfide concentration, oxygen concentration, temperature, and relative humidity over the previous 30 minutes, while the model output corresponded to the predicted methane concentration after the next 30 minutes. This configuration was selected to provide sufficient temporal context for identifying methane concentration trends and to support early warning in industrial safety applications.

Formally, the training dataset can be represented as:

$$(X_{t-n}, X_{t-n+1}, \dots, X_t) \rightarrow CH_4^{t+1} \quad (17)$$

where X_t - vector of gas environment parameters at time t.

The dataset was divided into three subsets: 70% for training, 15% for validation, and 15% for testing. The training dataset was used to fit the machine learning models, the validation dataset was used for hyperparameter tuning and model selection, and the testing dataset was reserved for the final evaluation of predictive performance.

To preserve the temporal structure of the measurements, the dataset was split chronologically rather than randomly. This approach prevents information leakage from future observations and provides a more realistic assessment of forecasting performance under industrial monitoring conditions.

The resulting dataset was subsequently used to train and validate the proposed hybrid methane concentration prediction model combining the LSTM recurrent neural network and the XGBoost algorithm.

Results

To evaluate the effectiveness of the proposed methane concentration prediction model, an analysis of the experimental dataset generated during the laboratory study was conducted. The primary objective of the analysis was to assess the ability of machine learning algorithms to predict changes in methane concentration based on time-series measurements of gas environment parameters.

The input data consisted of time-series measurements of gas concentrations and environmental parameters obtained from SENSOR - Mine 4GN gas analyzers. Prior to model training, the data underwent preprocessing, including noise filtering, parameter normalization, and the formation of training sequences.

To analyze the prediction performance, several machine learning algorithms were evaluated: Linear Regression, Random Forest, XGBoost, Long Short-Term Memory (LSTM) recurrent neural network.

The comparison of these models makes it possible to evaluate the advantages of neural networks and ensemble learning methods in analyzing time-series data of gas concentrations.

Methane Concentration Dynamics. During the experiment, time-series data of methane concentration were collected over a continuous 10-day monitoring period. Methane-air calibration gas mixtures were introduced into the experimental chamber according to predefined test scenarios, while methane concentration measurements were recorded every 15 seconds throughout the entire 10-day monitoring period. For visualization purposes, the methane concentration dynamics presented in Figure 3 were aggregated and displayed using 30-minute intervals in order to improve graph readability and highlight long-term concentration trends. The experimental protocol included both background monitoring periods and controlled gas injection events using certified methane-air calibration gas mixtures with methane concentrations of 0.20%, 2.10%, and 4.20% by volume.

As shown in Figure 3, methane concentration remained relatively stable during background monitoring periods and increased significantly during controlled gas injection events. The highest methane concentrations were observed during experiments performed using the 4.20% methane-air calibration gas mixture, while lower concentration peaks corresponded to the 0.20% and 2.10% methane-air mixtures.

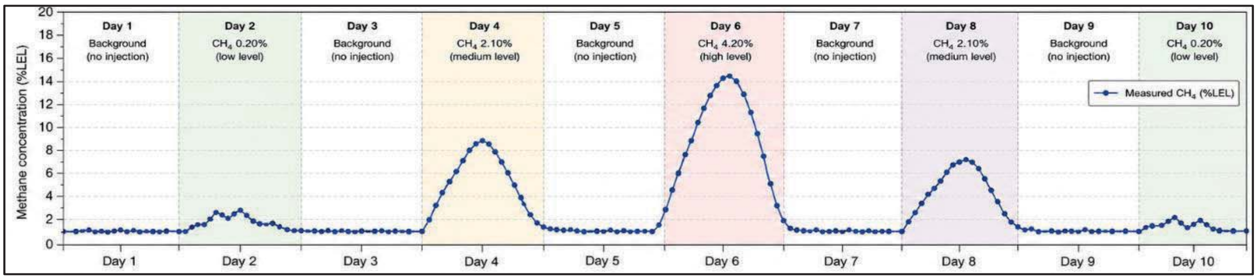

Figure 3. Methane concentration dynamics during the 10-day experimental period.

The obtained dataset contains both stable operating conditions and transient methane concentration changes, which are essential for training and evaluating machine learning models designed for gas concentration forecasting. Such concentration patterns are representative of industrial gas monitoring systems, where methane levels may change rapidly due to variations in ventilation conditions or localized gas release events.

The resulting dataset was subsequently used for machine learning model development, validation, and performance evaluation.

Methane Concentration Prediction. To evaluate the performance of the proposed forecasting approach, a methane concentration prediction task was carried out using the collected experimental dataset. The objective of the model was to predict future methane concentration values based on historical measurements and environmental parameters.

To assess the practical applicability of the proposed forecasting approach for industrial safety systems, several forecasting horizons were considered. In addition to one-step-ahead prediction, the model was evaluated for short-term forecasting horizons corresponding to 30 minutes. These forecasting horizons were selected because industrial safety personnel require sufficient time to assess hazardous situations, initiate warning procedures, and implement preventive measures before methane concentrations reach critical levels.

The prediction models were trained using methane concentration data together with auxiliary parameters, including carbon monoxide concentration, hydrogen sulfide concentration, oxygen concentration, temperature, and relative humidity. The inclusion of multiple environmental variables allows the model to capture complex nonlinear relationships affecting methane concentration dynamics.

Figure 4 presents an example comparison between the actual methane concentration values and the values predicted by the LSTM model.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

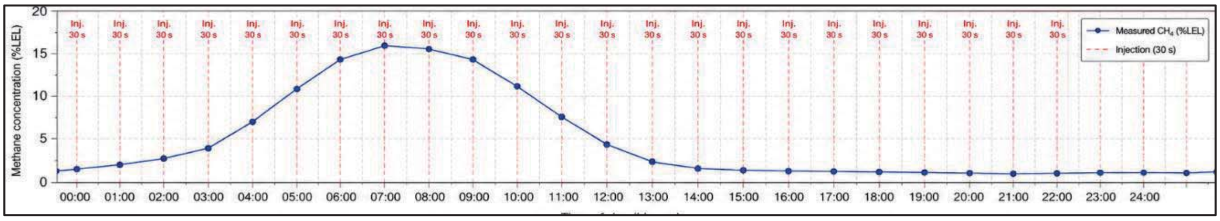

Figure 4. Comparison of actual and predicted methane concentration values.

As shown in Figure 4, the LSTM model accurately reproduces the dynamics of methane concentration changes and successfully captures both rapid concentration increases and subsequent decreases. The predicted values closely follow the actual methane concentration profile throughout the observation interval.

The results demonstrate that recurrent neural networks are capable of learning temporal dependencies in gas concentration measurements and can provide accurate short-term forecasts of methane concentration dynamics. This capability is particularly important for industrial safety applications, where early identification of increasing methane concentration may provide additional time for preventive actions and risk mitigation.

Furthermore, the prediction results confirm the suitability of machine learning methods for real-time gas monitoring systems and support the transition from traditional threshold-based monitoring toward predictive analytics and proactive industrial safety management.

Comparison of Machine Learning Models. For a quantitative evaluation of the performance of different machine learning models, the following metrics were used:

MAE (Mean Absolute Error):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{18}$$

RMSE (Root Mean Square Error):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{19}$$

R^2 (Coefficient of Determination):

$$R^2 = \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \tag{20}$$

where y_i - actual value of the parameter, $\bar{y}_i$ - mean value of the observations, $\hat{y}_i$ - predicted value. The comparison results of the models are presented in Table 6.

Table 6. Comparison of prediction model accuracy

Model	MAE	RMSE	R^2
Linear Regression	0.82	1.05	0.71
Random Forest	0.54	0.76	0.84
XGBoost	0.47	0.69	0.88

LSTM	0.31	0.48	0.93
Hybrid LSTM–XGBoost	0.19	0.30	0.97

As shown in Table 6, the Hybrid LSTM–XGBoost model achieved the highest prediction accuracy among all evaluated approaches, providing the lowest values of MAE and RMSE and the highest coefficient of determination ($R^2 = 0.97$).

Among the standalone models, the LSTM neural network demonstrated the best performance, outperforming Linear Regression, Random Forest, and XGBoost. This result confirms the ability of recurrent neural networks to effectively capture temporal dependencies in methane concentration dynamics.

The Hybrid LSTM–XGBoost model further improved forecasting performance through residual error correction. Compared with the standalone LSTM model, the hybrid approach reduced the Mean Absolute Error from 0.31 to 0.19 and the Root Mean Square Error from 0.48 to 0.30, while increasing the coefficient of determination from 0.93 to 0.97.

The performance improvement can be explained by the complementary strengths of the two algorithms. The LSTM network captures long-term temporal dependencies in methane concentration time series, whereas XGBoost learns nonlinear residual patterns associated with environmental conditions, gas injection scenarios, and sensor response characteristics. As a result, the hybrid model combines sequential learning and residual error correction within a single forecasting framework.

The obtained results confirm the suitability of machine learning methods for methane concentration forecasting and demonstrate the potential of predictive analytics for industrial gas monitoring systems. Such approaches enable a transition from traditional threshold-based monitoring toward proactive risk assessment and early warning of hazardous conditions.

Discussion

The results of the experimental study demonstrate that the proposed hybrid model is capable of identifying upward trends in methane concentration before critical levels are reached. This capability enables the detection of potentially hazardous situations at early stages and significantly improves the effectiveness of industrial safety systems.

Therefore, the application of predictive analytics methods combined with IoT-based gas analyzers provides an opportunity to move beyond traditional threshold-based gas concentration monitoring toward intelligent monitoring systems capable of forecasting the development of emergency situations and ensuring a higher level of safety at industrial facilities.

The obtained results are consistent with recent studies demonstrating the effectiveness of hybrid machine learning models for methane concentration forecasting and industrial gas monitoring [12,17]. Previous research has shown that recurrent neural networks are capable of capturing temporal dependencies in gas concentration time series [16-17], while recent studies on underground methane prediction also confirm the effectiveness of deep learning models for mining applications [20].

Compared with standalone machine learning approaches, the proposed Hybrid LSTM–XGBoost model achieved superior predictive performance, reducing forecasting errors and improving the coefficient of determination. This improvement can be attributed to the combination of temporal sequence learning and residual error correction, allowing the model to better represent complex gas concentration dynamics.

From a practical perspective, the proposed forecasting approach may support early warning systems in mining and industrial environments. The ability to predict methane concentration growth 30 minutes in advance provides additional time for ventilation control, evacuation planning, and preventive safety actions.

Nevertheless, the present study has several limitations. The experimental dataset was collected under laboratory conditions using controlled methane release scenarios. Although the obtained results demonstrated high forecasting accuracy and confirmed the effectiveness of the

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 26, JUNE 2026

proposed Hybrid LSTM–XGBoost model, the experiments do not fully reproduce the complexity of real industrial environments.

Factors such as mine ventilation conditions, equipment operation, geological characteristics, and irregular methane emissions may influence methane concentration dynamics in actual underground facilities. Therefore, additional validation of the proposed approach using real-world industrial datasets collected from operating mining enterprises is required before large-scale deployment. Future research will focus on long-term field testing and model adaptation to real industrial monitoring conditions.

Future research will focus on collecting long-term industrial datasets, expanding the model to multicomponent gas environments, and evaluating advanced deep learning architectures for real-time incident prediction.

Compared with the standalone LSTM model, the proposed Hybrid LSTM–XGBoost approach reduced MAE by 38.7%, RMSE by 37.5%, and increased R^2 from 0.93 to 0.97.

Another limitation of the present study is that statistical significance testing of the forecasting performance differences between the evaluated machine learning models was beyond the scope of the present study. Although the Hybrid LSTM–XGBoost model demonstrated lower MAE and RMSE values and a higher coefficient of determination compared with the standalone LSTM and XGBoost models, additional statistical validation using repeated experiments and cross-validation procedures is required to quantify the significance of these improvements. This aspect will be addressed in future research.

Conclusion

This study addressed the problem of predicting methane concentration in industrial atmospheres based on data obtained from IoT gas analyzers. An approach to intelligent gas environment monitoring was proposed, integrating sensor devices, a data transmission system, and machine learning algorithms for the analysis of gas concentration time series.

Within the framework of the study, a monitoring system architecture was developed based on SENSOR - Mine 4GN gas analyzers, which provide continuous measurement of methane, carbon monoxide, hydrogen sulfide, and oxygen concentrations. Data transmission from the devices to the monitoring system server is performed at 15-second intervals, enabling the formation of detailed time-series datasets of gas environment parameters.

To evaluate the effectiveness of the proposed approach, experimental studies were conducted under laboratory conditions. During the experiment, a 10-day time-series dataset of gas environment parameters was generated. Certified calibration gas mixtures were used to simulate emergency scenarios and reproduce controlled methane release events.

For data analysis and methane concentration prediction, a hybrid machine learning model combining a Long Short-Term Memory (LSTM) recurrent neural network and the XGBoost algorithm was proposed. The use of this model allows both temporal dependencies in the data and nonlinear relationships between gas environment parameters to be taken into account.

The modeling results showed that the application of machine learning algorithms significantly improves methane concentration prediction accuracy compared with traditional data analysis methods. Among the evaluated approaches, the proposed Hybrid LSTM–XGBoost model demonstrated the best performance, achieving the lowest prediction errors and the highest coefficient of determination ($R^2 = 0.97$). The standalone LSTM model showed the strongest performance among the individual machine learning algorithms, confirming the effectiveness of recurrent neural networks for methane concentration forecasting.

The use of predictive analytics methods in gas monitoring systems enables the detection of hazardous trends in gas concentration before critical levels are reached. This creates the foundation for the development of intelligent industrial safety systems capable of preventing emergency situations at early stages.

The scientific novelty of this research lies in the development of a methane concentration prediction model based on time-series data from IoT gas analyzers and the hybrid application of

LSTM and XGBoost algorithms. The proposed approach improves the accuracy of gas environment dynamics analysis and can be applied in the development of intelligent monitoring systems for industrial facilities.

The practical significance of the study lies in the possibility of applying the proposed model in industrial safety systems for mining and energy sector enterprises. The use of predictive analytics in combination with IoT gas analyzers can improve the efficiency of gas environment monitoring and reduce the risk of industrial accidents.

Future research will focus on expanding the experimental dataset and developing prediction methods for multicomponent gas environments using more advanced deep learning models.

References

[1] International Labour Organization. (2019). Safety and health at the heart of the future of work: Building on 100 years of experience. Geneva: International Labour Office. Available at: https://www.ilo.org/publications/safety-and-health-future-work-building-100-years-experience (accessed: 24 July 2026).

[2] Haleem, A., Javaid, M., & Singh, R. P. (2025). Encouraging Safety 4.0 to enhance industrial culture: An extensive study of its technologies, roles, and challenges. *Green Technologies and Sustainability*, 3, 100158. https://doi.org/10.1016/j.grets.2024.100158

[3] Zhang, J., Xu, K., You, G., Wang, B., & Zhao, L. (2022). Analysis of the interaction mechanism of the risk factors of gas explosion accidents in Chinese underground coal mines. *International Journal of Environmental Research and Public Health*, 19(2), 1002. https://doi.org/10.3390/ijerph19021002

[4] Huang, Y., Li, S., Fan, J., Yan, Z., & Li, C. (2023). A Spark Streaming-based early warning model for gas concentration prediction. *Processes*, 11(1), 220. https://doi.org/10.3390/pr11010220

[5] Gololo, M. G. D., Nyathi, C. W., Boateng, L., Nkadimeng, E. K., Mckenzie, R. P., Atif, I., Kong, J., Mahboob, M. A., Cheng, L., & Mellado, B. (2024). Review of IoT systems for air quality measurements based on LTE/4G and LoRa communications. IoT, 5(4), 711-729. https://doi.org/10.3390/iot5040032

[6] Centenaro, M., Vangelista, L., Zanella, A., & Zorzi, M. (2016). Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios. *IEEE Wireless Communications*, 23(5), 60-67. https://doi.org/10.1109/MWC.2016.7721743

[7] Chang, G., & Chang, H. (2023). Underground abnormal sensor condition detection based on gas monitoring data and deep learning image feature engineering. *Heliyon*, 9(11), e22026. https://doi.org/10.1016/j.heliyon.2023.e22026

[8] Cong, Y., Zhao, X., Tang, K., Wang, G., Hu, Y., & Jiao, Y. (2022). FA-LSTM: A novel toxic gas concentration prediction model in pollutant environment. *IEEE Access*, 10, 1591-1602. https://doi.org/10.1109/ACCESS.2021.3133497

[9] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). *Association for Computing Machinery*. https://doi.org/10.1145/2939672.2939785

[10] Munera, D., Aguirre, J., & Gomez, N. G. (2021). IoT-based air quality monitoring systems for smart cities: A systematic mapping study. *International Journal of Electrical and Computer Engineering*, 11(4), 3470-3482. https://doi.org/10.11591/ijece.v11i4.pp3470-3482

[11] Augustin, A., Yi, J., Clausen, T., & Townsley, W. M. (2016). A study of LoRa: Long range and low power networks for the Internet of Things. *Sensors*, 16(9), 1466. https://doi.org/10.3390/s16091466

[12] Majumder, R., Pollard, J., Salek, M. S., Werth, D., Comert, G., Gale, A., Khan, S. M., Darko, S., & Chowdhury, M. (2024). Development and evaluation of ensemble learning-based

environmental methane detection and intensity prediction models. *Environmental Health Insights*, 18. https://doi.org/10.1177/11786302241227307

[13] Jin, Z., He, J., Wang, W., et al. (2024). Monitoring methane concentrations with high spatial resolution over China by using Random Forest model. *Remote Sensing*, 16(14), 2525. https://doi.org/10.3390/rs16142525

[14] Zhang, W., Liu, H., & Chen, Y. (2025). Retrieval of atmospheric XCH4 via XGBoost method based on TROPOMI satellite data. *Atmosphere*, 16(3), 279. https://doi.org/10.3390/atmos16030279

[15] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. Cambridge, MA: The MIT Press. ISBN 978-0-262-03561-3.

[16] Liu, C., Li, S., Zhao, X., et al. (2023). LSTM-Pearson gas concentration prediction model based on data feature selection. *Energies*, 16(5), 2318. https://doi.org/10.3390/en16052318

[17] Meng, X., Chang, H., & Wang, X. (2022). Methane concentration prediction method based on deep learning and classical time series analysis. *Energies*, 15(6), 2262. https://doi.org/10.3390/en15062262

[18] Xu, N., Liu, Z., Ren, W., & Chen, X. (2022). Gas concentration prediction based on IWOA-LSTM-CEEMDAN residual correction model. *Sensors*, 22(12), 4412. https://doi.org/10.3390/s22124412

[19] Karagulian, F., Barbiere, M., Kotsev, A., Spinelle, L., Gerboles, M., Lagler, F., Redon, N., Crunaire, S., & Borowiak, A. (2019). Review of the performance of low-cost sensors for air quality monitoring. *Atmosphere*, 10(9), 506. https://doi.org/10.3390/atmos10090506

[20] Tutak, M., & Brodny, J. (2024). Predicting methane concentrations in underground coal mining using a multi-layer perceptron neural network based on mine gas monitoring data. *Sustainability*, 16(19), 8388. https://doi.org/10.3390/su16198388