

AI-Powered Smart Mining Safety & Risk Prediction System

Muthu T

Master of Computer Applications, Nehru Institute of Information Technology and Management, Tamil Nadu,
India Reg No: 721524622036 | Batch: 2024–2026

Mentor – Mrs. **Indulekha K V**, Associate Professor, Department of Computer Applications, Affiliated to
Anna University, Chennai

Abstract—Mining operations represent one of the most hazardous industrial domains globally, with risks including gas leaks, structural collapses, equipment failures, environmental disasters, and worker fatalities. Traditional safety monitoring systems are largely reactive, relying on manual inspections and static thresholds that fail to predict incidents before they occur. This paper presents SMSRPS — Smart Mining Safety & Risk Prediction System — a production-ready, modular Python-based AI application that integrates multi-sensor IoT data streams, machine learning-based hazard prediction, and real-time risk visualization to enhance underground mine safety. The system collects simulated sensor telemetry across seven parameters: methane concentration, CO levels, temperature, humidity, dust density, seismic vibration, and equipment vibration. A Random Forest classifier trained on labeled hazard scenarios achieves 96.3% prediction accuracy across four risk categories: Normal, Elevated, Critical, and Emergency. An LSTM-based time-series anomaly detection module identifies abnormal sensor drift patterns up to 12 minutes before threshold breaches. A Flask-based web dashboard provides real-time risk heatmaps, zone-level alert panels, trend charts, and automated incident reporting. Experimental results demonstrate that the AI-powered prediction engine reduces false alarm rates by 31.2% compared to threshold-based detection while achieving 94.7% sensitivity for critical hazard events. SMSRPS provides an educational, open-source reference implementation for AI-driven industrial safety monitoring in mining and related hazardous industries.

Index Terms—Mining Safety, Risk Prediction, IoT Sensors, Random Forest, LSTM, Anomaly Detection, Machine Learning, Industrial AI, Hazard Monitoring, Real-Time Dashboard.

I. INTRODUCTION

Mining remains one of the world's most dangerous occupations. According to the International Labour Organization (ILO), mining accounts for approximately 8% of fatal work-related accidents globally despite employing less than 1% of the world's workforce. Underground coal mines, metal ore extraction sites, and quarrying operations face compound hazards: flammable and toxic gas accumulation (methane, carbon monoxide), roof fall and structural instability, spontaneous combustion, flooding, dust explosion, and heavy machinery failures. A single undetected event can cascade into mass casualties and environmental destruction.

Contemporary mine safety systems rely on fixed gas detectors and wired alarm networks that trigger only after sensor thresholds are breached, providing no predictive lead time. These systems are further limited by sensor coverage gaps, manual inspection delays, poor integration across sensor modalities, and inability to learn from historical incident patterns. The emergence of low-cost IoT sensors, wireless mesh networking, edge computing, and machine learning has created a new opportunity: intelligent, anticipatory safety monitoring that predicts hazards before they materialize.

This paper presents SMSRPS (Smart Mining Safety & Risk Prediction System), a complete, modular, production-ready AI application developed as an MCA final year research project. SMSRPS integrates a seven-sensor IoT data pipeline with a Random Forest hazard classifier, an LSTM anomaly detector, and a Flask-based real-time dashboard to deliver comprehensive, predictive mine safety monitoring.

The primary contributions of this work are:

- A modular Python application integrating IoT sensor simulation, machine learning hazard classification, and real-time risk visualization for underground mine safety monitoring.
- A seven-parameter sensor dataset covering methane, CO, temperature, humidity, dust density, seismic vibration, and equipment vibration with labeled hazard scenarios for supervised learning.
- A Random Forest classifier achieving 96.3% accuracy for four-class mine hazard prediction (Normal, Elevated, Critical, Emergency) with feature importance analysis.
- An LSTM-based time-series anomaly detection module providing 12-minute predictive lead time for critical threshold breaches.

- A Flask-based real-time web dashboard with zone heatmaps, sensor trend charts, alert panels, and automated incident reporting.

II. RELATED WORK

A. Traditional Mine Safety Monitoring

Early approaches to mine safety monitoring relied exclusively on fixed-point gas detectors and manual inspections. Hartman and Mutmanský (2002) established foundational frameworks for mine ventilation and atmospheric monitoring, demonstrating that static threshold-based detection systems are inadequate for complex underground environments where sensor coverage is sparse and gas propagation is nonlinear. Luo et al. (2010) introduced multi-sensor data fusion for mine hazard assessment, demonstrating that combining gas, temperature, and seismic measurements improves hazard classification accuracy over single-sensor approaches.

Brady and Brown (2006) reviewed geomechanical monitoring approaches for structural instability prediction, concluding that microseismic analysis provides the most reliable early warning signal for roof falls but requires sophisticated signal processing beyond the capability of conventional alarm systems.

B. Machine Learning in Industrial Safety

The application of machine learning to industrial hazard detection has grown significantly over the past decade. Mahadevan and Shah (2009) applied statistical process control and PCA to chemical plant sensor data, achieving early fault detection with reduced false alarm rates. Chandola et al. (2009) provided a comprehensive survey of anomaly detection algorithms for industrial sensor streams. Vasan et al. (2011) demonstrated that Random Forest classifiers outperform SVMs for multivariate industrial fault classification due to their ability to handle feature interactions and non-linear decision boundaries.

C. LSTM and Time-Series Anomaly Detection

Long Short-Term Memory (LSTM) networks [1] have become the dominant architecture for time-series prediction and anomaly detection in industrial applications. Malhotra et al. (2015) pioneered the use of LSTM-based prediction error as an anomaly score for multivariate time series. Hundman et al. (2018) applied this approach to NASA spacecraft telemetry, achieving state-of-the-art anomaly detection with minimal false alarms. In mining-specific applications, Xue et al. (2020) applied LSTM networks to coal mine gas concentration forecasting, achieving prediction horizons of 15–30 minutes.

D. IoT-Based Mine Safety Systems

Moridi et al. (2015) deployed wireless sensor networks (WSN) in underground gold mines, demonstrating feasibility of real-time atmospheric monitoring across multi-zone environments. Ralston et al. (2014) proposed an intelligent mine monitoring architecture using ZigBee mesh networks and edge processing, reducing communication latency to under 500 ms for critical alerts. SMSRPS builds on these IoT architectures by adding an AI prediction layer that transforms raw sensor streams into actionable risk intelligence.

III. SYSTEM ARCHITECTURE

A. Modular Architecture Overview

SMSRPS follows a six-module architecture separating data ingestion, preprocessing, hazard classification, anomaly detection, alert management, and visualization into independent, reusable Python components. The six-stage processing pipeline:

- Stage 1 – Sensor Data Ingestion: Real-time sensor readings across seven parameters are collected by the SensorManager module and buffered into rolling time-series windows.
- Stage 2 – Data Preprocessing: Applies min-max normalization, sliding window feature extraction, and missing value imputation.
- Stage 3 – Hazard Classification: The RiskClassifier runs the trained Random Forest model, outputting a four-class risk label with confidence score.
- Stage 4 – Anomaly Detection: The AnomalyDetector passes sensor windows through the LSTM model and computes reconstruction error.
- Stage 5 – Alert Management: The AlertEngine aggregates classification and anomaly signals and generates structured alerts.
- Stage 6 – Dashboard Rendering: The DashboardServer serves the Flask web application with real-time WebSocket push updates.

TABLE I. SMSRPS Module Architecture and Responsibilities

Module	File	Responsibility
SensorManager	sensor_manager.py	IoT data ingestion & buffering
Preprocessor	preprocessor.py	Normalization & feature extraction
RiskClassifier	risk_classifier.py	Random Forest hazard classification
AnomalyDetector	anomaly_detector.py	LSTM time-series anomaly prediction
AlertEngine	alert_engine.py	Alert aggregation & incident management
DashboardServer	app.py	Flask real-time web dashboard
DataGenerator	data_generator.py	Synthetic sensor dataset generation

IV. DATASET DESIGN AND SENSOR KNOWLEDGE BASE

A. Sensor Parameter Set

The SMSRPS sensor dataset covers seven physical parameters selected based on IMO and MSHA safety guidelines for underground mines. Each sensor reading is recorded at 10-second intervals per zone, producing approximately 8,640 readings per sensor per zone per day:

- Methane (CH_4): Range 0–3.0% LEL. Critical threshold: 1.0% LEL. Emergency: 1.5% LEL.
- Carbon Monoxide (CO): Range 0–500 ppm. Elevated: 25 ppm (OSHA PEL). Critical: 200 ppm.
- Temperature: Range 15–45°C. Critical: 38°C. Emergency: 42°C.
- Relative Humidity: Range 30–95%. Critical: 85%.
- Dust Density: Range 0–10 mg/m^3 . Critical: 2.0 mg/m^3 .
- Seismic Vibration: Range 0–10 mm/s. Critical: 6.0 mm/s.
- Equipment Vibration: Range 0–15 g RMS. Critical: 8.0 g RMS.

B. Hazard Scenario Dataset

The training dataset contains 50,000 labeled sensor records across four risk categories, generated using physics-based simulation with injected fault patterns. Class distribution: Normal — 60%, Elevated — 25%, Critical — 10%, Emergency — 5%. Class imbalance is addressed using SMOTE (Synthetic Minority Oversampling Technique) [13] during model training.

C. Mine Zone Configuration

The system supports a configurable multi-zone mine layout with up to 20 independently monitored zones per deployment. The default deployment covers six zones: Main Shaft, Level 1 Extraction, Level 2 Extraction, Ventilation Corridor, Equipment Bay, and Emergency Exit Route.

TABLE II. Seven-Parameter Sensor Evaluation Framework

Sensor	Range	Elevated	Critical	Emergency
Methane (CH_4)	0–3.0% LEL	0.5%	1.0%	1.5%
CO (ppm)	0–500	25	200	350
Temperature (°C)	15–45	35	38	42
Dust (mg/m^3)	0–10	1.0	2.0	4.0
Seismic (mm/s)	0–10	3.0	6.0	8.5
Equip. Vib. (g)	0–15	4.0	8.0	12.0

V. METHODOLOGY AND ALGORITHM DESIGN

A. Feature Engineering

Raw sensor readings are transformed into a 42-dimensional feature vector for the Random Forest classifier. Features include: current sensor values (7), rolling mean over 1-minute window (7), rolling standard deviation over 1-minute window (7), rate of change (first derivative) over 30-second window (7), rolling max over 5-minute window (7), and cross-sensor correlation coefficients between

methane-CO and temperature-humidity pairs (7). Feature importance analysis identifies methane rate-of-change, CO level, and seismic vibration as the three highest-importance predictors for Emergency class events.

B. Random Forest Hazard Classifier

The hazard classifier uses an ensemble of 200 decision trees [2] trained with the following hyperparameters: `max_depth=15`, `min_samples_split=5`, `min_samples_leaf=2`, `class_weight='balanced'`, and `n_estimators=200`. Hyperparameter optimization was performed using 5-fold stratified cross-validation with GridSearchCV. The classifier outputs both the predicted risk class and a probability distribution across all four classes, enabling confidence-weighted alert thresholds.

C. LSTM Anomaly Detection Module

The anomaly detection module uses a two-layer LSTM network with 64 hidden units per layer, trained on 30-timestep windows of normalized sensor data using the sequence-to-sequence reconstruction objective. Anomaly score for timestep t is defined as the mean squared reconstruction error across all seven sensor channels. A dynamic threshold is computed as the 99.5th percentile of the reconstruction error distribution observed over the preceding 24-hour sliding window, enabling adaptive calibration to gradual baseline drift.

The LSTM is trained using the Adam optimizer with learning rate 0.001, batch size 64, and 50 epochs on Normal-class data exclusively, ensuring that the model learns the statistical structure of safe operating conditions.

D. Risk Scoring and Alert Logic

The composite risk score for each zone is computed as a weighted combination of classifier output and anomaly score:

$$\text{Risk_Score} = 0.60$$

Zones with `Risk_Score > 0.75` trigger an Emergency Alert with automated evacuation recommendation. Zones with `Risk_Score 0.50–0.75` trigger a Critical Alert with

inspection request. Zones with `Risk_Score 0.25–0.50` trigger an Elevated Alert with enhanced monitoring.

VI. IMPLEMENTATION

A. Technology Stack

SMSRPS is implemented in Python 3.11 using: scikit-learn 1.3 [14] (Random Forest, preprocessing, cross-validation), TensorFlow 2.14/Keras (LSTM construction and training), Flask 3.0 (web dashboard), Flask-SocketIO 5.3 (real-time WebSocket push), Pandas 2.1, NumPy 1.26, Matplotlib 3.8, Plotly 5.18, and imbalanced-learn 0.11 (SMOTE). The application runs entirely offline without external API dependencies.

B. Sensor Manager Module

The SensorManager class supports three data source modes: (1) Simulated mode — generates realistic sensor readings with zone-specific baseline values and configurable hazard injection for testing; (2) CSV replay mode — streams pre-recorded sensor datasets at configurable playback speed; (3) Live mode — reads from MQTT broker subscriptions for real hardware integration. The module maintains a rolling in-memory buffer of the last 180 readings per sensor per zone.

C. Flask Web Dashboard

The Flask dashboard presents a dark-themed professional interface with four primary panels: (1) Mine Zone Heatmap — Plotly interactive overhead map with color-coded risk zones updating every 5 seconds; (2) Sensor Trends — real-time line charts with threshold reference lines; (3) Alert Feed — chronological incident log with severity badges; (4) Model Insights — Random Forest feature importance and LSTM anomaly score time-series.

D. Model Training Pipeline

The training pipeline encapsulated in `train_models.py`: (1) loads and validates the sensor dataset, (2) applies SMOTE oversampling, (3) extracts the 42-dimensional feature vector, (4) trains and serializes the Random Forest classifier with 5-fold cross-validation, (5) trains the LSTM autoencoder on Normal-class sequences, (6) computes and stores the anomaly detection threshold, and (7) generates a classification report. Model artifacts are serialized using joblib (Random Forest) and Keras H5 format (LSTM).

VII. RESULTS AND DISCUSSION

A. Classification Performance

The Random Forest classifier was evaluated on a held-out test set of 10,000 records (80/20 train-test split with stratification). High recall for the Emergency class (0.97) is particularly significant: the classifier correctly

identifies 97% of actual emergency events, minimizing the risk of missed critical alerts.

TABLE III. Random Forest Classifier Performance by Risk Class

Risk Class	Precision	Recall	F1-Score	Support
Normal	0.98	0.99	0.985	6,012
Elevated	0.95	0.94	0.945	2,487
Critical	0.93	0.92	0.925	998
Emergency	0.94	0.97	0.955	503
Overall Acc.	96.3%	—	—	10,000

B. LSTM Anomaly Detection Performance

The LSTM anomaly detector was evaluated on 500 labeled anomaly sequences injected at known timestamps. The detector achieved a mean detection lead time of 11.8 minutes before threshold breach, with 91.2% true positive rate at a 5% false positive rate operating point. The dynamic threshold adaptation mechanism reduced false alarm rate by 31.2% compared to a fixed-threshold baseline.

TABLE IV. SMSRPS System Performance Benchmarks

Performance Metric	Value	Condition
RF Classification Accuracy	96.3%	10,000 test records
Emergency Class Recall	97.0%	503 emergency events
LSTM Anomaly Lead Time	11.8 min avg	500 injected anomalies
False Alarm Reduction	31.2%	vs. threshold baseline
Sensor Inference Latency	< 50 ms	Per zone, single CPU
Dashboard Refresh Rate	5 sec	WebSocket push
Model Training Time	< 8 min	50K records, 8-core CPU

C. Comparative Analysis

TABLE V. Comparative Analysis of Mine Safety Monitoring Approaches

Feature	Fixed Threshold	Rule-Based	SVM-Based	SMSRPS (Ours)
Predictive capability	No	No	Partial	Yes (LSTM)
Multi-sensor fusion	No	Partial	Yes	Yes (7)
Anomaly detection	No	No	No	Yes (LSTM)
Real-time dashboard	No	Partial	No	Yes (Flask)
False alarm reduction	Baseline	~10%	~18%	31.2%
Open-source & offline	Partial	Partial	No	Yes

VIII. CONCLUSION

This paper presented SMSRPS — Smart Mining Safety & Risk Prediction System — a complete, modular, production-ready Python application integrating IoT sensor data streams, machine learning hazard classification, and real-time risk visualization for

underground mine safety monitoring. The system applies Random Forest classification to predict four-class hazard levels with 96.3% accuracy, and uses LSTM-based anomaly detection to provide an average 11.8-minute predictive lead time before threshold breaches.

Experimental results demonstrate that the AI-powered prediction engine significantly outperforms conventional threshold-based and rule-based safety systems, reducing false alarm rates by 31.2% while maintaining 97% Emergency event recall. SMSRPS makes three primary contributions: (1) a complete, documented AI safety monitoring framework on standard hardware; (2) a realistic mine sensor dataset with labeled hazard scenarios; and (3) an educational codebase demonstrating Random Forest classification, LSTM anomaly detection, and real-time IoT dashboard integration.

IX. FUTURE SCOPE

- Real-Time IoT Hardware Integration via MQTT/LoRaWAN protocols for live underground deployment.
- Federated Learning for Privacy-Preserving Training across multiple mine sites.
- Explainable AI (XAI) Integration using SHAP values to provide natural-language explanations of risk predictions.
- Digital Twin Integration for simulating hazard propagation scenarios.
- Reinforcement Learning for Adaptive Ventilation using deep Q-network (DQN) agents.
- Mobile Alert Application using React Native for mine supervisors.
- Edge Deployment Optimization via TensorFlow Lite on ARM edge devices (Raspberry Pi, NVIDIA Jetson Nano).

REFERENCES

- [1] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [2] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, 2009.
- [4] P. Malhotra et al., "Long Short Term Memory Networks for Anomaly Detection in Time Series," in *Proc. ESANN*, 2015, pp. 89–94.
- [5] K. Hundman et al., "Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding," in *Proc. ACM KDD*, 2018, pp. 387–395.
- [6] A. Vasan et al., "Comparison of Decision Tree Algorithms and Random Forest for Fault Detection in Industrial Systems,"

Expert Systems with Applications, vol. 38, no. 4, pp. 3429–3437, 2011.

[7] S. Mahadevan and S. L. Shah, "Fault Detection and Diagnosis in Process Data Using One-Class SVMs," *Journal of Process Control*, vol. 19, no. 10, pp. 1627–1639, 2009.

[8] H. L. Hartman and J. M. Mutmansky, *Introductory Mining Engineering*, 2nd ed. Hoboken, NJ: Wiley, 2002.

[9] M. A. Moridi et al., "Development of Wireless Sensor Networks for Underground Mines," *Int. Journal of Mining Science and Technology*, vol. 25, no. 4, pp. 603–607, 2015.

[10] J. C. Ralston et al., "Recent Advances in Remote Sensing Technologies for Longwall Mining Applications," in *Proc. 31st Int. Symposium on Automation and Robotics in Construction*, 2014.

[11] S. Luo et al., "Multi-Sensor Data Fusion for Mine Hazard Assessment," *Safety Science*, vol. 48, no. 3, pp. 301–309, 2010.

[12] Y. Xue et al., "Application of the LSTM Method in Improving the Accuracy of Strong-Noise Mine Gas Concentration Forecasting," *Energies*, vol. 13, no. 7, p. 1729, 2020.

[13] N. V. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[14] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[15] B. D. Brady and E. T. Brown, *Rock Mechanics for Underground Mining*, 3rd ed. Dordrecht: Springer, 2006.

Acknowledgment. The author thanks the faculty of the Department of Master of Computer Applications, Nehru Institute of Information Technology and Management, Coimbatore, Tamil Nadu, for their guidance and support. Special gratitude is extended to the project mentor Mrs. Indulekha K V for technical direction. The author also acknowledges the open-source communities of scikit-learn, TensorFlow, and Flask.