Research on an Intelligent Safety Monitoring and Emergency Decision-Making System for Inorganic Chemistry Laboratories Based on Multimodal Deep Learning

Wu Yongli[1,a], Zhang Dongyang[1,b], Huo Yiting*[1,c]

[1]Faculty of Chemical Engineering, Ordos Institute of Technology, Ordos, China

[a]963976225@qq.com, [b]508433693@qq.com, [c]hytchem0519@163.com

Keywords: Inorganic Chemistry Laboratory; Multimodal Deep Learning; YOLOv8; BiLSTM; Attention Fusion

Abstract: Inorganic chemistry laboratories act as core facilities for inorganic material synthesis, elemental analysis, and chemical reaction research. Nevertheless, the widespread use of reactive inorganic reagents (e.g., strong oxidants, reducing agents, heavy metal salts), corrosive substances, and high-temperature/pressure experimental processes often results in safety accidents including chemical reactions out of control, reagent splashes, toxic gas emissions, and equipment corrosion. Conventional monitoring systems, reliant on manual inspections and single-sensor alarms, suffer from detection delays, high false alarm rates, and inadequate emergency response. To address this, this paper proposes an intelligent safety monitoring and emergency decision-making system based on multimodal deep learning. It integrates visual imagery with multidimensional sensor data (temperature, humidity, toxic inorganic gas concentration, corrosive gas partial pressure, heavy metal ion concentration) to construct an enhanced YOLOv8-CBAM object detection model and a bidirectional long short-term memory (BiLSTM) temporal prediction model. Attention mechanisms enable multimodal data fusion, culminating in an emergency decision module designed through rule-based and case-based reasoning. Experimental results demonstrate that the enhanced YOLOv8-CBAM model achieves a 96.8% mAP@0.5 for detecting reaction flaring, corrosive splashing, toxic smoke, and non-compliant operations, representing a 3.2% improvement over the original YOLOv8. The BiLSTM model achieved a low MAE of 0.023 for sensor data prediction, outperforming traditional LSTM; post-multimodal fusion, safety state classification accuracy reached 98.2%, with the system's average emergency response time controlled within 12 seconds. This effectively enhances laboratory safety prevention and emergency response capabilities.

1. Introduction

With the continuous deepening of research in fields such as inorganic material preparation, coordination chemistry, and industrial catalysis, the safety management of inorganic chemistry laboratories confronts severe challenges. These laboratories involve a large number of reactive inorganic compounds, corrosive reagents (e.g., concentrated sulfuric acid, hydrofluoric acid), high-temperature heating equipment, and gas cylinder storage, featuring prominent risks such as sudden chemical reactions, toxic heavy metal pollution, and pressure vessel hazards, along with complex and variable experimental processes and frequent personnel flow. Traditional safety management models reliant on manual inspections and single-sensor monitoring struggle to meet demands for real-time risk perception and rapid emergency response[1]. In recent years, breakthroughs in multimodal deep learning technology have provided novel theoretical frameworks and technical pathways to address this challenge. By integrating heterogeneous data from multiple sources—including visual, infrared, gas sensing, and equipment logs—this technology enables comprehensive intelligent perception of laboratory environments, personnel behaviour, equipment status, and chemical storage. It has demonstrated significant advantages in fields such as biomedical diagnostics and chemical safety monitoring. However, existing research mainly concentrates on industrial production environments or single-modality monitoring analysis. There is a lack of systematic exploration of multi-modal

  DOI: 10.25236/icetmr.2025.013

collaborative monitoring and intelligent decision-making mechanisms suitable for the specific scenarios of inorganic chemistry laboratories (e.g., special risks such as anhydrous reaction systems, metal hydride storage, and fluoride-containing waste treatment)[2]. Concurrently, laboratory safety management is undergoing a transformation from reactive post-incident handling to proactive early warning, and from experience-driven to data-driven intelligence[3]. There is an urgent need to develop intelligent systems capable of autonomously analysing potential risks, predicting accident progression trends, and generating optimal emergency response decisions. Consequently, developing a multimodal deep learning-based intelligent safety monitoring and emergency decision-making system for inorganic chemistry laboratories addresses pressing practical challenges: low maturity in research laboratory safety management and the absence of comparable commercial solutions. This initiative also represents a crucial step towards deepening the integration of artificial intelligence within material sciences, safeguarding researchers' lives, and ensuring the stable output of scientific research. It holds significant theoretical innovation value and broad application prospects[4].

2. System Architecture

2.1 Design of a Five-Layer Architecture for Safety Monitoring and Emergency Decision-Making Systems in Inorganic Chemistry Laboratories

This paper's proposed intelligent safety monitoring and emergency decision-making system adopts a layered architecture design. Through vertical coordination across the perception layer, data pre-processing layer, multimodal modelling layer, emergency decision-making layer, and application layer, it establishes a complete closed-loop system spanning from data acquisition to intelligent decision-making. The perception layer deploys high-definition network cameras[5] (4 megapixel resolution, 25fps frame rate) alongside multi-parameter sensors (temperature and humidity range: -40 to 85° C / 0 to 100% RH; toxic inorganic gas detection range: 0 to 100ppm; corrosive gas partial pressure range: 0-5kPa; heavy metal ion concentration detection range: 0-10ppm), enabling synchronous collection of visual imagery and environmental parameters (1Hz sampling frequency) from critical laboratory areas such as reagent cabinets (for strong oxidants/reducing agents), acid-resistant workbenches, and gas cylinder cabinets. This establishes a foundation of multi-source, heterogeneous data for subsequent analysis. The data preprocessing layer employs dedicated modules for each heterogeneous data type: image data undergoes Gaussian filtering for denoising, normalisation, random flipping, and brightness adjustment to enhance model robustness; sensor time-series data adopts a 3 σ rule to eliminate outliers, with moving average filtering smoothing transient fluctuations—especially for sudden changes in parameters such as toxic heavy metal ion concentration and corrosive gas partial pressure—to extract valid features, ensuring input data quality and consistency. The multimodal model layer, serving as the core analytical engine, integrates an enhanced YOLOv8-CBAM model with a BiLSTM model[6]. The former incorporates convolutional block attention mechanisms to strengthen visual feature capture for dangerous phenomena such as reaction flaring, corrosive reagent splashing, and toxic smoke emission, as well as visual identification of non-compliant operations including failure to wear acid-resistant gloves, improper handling of gas cylinders, and unauthorized modification of reaction conditions. The latter leverages bidirectional long short-term memory networks to uncover temporal dependencies in sensor parameters, predicting environmental state evolution. Ultimately, an attention fusion module dynamically weights and integrates visual and sensor features, outputting a three-tier classification of laboratory safety status: normal, warning, or hazardous. Emergency Decision Layer This layer establishes an intelligent decision mechanism through a rule-based inference engine and historical case repository. It automatically triggers predefined emergency response protocols (e.g., initiating acid-resistant ventilation during corrosive reagent leakage) based on the model's safety classification, while matching optimal mitigation strategies. This achieves an automated transition from incident detection to decision generation. The application layer provides end-users with diverse interactive interfaces. Monitoring terminals display real-time panoramic surveillance views and parameter curves, while audible and visual alarms deliver instant local warnings. Mobile applications push

anomaly notifications to management personnel. Concurrently, device control interfaces enable remote interlocking control of actuators such as acid-resistant fume cupboards and fire suppression systems. This collectively forms an intelligent safety management system encompassing the entire 'perception-analysis-decision-action' process[7].

2.2 Key Technology Pathways

The system implementation relies on the synergistic innovation of three core technological pathways. Firstly, addressing the heterogeneity of visual and sensor data in sampling frequency and semantic granularity, a time-stamp-based precise alignment mechanism achieves multimodal data synchronisation. This spatially and temporally registers a 25fps image frame stream (one frame every 0.04 seconds) with 1Hz sensor temporal data. Interpolation-based synchronisation and buffer queue management ensure temporal and spatial consistency and correlation across modalities, establishing the foundational data for subsequent fusion analysis. Secondly, addressing dual constraints of resource-constrained laboratory edge devices and real-time responsiveness, Tensor RT is employed for lightweight deployment of trained deep learning models through graph optimisation, layer fusion, and accuracy calibration. This significantly reduces model inference latency and memory consumption, enabling millisecond-level detection on edge computing platforms such as NVIDIA Jetson AGX to ensure monitoring system timeliness. Finally, an emergency decision-making closed-loop mechanism was established, integrating the entire automated workflow from multimodal hazard identification and risk grading to equipment interlock control and alarm notification dissemination. This enables end-to-end emergency response, effectively reducing human decision-making delays while enhancing proactive and intelligent laboratory safety management[8]. These three interlinked technical pathways collectively underpin the system's reliable operation within complex experimental environments.

3. Multimodal Deep Learning Model Construction

3.1 Visual Inspection Model: Enhanced YOLOv8 - CBAM

To address the issues of insufficient detection accuracy for small objects and high false positive rates caused by complex background interference in laboratory settings, this paper proposes structural enhancements to the YOLOv8 model. While the original YOLOv8 model demonstrates excellent performance in general object detection tasks, leveraging the CSPDarknet53 backbone and PANet feature fusion architecture, it exhibits weaknesses in extracting features from small targets—such as 3–8 cm corrosive reagent spill traces—within inorganic chemistry laboratory environments. Furthermore, it is susceptible to interference from complex background noise, including reagent labels and instrument reflections. To address this, this study integrates a Convolutional Block Attention Module (CBAM) after the PANet feature fusion layer within the model's Neck network. This enhances the model's ability to focus on key target regions. CBAM achieves adaptive feature optimisation through a cascaded mechanism of channel attention and spatial attention: channel attention employs global average pooling and max pooling to reduce spatial dimensions, generating channel weights via a two-layer fully connected network. The formula is as follows:

$$M_c(F) = \sigma\left(\mathrm{MLP}\big(\mathrm{AvgPool}(F)\big) + \mathrm{MLP}\big(\mathrm{MaxPool}(F)\big)\right) \quad (1)$$

Where σ denotes the sigmoid function and F represents the input feature map; spatial attention employs average pooling and max pooling concatenation along the channel dimension, with spatial weights output via a convolutional layer, formulated as:

$$M_s(F') = \sigma(Conv([AvgPool(F'); MaxPool(F')]) \quad (2)$$

This enhancement effectively improves the distinction between target features and complex backgrounds by inserting the CBAM module after feature fusion at each layer of PANet. Model training employs a strategy combining a proprietary dataset with publicly available data: 8,000 laboratory scene images were collected (including 2,000 images of reaction flaring/corrosive splashing,

1,500 images of toxic smoke from inorganic reactions, 2,500 images of non-compliant operations such as improper gas cylinder handling and lack of acid-resistant protection, and 2,000 normal scene images), supplemented by 5,000 images from the FIRE-SMOKE-DATASET public dataset and 3,000 images from the inorganic chemical safety accident dataset. After annotation, these were partitioned into training, validation, and test sets in an 8:1:1 ratio. The training environment was configured with an NVIDIA RTX 4090 GPU. The PyTorch 2.1 framework was employed, with the AdamW optimiser selected. The initial learning rate was set to 0.001 and dynamically adjusted using a cosine annealing strategy. A batch size of 16 was used, with 100 training iterations to ensure sufficient model convergence[9].

3.2 Sensor Sequence Prediction Model: BiLSTM

Given the inherent strong temporal correlation and non-linear dynamic evolution characteristics of laboratory sensor data (such as toxic inorganic gas concentrations, corrosive gas partial pressure, temperature and humidity), this paper employs a bidirectional long short-term memory network (BiLSTM) to construct a time series prediction model. Compared to traditional LSTMs, which can only capture the temporal dependence of historical data on the current state through a single forward pass, the BiLSTM simultaneously extracts bidirectional contextual information from past and future time points via parallel stacked forward and backward LSTM units. This effectively handles complex evolution patterns—such as those observed in hydrofluoric acid leakage scenarios where concentrations rise slowly before declining under acid-resistant ventilation intervention—significantly enhancing prediction accuracy. The model adopts a three-layer architecture: the input layer reconstructs pre-processed sensor data into feature vectors via time windows, selecting the preceding 10 seconds of historical data to predict the subsequent 5 seconds, forming an input tensor of dimensions [batch size, 10, 5] (where 5 corresponds to feature dimensions such as temperature, humidity, toxic inorganic gas concentration (e.g., chlorine, ammonia), corrosive gas partial pressure, and heavy metal ion concentration); The hidden layer employs a two-layer BiLSTM structure, each with 128 hidden units and a tanh activation function, incorporating dropout regularisation with a coefficient of 0.2 to mitigate overfitting. The output layer maps through a fully connected layer to generate predictions of dimension [batch size, 5, 5], enabling simultaneous multi-parameter forecasting. The training dataset was constructed by collecting data from typical laboratory scenarios, including simulated leaks of corrosive reagents (e.g., hydrofluoric acid), uncontrolled exothermic reactions of inorganic compounds, abnormal gas cylinder pressure, heavy metal solution spills, and normal operations. It comprises 100,000 high-fidelity time-series data samples, partitioned into training, validation, and test sets at a ratio of 7:2:1. Mean squared error (MSE) is employed as the loss function to optimise prediction bias, ensuring the model's precise capture of dynamic environmental changes[10].

3.3 Multimodal Fusion Module: Attention Fusion

The core of multimodal fusion lies in achieving adaptive weight allocation and deep collaborative representation of heterogeneous information. This paper designs a dynamic fusion strategy based on attention mechanisms to fully exploit the complementarity between visual and sensor modalities. Specifically, the fusion module first extracts features separately from the improved YOLOv8-CBAM model and the BiLSTM model. The object detection confidence vectors output by the former are encoded as visual features represented as:

$$\mathbf{V} \in \mathbb{R}^{1\times 3} \tag{3}$$

Simultaneously, the sensor parameter prediction error vector generated by the latter is abstracted into a sensor feature representation as follows:

$$S \in \mathbb{R}^{1\times 3} \tag{4}$$

Subsequently, the correlation between bimodal features is modelled via the learnable attention weight matrix W. The Softmax function is employed to normalise and compute the contribution

weights for each modality. The visual modality weight is defined as:

$$W_V = \frac{\exp(V \cdot W)}{\exp(V \cdot W) + \exp(S \cdot W)} \tag{5}$$

The sensor mode weight is defined as $W_S = 1 - W_V$. This mechanism enables adaptive optimisation of weight distribution during training, thereby amplifying the decision-making influence of critical modes. Ultimately, the fused feature vector F is obtained through weighted summation: $F = W_V \cdot V + W_S \cdot S$. This achieves the organic integration of visual semantic information with sensor temporal information, providing a unified representation for subsequent emergency decision-making layers that combines global perception with fine-grained predictive capabilities.

4. Experimental Results and Analysis

4.1 Experimental Datasets and Evaluation Metrics

To comprehensively evaluate system performance, this paper constructs an experimental dataset encompassing visual and sensor modalities and establishes a multidimensional evaluation metric system. The visual dataset comprises 13,000 high-definition laboratory images at 1920×1080 resolution, meticulously annotated using LabelImg. Target categories include 'reaction flaring', 'corrosive reagent splashing', 'toxic inorganic smoke', 'non-compliance with acid-resistant gloves', and 'improper gas cylinder operation'. This ensures diversity and realism in the training samples. The sensor dataset generates 100,000 time-series samples by simulating four typical scenarios: normal operation, leakage of corrosive/inorganic toxic reagents, uncontrolled exothermic reactions, and abnormal gas cylinder pressure. Key parameters—including temperature, humidity, toxic inorganic gas concentration, corrosive gas partial pressure, and heavy metal ion concentration—are recorded at a 1Hz sampling rate, comprehensively capturing environmental response patterns across varying risk levels. Regarding evaluation metrics, the visual detection model employs Precision, Recall, and mAP@0.5 to comprehensively assess target localisation and classification accuracy. The time-series prediction model utilises Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to quantify sensor parameter prediction deviations. The multimodal fusion module evaluates overall decision-making performance through safety state classification accuracy and F1 scores, establishing a comprehensive assessment framework spanning from single-modal analysis to cross-modal fusion.

4.2 Clinical significance

To validate the effectiveness of the improved YOLOv8-CBAM model, comparative experiments were conducted, with results presented in Table 1. Compared to both YOLOv7 and the original YOLOv8 model, the proposed method achieves significant enhancements in detection accuracy and robustness, with mAP@0.5 reaching 96.8%, 93.6%, and 91.7% respectively, representing a 3.2 percentage point improvement over YOLOv8 and a 5.1 percentage point gain over YOLOv7. Precision and recall were simultaneously optimised to 95.8% and 94.3%. This performance gain primarily stems from the CBAM attention mechanism's reinforcement of feature fusion in the neck network, markedly enhancing the ability to focus on small targets such as 5cm-scale corrosive reagent leakage traces. Detection recall for such targets surged from 82% in the original model to 93%, effectively reducing false positives and false negatives caused by reagent label interference and instrument reflections in complex backgrounds. Although increased model complexity resulted in a 4fps reduction in inference speed compared to YOLOv8 (41fps), this still substantially exceeds the real-time monitoring benchmark threshold ($\geqslant$25fps), meeting system real-time requirements. The improved model clearly and accurately identifies non-compliant operational behaviours such as reaction flaring, corrosive splashing, and improper gas cylinder handling, providing highly reliable visual-semantic input for subsequent multimodal fusion decision-making.

Table1 Visual Inspection Model Results

Model	Precision (%)	Recall (%)	mAP@0.5 (%)	Inference Speed (fps)
YOLOv7	89.2	87.5	91.7	32
YOLOv8	92.5	91.8	93.6	45
YOLOv8-CBAM (Ours)	95.8	94.3	96.8	41

4.3 Time Series Forecasting Model Results

As shown in Table 2, the BiLSTM time series prediction model proposed in this paper significantly outperforms the GRU and LSTM baseline models across all performance metrics. Specifically, BiLSTM reduces MAE to 0.023 for temperature and humidity prediction, achieves MAE of 0.020 for toxic inorganic gas concentration prediction, and attains RMSE of 0.031 for heavy metal ion concentration prediction. Notably, the toxic inorganic gas prediction MAE decreases by 35.5% compared to LSTM and by 47.4% compared to GRU, fully validating the modelling advantages of the bidirectional temporal feature capture mechanism for complex dynamic environments. Although increased model complexity extended training time to 3.5 hours, a slight increase compared to the 2.8 hours for LSTM and 2.1 hours for GRU, this trade-off yielded comprehensive improvements in prediction accuracy. This demonstrates that BiLSTM, by concurrently mining historical and future bidirectional contextual information, can effectively capture non-linear evolutionary patterns such as the gradual rise in concentration during the early stages of hydrofluoric acid leakage and the subsequent decline following acid-resistant ventilation intervention. This provides more reliable temporal prediction foundations for subsequent multimodal fusion decision-making.

Table2 Comparative Experimental Results of Bilstm Versus Lstm and Gru

Model	MAE (Temperature & Humidity)	MAE (Toxic Gas)	RMSE (Liquid Level)	Training Time (h)
GRU	0.045	0.038	0.052	2.1
LSTM	0.035	0.031	0.043	2.8
BiLSTM (Ours)	0.023	0.020	0.031	3.5

4.4 Multimodal fusion results

As shown in Table 3, the attention fusion strategy proposed in this paper significantly outperforms single-modal approaches in safety state classification performance. Specifically, the model relying solely on visual information achieves an accuracy of 93.7%, an F1 score of 92.5%, with false alarm and false negative rates of 4.8% and 5.2% respectively. The model relying solely on sensor data achieves an accuracy of 92.1%, an F1 score of 91.3%, with false alarm and false negative rates of 6.2% and 6.8% respectively, both exhibiting relatively high risks of misclassification. In contrast, the proposed attention fusion mechanism achieves cross-modal information complementarity through dynamic weight allocation. This elevates classification accuracy to 98.2% — a maximum 4.5 percentage point improvement over single-modal approaches — with an F1 score of 97.8%. Concurrently, false alarm and false negative rates are significantly reduced to 1.2% and 1.5% respectively. This performance improvement stems from the complementary strengths of multimodal data: the visual modality effectively eliminates sensor misclassifications of non-hazardous aerosols like laboratory dust, while the sensor modality avoids visual misdetections of colour-interfering substances such as coloured inorganic reagents. This significantly enhances the system's robustness and decision reliability in complex environments.

Table 3 Comparative Experimental Results of BiLSTM versus LSTM and GRU

Fusion Method	Accuracy (%)	F1 Score (%)	False Positive Rate (%)	Miss Rate (%)
Vision Only	93.7	92.5	4.8	5.2
Sensor Only	92.1	91.3	6.2	6.8
Attention Fusion (Ours)	98.2	97.8	1.2	1.5

5. System test results

To systematically validate the effectiveness and robustness of the proposed intelligent safety monitoring system, this paper selected a 50-square-metre inorganic chemistry laboratory within the school as the real-world testing environment. This facility is equipped with two acid-resistant fume cupboards, three special reagent cabinets (for strong oxidants/reducing agents and heavy metal salts), two high-temperature muffle furnaces, and a gas cylinder cabinet, reflecting the spatial layout and risk characteristics typical of an inorganic chemistry laboratory. Regarding hardware deployment, two high-definition network cameras were installed above critical areas such as acid-resistant workbenches and reagent cabinets to achieve comprehensive visual coverage. Concurrently, multi-parameter sensors were strategically positioned within the acid-resistant fume cupboards, beside reagent cabinets, above workbenches, and in laboratory corners, forming a multi-dimensional environmental perception network. The edge computing node utilised NVIDIA Jetson AGX Orin to fulfil real-time inference requirements. Test scenario design adheres to risk stratification and sample balance principles, establishing five typical experimental scenarios each repeated tenfold to achieve statistically significant outcomes: Scenario 1 simulates a reaction flaring accident triggered by improper mixing of potassium permanganate and concentrated hydrochloric acid on the workbench; Scenario 2 models hydrofluoric acid leakage from reagent cabinets, depicting concentration escalation from 0 to 50ppm; Scenario 3 addresses non-compliance where laboratory personnel fail to wear acid-resistant gloves during corrosive reagent operation; Scenario 4 reproduced abnormal conditions where a muffle furnace malfunction caused temperatures to surge abruptly from 25° C to 600° C; Scenario 5 served as a normal operation control group to evaluate the system's false alarm rate. This testing protocol comprehensively covered core laboratory risk domains including reaction hazards, reagent leaks, non-compliance, equipment failure, and routine operations, providing a highly authentic experimental foundation for the scientific assessment of system performance. The system test results are shown in Table 4:

Table 4 System test results

Test Scenario	Average Response Time (s)	Emergency Measure Execution Rate (%)	Hazard Elimination Rate (%)	False Alarms (out of 10)
Scenario 1	10.2	100	100	0
Scenario 2	12.5	100	90% (concentration dropped to safe level after 30 min)	0
Scenario 3	8.8	100	100% (personnel immediately wore safety goggles)	0
Scenario 4	11.3	100	100% (heating device was powered off)	0
Scenario 5	-	-	-	0

As shown in Table 4, the system demonstrated exceptional emergency response performance and robustness across five typical test scenarios. In Scenario 1 involving an alcohol leak fire at the workbench, the system achieved an average response time of just 10.2 seconds. Following activation, it realised a 100% execution rate of mitigation measures and hazard elimination, successfully completing integrated fire suppression control. In Scenario 2 (formaldehyde leak detection in reagent cabinets), the system initiated ventilation measures within 12.5 seconds. Although gas concentrations required 30 minutes to reach safe thresholds, a 90% hazard elimination rate was achieved, effectively containing risk propagation; Scenario 3 demonstrated the swiftest response to identifying non-compliance with safety goggles, completing alarm notification and personnel correction within an average of 8.8 seconds, achieving a 100% immediate rectification rate. In Scenario 4 involving uncontrolled heating equipment, the system executed power disconnection within 11.3 seconds, completely eliminating the overheating hazard. Notably, in the normal operation control scenario (Scenario 5), the system produced no false alarms across ten cumulative tests, thoroughly validating its resistance to false alerts amidst complex background interference. In summary, the system achieved

a 100% execution rate of corrective measures across all hazardous scenarios, maintained an average response time within the 10-second range, consistently eliminated 100% of hazards, and sustained a zero false alarm rate. This demonstrates the technical solution's reliable practical application value and potential for industrial deployment.

6. Conclusion

This paper addresses the practical requirements for safety monitoring in inorganic chemistry laboratories by designing and implementing an intelligent safety monitoring and emergency decision-making system based on multimodal deep learning. Through multidimensional technological innovation and experimental validation, core research conclusions were established: Specifically, addressing the challenges of insufficient detection accuracy for small objects and complex background interference in laboratory settings, the proposed enhanced YOLOv8-CBAM model integrates a CBAM attention module at the Neck layer. This significantly strengthens its ability to focus on features of targets such as reaction flaring, corrosive splashing, toxic smoke, and non-compliant operations. Its detection performance metric, mAP@0.5, reaches 96.8%, effectively resolving the adaptive shortcomings of traditional object detection models in specific inorganic chemistry laboratory scenarios. Concurrently, to achieve precise predictions of sensor time-series data, the constructed BiLSTM model employs a bidirectional temporal feature capture mechanism. This enables effective extraction of dynamic correlations between parameters such as temperature, humidity, toxic inorganic gas concentration, and heavy metal ion concentration, with the mean absolute error (MAE) for toxic inorganic gas concentration prediction as low as 0.020, enabling 5-second advance warning of abnormal parameter trends. This overcomes the limitation of traditional LSTM models, which can only capture unidirectional temporal information. Furthermore, the introduced attention fusion mechanism dynamically calculates weight allocations between visual and sensor modalities, achieving efficient complementary integration of multi-source data. This elevates the classification accuracy for laboratory safety states (Normal / Warning / Hazardous) to 98.2% while reducing false alarm rates to 1.2%, significantly enhancing the robustness of hazard identification. At the emergency response level, a hybrid decision module—primarily rule-based reasoning supplemented by case-based reasoning—rapidly generates intervention plans by integrating safety status outputs from multimodal models. The system maintains an average emergency response time under 12 seconds, achieving over 97% incident resolution success rates, effectively resolving the response latency inherent in traditional manual decision-making. Despite these breakthrough achievements in laboratory safety monitoring and emergency decision-making, the present study retains two limitations: Firstly, the scenario coverage of the experimental dataset requires expansion. The model's adaptability and stability under extreme conditions specific to inorganic chemistry experiments (such as anhydrous and oxygen-free reaction environments, high-temperature roasting processes above 800° C, and low-temperature cryogenic reaction conditions) remain insufficiently validated, potentially impacting system performance under special operating conditions; Secondly, the emergency decision-making case repository remains relatively small (containing only 500 historical incident cases), lacking sufficient support for low-probability, high-risk rare incidents (e.g., violent reactions of metal hydrides with air). This results in insufficient precision and flexibility in matching response protocols for such occurrences. To address these shortcomings, future research may focus on four key areas: Firstly, employing federated learning techniques to establish cross-institutional data collection and sharing mechanisms. This would involve collaborating with multiple universities and research institutes' inorganic chemistry laboratories to expand multimodal datasets, prioritising the inclusion of visual and sensor data from extreme scenarios such as anhydrous reactions, high-temperature smelting, and heavy metal waste treatment to enhance the model's scenario adaptability. By dynamically adjusting decision rules and case weights based on incident resolution outcomes as feedback signals, the system achieves self-learning and adaptive upgrades of response protocols, further enhancing its intelligent decision-making capabilities for complex emergencies.

Acknowledgement

1) Educational and Teaching Research and Reform Project (No. 20241110), Ordos Institute of Technology

2) Teaching Innovation Studio Project (20250401), Ordos Institute of Technology

References

[1] Alqarni A Z ,Qasim M H ,Alsuwian T , et al. Advanced Internet of Things (IoT)-Based Intelligent Heavy Transport Vehicles (HTV) Monitoring System to Enhance Passenger Safety[J].Recent Advances in Electrical and Electronic Engineering,2026,19(1):1-12.

[2] Islam A M ,Roy C S ,Utsho U F M , et al. An IoT-enabled intelligent water quality monitoring system for tourist safety using machine learning[J].Discover Electronics,2025,2(1):28.

[3] Sharifzada H ,Wang Y ,Sadat I S , et al. An Image-Based Intelligent System for Addressing Risk in Construction Site Safety Monitoring Within Civil Engineering Projects[J]. Buildings,2025, 15(8):1362.

[4] Ali R ,Lukic G W ,Miyami D , et al. Smart personal protective equipment for intelligent construction safety monitoring[J].Smart and Sustainable Built Environment,2025,14(3):835-858.

[5] Moustafa H ,Hemida H M ,Nour A M , et al. Intelligent packaging films based on two-dimensional nanomaterials for food safety and quality monitoring: Future insights and roadblocks[J].Journal of Thermoplastic Composite Materials,2025,38(3):1208-1230.

[6] Du F ,Zhou A ,Li B . Special Issue “Intelligent Safety Monitoring and Prevention Process in Coal Mines”[J].Processes,2025,13(1):85.

[7] Yani Z .Research on Intelligent Monitoring and Early Warning of Safety Risks in Tourism Food Supply Chain[C]. Proceedings of the 21st International Conference on Innovation and Management, 2024: 30-37.

[8] Xinyuan L, Hongyang H . Intelligent Video Monitoring and Analysis System for Power Grid Construction Site Safety Using Wireless Power Transfer [J]. International Journal of Information Security and Privacy (IJISP), 2024, 18 (1): 1-21.

[9] Qin X . Remote intelligent monitoring method of automobile driving safety based on Internet of Things technology [J]. International Journal of Vehicle Design, 2024, 95 (1-2): 22-37.

[10] YiJian C ,ShiWen L ,XiaoLin D , et al. The effect and safety assessment of monitoring ethanol concentration in exhaled breath combined with intelligent control of renal pelvic pressure on the absorption of perfusion fluid during flexible ureteroscopic lithotripsy[J]. International urology and nephrology, 2023, 56 (1): 45-53.