MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6453-6463

PREDICTING FIRE RISK USING MACHINE LEARNING AND REAL-TIME ACCESS CONTROL SYSTEM DATA: EVIDENCE FROM KIGALI SPECIAL ECONOMIC ZONE (KSEZ)

Adventist University of Central Africa (AUCA)
Author:
Elisabeth GIRAMATA
elise.giramata@gmail.com
Co-authors:
Prof. Sebagenzi Jason
Associate Professor
Adventist University of Central Africa-AUCA
Jason.sebagenzi@auca.ac.rw
Dr. Hategekimana Jean Paul, PhD
Lecturer of R. Methodology at BU, Delaware State, USA
Editorial Member of BJBST-Delaware State, USA
Examiner of Africa Leadership University (ALU)
Senior Researcher and Advisor in Research Department, Brainae University
hategejp6@gmail.com
Dr. Oswald Rudahigwa
Senior Lecturer
Kigali Independent University
Rudahigwa77@gmail.com
Ineza Yves
Assistant Lecturer-ICT
Rwanda Polytechnic-Kigali College
ineyves87@gmail.com
Nobel Uhagaze
Assistant Lecturer in the Department of Computer Science
Kigali Independent University
unobel13@gmail.com

ABSTRACT
This study investigates the application of machine learning techniques in predicting fire risk within industrial environments, using the Kigali Special Economic Zone (KSEZ) as a case study. The research integrates environmental monitoring data and real-time access control system data to develop predictive models capable of identifying fire risk levels. A dataset comprising 80,000 observations and 11 variables was analyzed using Python-based machine learning models, including Logistic Regression, Random Forest, and XGBoost. The findings reveal that fire risk is not random but can be predicted using key indicators such as equipment heat index, temperature, hazard level, and abnormal access patterns. Among the models tested, Logistic Regression achieved the highest accuracy of 99.81%, demonstrating strong predictive capability. The study further shows that integrating environmental and behavioral data significantly improves model performance compared to single-source models.

1. Introduction

In the era of Industry 4.0, technological innovation has become a cornerstone of risk management strategies across industrial zones. High-value infrastructure, human capital, and raw materials are increasingly concentrated within these zones, making them highly vulnerable to disasters such as fires. In response, global industrial hubs have turned to **integrated technological solutions** to move from reactive to proactive fire risk management models. Notably, countries such as the United States, Germany, and Japan have pioneered the use of **Internet of Things (IoT)** technologies and **simulation-driven e-learning platforms** to revolutionize how fire hazards are detected, prevented, and mitigated (National Fire Protection Association, 2021; EFSA, 2022). The deployment of **real-time access control systems** featuring surveillance cameras, biometric scanners, motion detectors, and RFID-enabled entry points has enabled these nations to **track personnel movement, detect anomalies, and control access** to sensitive areas in real time. When paired with predictive analytics, these tools provide not only event monitoring but also **forecasting capabilities**, allowing organizations to detect patterns that signal potential fire risks before they escalate (Zhang *et al.,* 2021). Rwanda have yet to **fully capitalize on the synergistic potential** of these technologies. The **Kigali Special Economic Zone (KSEZ),** Rwanda's flagship industrial park, has witnessed significant growth in manufacturing, logistics, and warehousing activities. However, the increased complexity and density of operations have also heightened the risk of fire incidents. Between 2018 and 2022, the zone recorded an average of **five fire outbreaks annually**, often resulting in substantial financial and material losses (Rwanda National Police, 2023). While some fire safety measures exist in KSEZ, including **manual inspections and occasional training sessions**, these are largely reactive and fragmented.

Problem Statement: Despite being Rwanda's primary industrial hub, KSEZ lacks a comprehensive and integrated fire risk management system. Current approaches are largely reactive, relying on basic fire detection equipment and periodic physical training sessions. These systems are insufficient for identifying early warning signs of fire risk. Previous studies have highlighted key gaps such as Limited use of real-time monitoring systems; lack of predictive analytics in fire risk management; and Minimal integration of behavioral data such as access control systems. Furthermore, existing research has not adequately explored the predictive capability of real-time access control data; context-specific applications in African industrial zones; and Integration of environmental and behavioral data for fire prediction. As a result, fire incidents continue to occur due to the inability of traditional systems to detect risk patterns early. This study addresses these gaps by developing a machine learning-based predictive model that integrates environmental and access control data to enhance fire risk prediction in KSEZ.

Research Questions: This study answered the questions follows:

(i) What environmental factors significantly influence fire risk in industrial settings?
(ii) How can access control system data contribute to predicting fire risk?
(iii) Which machine learning models are most effective in predicting fire risk based on environmental and access control data?
(iv) Does integrating environmental and access control data improve the accuracy of fire risk prediction compared to using either data source alone?
(v) Which factors, identified through feature importance analysis, have the greatest impact on fire risk?
(vi) How well does the developed predictive model perform under real-world and scenario-based conditions?

Objectives of the Study:

To develop a machine learning model for predicting fire risk using environmental and access control system data. While, specific objectives are:

(i) To analyze environmental factors influencing fire risk in industrial settings
(ii) To evaluate the role of access control system data in predicting fire risk
(iii) To develop and evaluate machine learning models for fire risk prediction
(iv) To assess the effectiveness of integrating environmental and access control data
(v) To identify key predictors of fire risk using feature importance analysis
(vi) To validate the predictive model using real-world and scenario-based conditions

Hypotheses of the Study

The study tests the following null hypotheses:

(i) H_{01}: Environmental factors do not significantly influence fire risk

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6453-6463

(ii) H_{02}: Access control data does not significantly contribute to fire risk prediction
(iii) H_{03}: Machine learning model type does not affect prediction accuracy
(iv) H_{04}: Data integration does not improve prediction accuracy
(v) H_{05}: Feature importance analysis does not identify significant predictors
(vi) H_{06}: The predictive model does not perform significantly under real-world conditions

2. Review of Literature

Technological innovation has become indispensable in addressing the complexities of fire risk management in industrial zones. As industrial hubs grow in scale and sophistication, the need for integrated safety solutions has escalated. Two technologies including Real-Time Access Control Systems (RTACS) and E-Learning Platforms have emerged as critical tools in the prevention, prediction, and management of fire incidents. However, while each has shown measurable individual impact, there is a lack of scholarly investigation into their combined application, particularly in emerging economies such as Rwanda. Jones & Clark (2022) observed that RTACS improve incident response time and minimize unauthorized access in sensitive industrial areas. Smith *et al*., (2023**)** emphasized the potential of predictive analytics, derived from real-time data, to identify fire risks before they materialize.

Real-Time Access Control Systems (RTACS) and Fire Risk Mitigation.

The RTACS refer to integrated technologies designed to monitor and regulate access to physical areas in real-time. These systems ensure that only authorized personnel can enter sensitive locations, while also tracking patterns of movement and triggering alerts during anomalies. Their value in fire risk management lies in their capacity to provide predictive intelligence, access accountability, and rapid emergency response.

Surveillance Cameras

Surveillance systems serve as the foundation of RTACS. In fire-sensitive zones, AI-powered CCTV systems can detect visual cues such as smoke, heat waves, and sudden motion, prompting early warning signals. According to Zhang *et al.* (2021), the integration of surveillance with machine vision has reduced fire-related damages by up to 32% in Chinese industrial parks.

Biometric Scanners

These devices enhance site security by using unique physical attributes (e.g., fingerprints, retina scans, facial recognition) to grant or deny access. In high-risk areas, this ensures that only trained and certified personnel operate within zones containing combustible materials or heavy machinery (Jia & Thangavel, 2022).

RFID Card Readers

RFID-based entry systems offer real-time monitoring of personnel movement. Access logs can be analyzed to determine whether individuals are spending unusual amounts of time in risky areas or entering zones outside their clearance level. Kuo *et al.* (2023) suggest that RFID systems have enhanced accountability and reduced workplace accidents in Japanese manufacturing hubs.

Motion Detectors

Motion sensors augment safety by identifying abnormal movement or human presence in restricted areas, particularly during off-hours. These are often integrated with fire suppression systems to activate alarms, sprinklers, or lockdown protocols (Mbatha *et al.,* 2021). While such systems are increasingly implemented in developed countries, adoption remains limited in Rwanda. RTACS in KSEZ are often standalone and lack the sophistication necessary to generate predictive insights or interface with other safety systems (Rwanda Development Board, 2022)..

Technology Acceptance Model (TAM)

Originally developed by Davis (1989) and refined in subsequent studies, TAM explains why people decide to embrace or reject new technologies. The model highlights two important perceptions: how useful the technology is perceived to be and how easy it is to use. In the context of this study, TAM suggests that if employees and managers at KSEZ view the real-time access control systems and e-learning platforms as practical and user-friendly, they are more likely to adopt them successfully. Recent studies (e.g., Venkatesh et al., 2020) confirm that these perceptions remain strong predictors of technology adoption in workplace safety systems.

Diffusion of Innovation Theory.Proposed by Rogers (2003), this theory describes the process through which new ideas and technologies spread within a community or organization. It identifies key factors such as the relative advantage of the innovation, how compatible it is with existing practices, the complexity of the innovation, opportunities for trial, and visibility of its benefits. Applying this to KSEZ, the theory helps explain how introducing access control and e-learning technologies can gradually change safety culture and practices, encouraging broader acceptance and sustained use. Recent applications (Greenhalgh *et al.,* 2017) emphasize the importance of these factors in adopting safety innovations in industrial settings.

Behavioral Learning Theory.Grounded in the work of Skinner (1953) and Bandura (1977), behavioral learning theory focuses on how behaviors are acquired or changed through conditioning and reinforcement. This theory supports the use of e-learning as an effective tool for improving fire safety awareness. Regular training, assessments, and simulated scenarios help reinforce safe practices, making individuals more prepared to respond appropriately during fire emergencies. Contemporary research (Zimmerman & Schunk, 2019) highlights the effectiveness of digital learning platforms in reinforcing behavioral change in workplace safety.

Risk Management Theory.Risk management theory offers a systematic approach to identifying, analyzing, and mitigating potential hazards (Hopkin, 2018). It emphasizes proactive strategies, including the use of predictive analytics and real-time monitoring, to reduce risks before they materialize. Within this study, risk management theory underpins the use of data-driven models like the Python-based predictive tool to forecast fire hazards and improve prevention strategies in KSEZ. This approach aligns with recent industry trends towards integrating technology for enhanced safety outcomes (Aven, 2020). These theories collectively provide a strong foundation to explore how digital technologies and e-learning interventions can enhance fire risk prediction and safety management in industrial contexts. They offer insights into technology adoption, the spread of innovation, behavioral change through learning, and structured risk reduction all critical elements for this study.

Conceptual Framework

The conceptual framework for this study illustrates the relationship between environmental factors and behavioral indicators (independent variables) and fire risk (dependent variable), mediated through machine learning models.

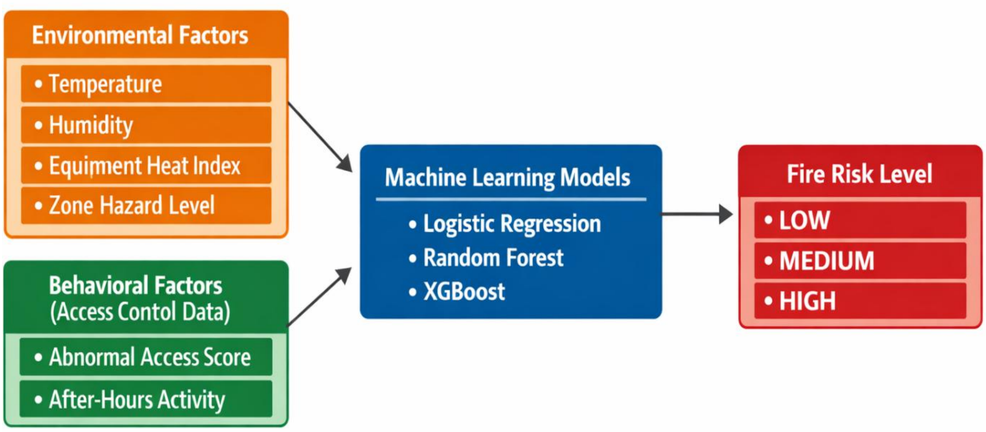


Figure 1: Visual Representation

3. Research Design and Methodology: Fire risk is treated as a measurable phenomenon that can be predicted through observable environmental and behavioral indicators. The positivist approach is appropriate because the study relies on structured datasets; variables are measurable and quantifiable; hypotheses are statistically testable; and results are validated through performance metrics.

Research Design: The study adopts a **quantitative predictive research design**. Therefore, the design integrates descriptive statistics (to understand data patterns), Correlational analysis (to examine relationships between variables), Supervised machine learning classification (to predict outcomes). The dependent variable is categorical (LOW, MEDIUM, HIGH), which makes classification algorithms the most suitable analytical technique. The overall research design follows a structured pipeline: data collection; data cleaning; feature selection; model training; model testing; model comparison; and model validation.

Population and Sample Size: The target population of this study consists of all fire-related environmental and access control records generated within the Kigali Special Economic Zone (KSEZ) during the study period. Specifically, the population includes all system-generated observations containing environmental variables (such as temperature, humidity, equipment heat index, and zone hazard level), behavioral indicators (such as abnormal access score and after-hours activity), and the corresponding fire risk classifications (LOW, MEDIUM, HIGH). In practical terms, the population represents the complete set of recorded operational events within KSEZ that meet the inclusion criteria for fire risk analysis. Because the study focuses on predictive modeling using structured system data, the population is defined in terms of data records rather than individuals. This ensures that all relevant industrial conditions and behavioral patterns are captured for comprehensive analysis. The study utilizes a large dataset consisting of 80,000 observations and 11 variables. The sample size for this research is N = 80,000 records. Since this dataset includes all available and eligible records extracted from the system during the selected period, the study applies a full-data approach rather than a limited random sampling technique. After data cleaning and preprocessing, the complete dataset is used for machine learning modeling. To ensure proper model training and unbiased evaluation, the sample is divided into two subsets: Training Set: 64,000 observations (80%); and Testing Set: 16,000 observations (20%).

Data Processing Procedures

The several systematic steps are undertaken to transform raw data into a clean and structured format suitable for analysis. The first step involves data cleaning. At this stage, the dataset is examined to identify and address errors or irregularities. Duplicate records are removed to prevent repetition that could bias the results. Missing values are carefully handled using appropriate statistical techniques to avoid distortions in the analysis. In addition, inconsistencies within the data such as incorrect entries, formatting errors, or contradictory information are corrected. This step ensures that the dataset accurately reflects real conditions and maintains integrity throughout the analysis process. After cleaning, the data undergoes normalization. Since the variables in this study are measured using different scales and units, normalization is necessary to ensure fairness during model training. By scaling the variables to a comparable range, no single variable dominates the analysis due to its numerical magnitude. This process enhances the stability and performance of machine learning algorithms, particularly those that are sensitive to differences in variable scale. The next step involves feature engineering, which focuses on improving the predictive power of the dataset. During this process, new variables may be created from existing data to better capture hidden patterns or relationships. For example, derived indicators may be constructed to summarize complex behaviors or environmental interactions. Feature engineering strengthens the model's ability to detect meaningful trends and improves overall prediction accuracy. Finally, the dataset is divided into two main subsets to support proper model evaluation. Seventy percent (70%) of the data is allocated for training the machine learning models, while the remaining thirty percent (30%) is reserved for testing. The training set is used to teach the model how to recognize patterns and relationships among variables. The testing set, on the other hand, is used to evaluate how well the model performs on new, unseen data.

Machine Learning Models

To develop an accurate and reliable fire risk prediction system, this study applies three different classification algorithms. The first model is Logistic Regression, which is used as a baseline model. It is a widely recognized statistical technique for classification problems and is particularly useful when the dependent variable is categorical. The second model is Random Forest, which is an ensemble learning method. The third model is XGBoost (Extreme Gradient Boosting), which is a highly efficient and powerful machine learning algorithm. XGBoost is known for its strong predictive performance, speed, and ability to handle structured data effectively.

Mathematical Formulation of the Models

Logistic regression estimates the probability that an observation belongs to a particular class. The model uses the logistic (sigmoid) function to transform linear combinations of predictors into probability values between 0 and 1. The probability of fire risk occurrence is expressed as:

$$P(Y=1)=\frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n)}}$$

In this equation: P(Y=1) represents the probability of a specific fire risk category. $X_1, X_2, \ldots, X_n$ represent the independent variables; β_0 is the intercept; $\beta_1, \beta_2, \ldots, \beta_n$ are the model coefficients; e represents the exponential function

XGBoost Mathematical Framework

XGBoost is based on gradient boosting principles. It builds a sequence of decision trees where each new tree attempts to correct the errors made by the previous ones. The prediction for an observation is defined as:

$$\hat{y}_i=\sum_{k=1}^{K} f_k(x_i)$$

Where, yi is the predicted value; fk represents individual decision trees; K is the total number of trees
xi represents input variables; The objective function that XGBoost seeks to minimize is:

$$Obj=\sum_{i=1}^{n} L(y_i,\hat{y}_i)+\sum_{k=1}^{K}\Omega(f_k)$$

In this equation: L (yi, y^i) is the loss function, measuring prediction error; Ω(fk) is the regularization term that controls model complexity; The regularization term is expressed as:

$$\Omega(f)=\gamma T+\frac{1}{2}\lambda||w||^2$$

Where, *T* is the number of leaves in the tree; *w* represents the leaf weights; γ and λ are regularization parameters. These combination of Logistic Regression, Random Forest, and XGBoost provides a strong analytical framework for fire risk prediction. The mathematical foundations of these models enhance transparency, scientific validity, and methodological rigor in this study.

4. Results and Discussion of Findings

The results were generated from environmental monitoring data and real-time access control system data integrated into a predictive framework. The analysis includes dataset overview, model performance evaluation, feature importance, correlation analysis, and scenario-based validation. The data were analyzed using Python with machine learning libraries such as Scikit-Learn and XGBoost. The dataset contains 80,000 observations and 11 variables, representing environmental conditions, operational indicators, and access control signals within KSEZ.

Table 1: Sample of observations extracted from Load the Dataset

N	Event Hour	Day of Week	Zone Hazard Level	Temperature (°C)	Humidity (%)	Equipment Heat Index	Occupancy	Abnormal Access Score	Smoke Indicator	After Hours Flag	Fire Risk Level
0	6	Sunday	2	22.72	43.43	0.606	13	0.222	0	0	Low
1	19	Saturday	2	29.79	57.55	0.401	11	0.719	0	0	Medium
2	14	Saturday	3	18.85	39.21	0.957	8	0.745	0	0	Medium
3	10	Saturday	1	31.58	36.43	0.64	16	0.065	0	0	Low
4	7	Monday	4	24.19	25.81	0.754	18	0.13	0	0	Medium

The first observation was recorded at 06:00 hours on Sunday in a zone with hazard level 2. At that time, the temperature measured 22.72°C, while humidity was 43.43%, reflecting relatively moderate environmental conditions. The equipment heat index was 0.606, indicating moderate machinery heat output, and

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6453-6463

the occupancy level was estimated at 13 individuals. The abnormal access score was relatively low at 0.222, suggesting normal access activity. With no smoke detected and no after-hours operation, the predictive system classified the situation as LOW fire risk. The second observation occurred at 19:00 hours on Saturday, also in a hazard level 2 zone. Environmental conditions were warmer, with temperature increasing to 29.79°C and humidity rising to 57.55%. The equipment heat index decreased slightly to 0.401, while the occupancy level was recorded at 11 individuals. However, the abnormal access score increased significantly to 0.719, indicating unusual access activity within the facility. Although no smoke was detected and the event occurred during operational hours, the model classified this situation as MEDIUM fire risk, largely due to the higher temperature and abnormal access behavior.The third observation was recorded at 14:00 hours on Saturday within a hazard level 3 zone, representing a relatively higher-risk industrial area. The temperature measured 18.85°C, with humidity at 39.21%, which are relatively moderate environmental conditions. However, the equipment heat index was extremely high at 0.957, indicating significant machinery heat generation. The occupancy level was 8 individuals, and the abnormal access score reached 0.745, suggesting irregular facility access. The fourth observation occurred at 10:00 hours on Saturday in a hazard level 1 zone, which represents the lowest industrial risk category in the dataset. During this observation, temperature was relatively high at 31.58°C, while humidity measured 36.43%. The equipment heat index was 0.640, and occupancy reached 16 individuals, indicating active operational activity. However, the abnormal access score was extremely low at 0.065, suggesting normal access behavior. The fifth observation was recorded at 07:00 hours on Monday in a hazard level 4 zone, which represents one of the most sensitive industrial environments in the facility. Environmental measurements indicated a temperature of 24.19°Cand humidity of 25.81%. The equipment heat index was relatively high at 0.754, while the occupancy level reached 18 individuals, which was the highest recorded among the sample observations.

Distribution of Fire Risk Levels

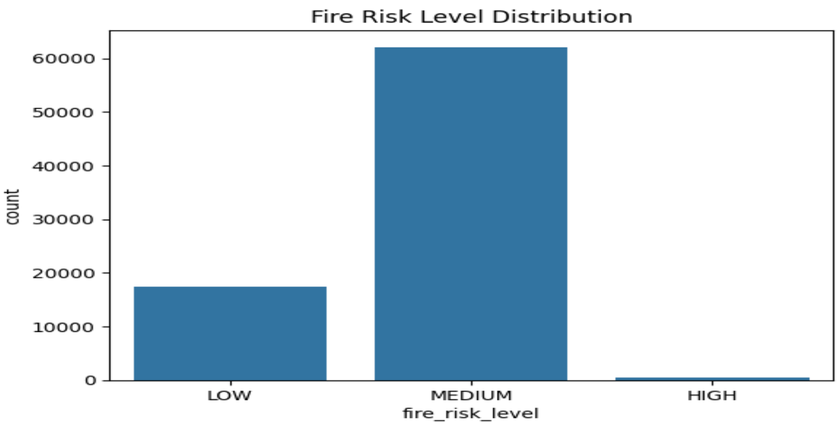
Examining this distribution helps determine whether the dataset is balanced and provides insight into how frequently different risk conditions occur within the industrial environment.

Table 2: Distribution of Fire Risk Levels

Fire Risk Level	Frequency	Percentage (%)
Low	17,415	21.77
Medium	62,121	77.65
High	464	0.58
Total	**80,000**	**100**

The results show that the majority of observations fall under the Medium fire risk level, representing 62,121 records, or approximately 77.65% of the dataset. This indicates that most operational conditions within the industrial zone are neither completely safe nor highly dangerous, but instead fall within moderate risk situations that require continuous monitoring and preventive attention. The Low fire risk level accounts for 17,415 observations, representing 21.77% of the total data. These cases reflect normal operational conditions where environmental and behavioral indicators remain within safe thresholds. In contrast, the High fire risk level represents only 464 observations, which is approximately 0.58% of the dataset. Although this percentage is very small, it is extremely significant from a safety perspective.

Figure 1: Fire Risk Level Distribution



The dataset is highly imbalanced, with most cases classified as MEDIUM risk. High-risk events are rare but critical.

Data Splitting: This step is very important in machine learning because it ensures that the model is trained on one portion of the data and tested on completely new, unseen data.

Table 3: Train-Test Split

Dataset	Observations	Percentage
Training Set	64,000	80%
Testing Set	16,000	20%

The training set consists of 64,000 observations, representing 80% of the total data. This portion was used to train the machine learning models, allowing them to learn patterns and relationships between environmental, behavioral variables, and fire risk levels. By exposing the models to a large amount of data during training, they are able to better understand underlying trends and improve their predictive capability. The remaining 16,000 observations, representing 20% of the dataset, were assigned to the testing set. The 80/20 split is widely accepted in predictive modeling and provides a balanced approach between learning and validation.

Model Performance Comparison: The performance of the three machine learning models such as Logistic Regression, Random Forest, and XGBoost was evaluated to determine which algorithm most effectively predicts fire risk levels within the dataset. Model comparison is an important step because it helps identify the most reliable and accurate approach for practical implementation.

Table 4: Machine Learning Model Performance

Model	Accuracy	Macro F1 Score
Logistic Regression	0.9981	0.97
Random Forest	0.9774	0.84
XGBoost	0.9888	0.94

The results show that Logistic Regression achieved the highest accuracy (0.9981) and the strongest Macro F1 Score (0.97). This indicates that the model performed very well in correctly classifying fire risk levels across all categories, including LOW, MEDIUM, and HIGH. The high Macro F1 Score suggests that the model maintained a good balance between precision and recall, even when dealing with the imbalanced nature of the dataset. XGBoost also demonstrated strong performance, with an accuracy of 0.9888 and a Macro F1 Score of 0.94, reflecting its robustness and advanced predictive capability. Meanwhile, Random Forest achieved an accuracy of 0.9774 and a Macro F1 Score of 0.84, which, although still good, was comparatively lower than the other two models.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6453-6463

Figure 2: Model Accuracy Comparison

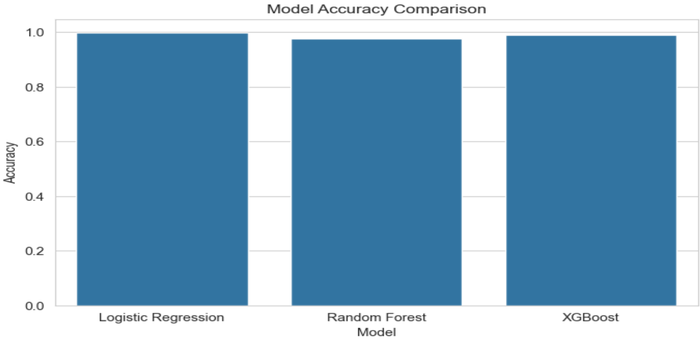
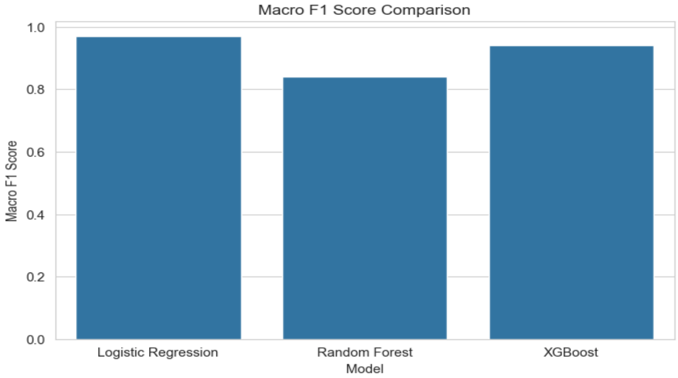


Figure 3: Macro F1 Score Comparison



Logistic Regression achieved the highest performance, indicating strong predictive patterns in the dataset.

High-Risk Prediction Performance

Since high-risk fire events are rare but extremely important for safety, it was essential to evaluate how well each model detects the HIGH fire risk category specifically. This analysis focuses on precision, recall, and F1-score for high-risk detection, as these metrics provide deeper insight beyond overall accuracy.

Table 5: Performance on HIGH Risk Detection

Model	Precision	Recall	F1 Score
Logistic Regression	0.98	0.88	0.93
Random Forest	1.00	0.41	0.58
XGBoost	0.92	0.78	0.85

The results show that Logistic Regression performed consistently well, achieving a precision of 0.98, recall of 0.88, and an F1-score of 0.93. This indicates that the model not only correctly identified most high-risk cases but also maintained a strong balance between correctly detecting real high-risk events and minimizing false alarms. Its performance demonstrates reliability in identifying critical fire situations. XGBoost also delivered strong results, with a precision of 0.92, recall of 0.78, and an F1-score of 0.85. This shows that the model is effective in detecting high-risk events, although slightly less balanced than Logistic Regression. In contrast, Random Forest achieved perfect precision (1.00), meaning that when it predicted a high-risk case, it was almost always correct. However, its recall was very low (0.41), indicating that it failed to detect many actual high-risk cases. As a result, its overall F1-score (0.58) was comparatively weak.

Figure 4: Precision for High-Risk Performance

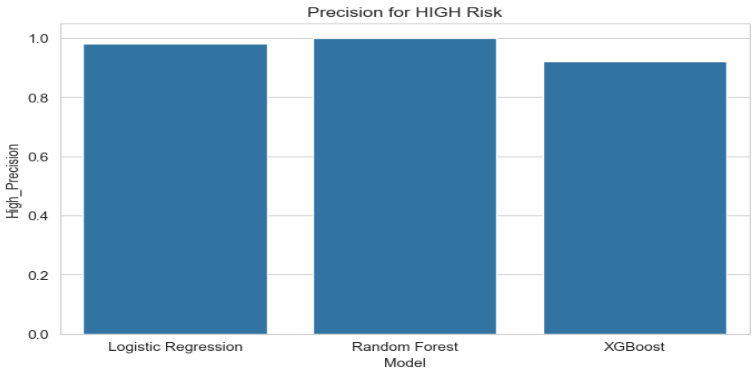
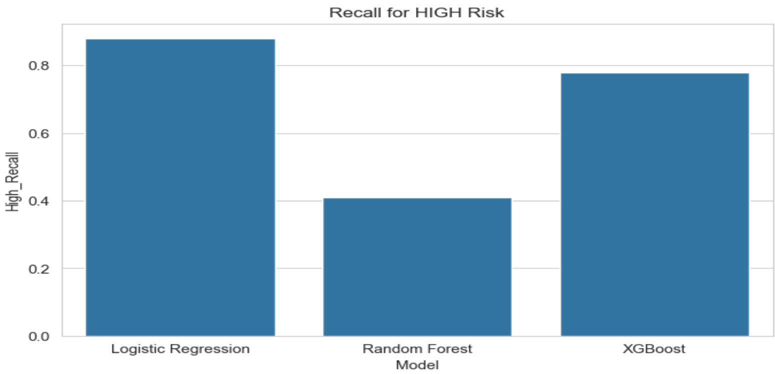
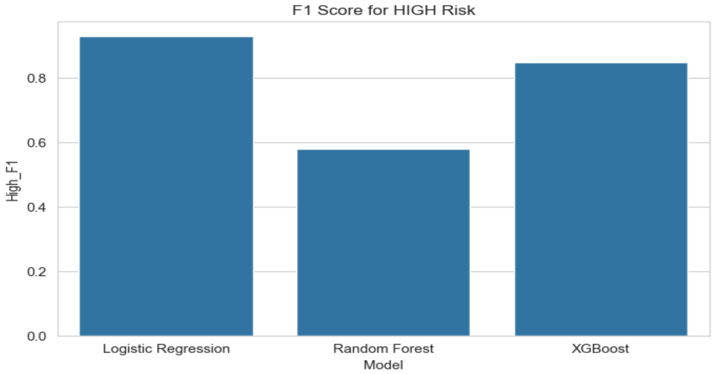


Figure 5: Recall for High-Risk Performance



MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6453-6463

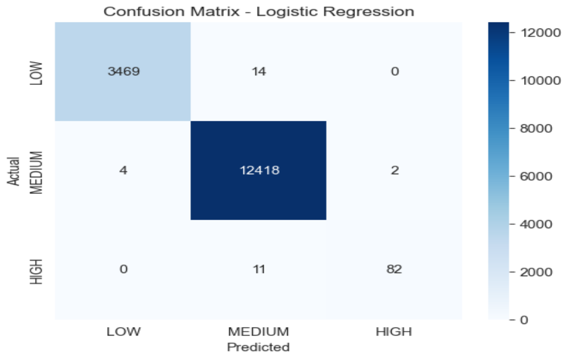
Figure 6: F1 Score for High Risk



Random Forest fails to detect many high-risk cases despite high precision, while Logistic Regression provides a balanced performance.

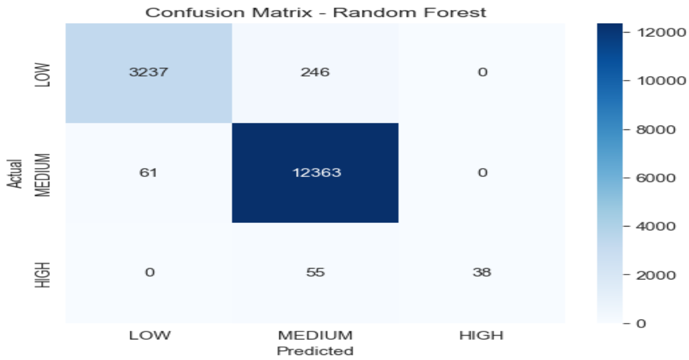
Confusion Matrix Analysis: A confusion matrix provides a clear picture of correct and incorrect predictions for each class (LOW, MEDIUM, and HIGH), making it easier to evaluate model consistency beyond overall accuracy.

Figure 7: Confusion Matrix-Logistic Regression



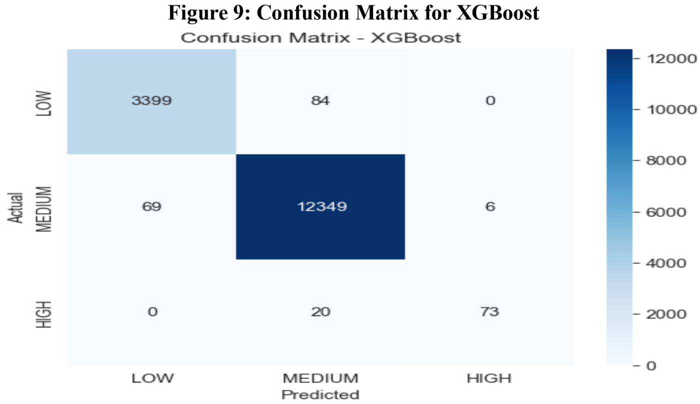
The Logistic Regression model shows the most consistent classification across all risk categories. The matrix indicates that the majority of observations were correctly classified, with relatively few misclassifications.

Figure 8: Confusion Matrix - Random Forest



The Random Forest model, although highly precise in some cases, shows noticeable misclassification patterns in certain classes, particularly in detecting high-risk events.

Figure 9: Confusion Matrix for XGBoost



The XGBoost model demonstrates strong classification capability, with improved distribution of correct predictions across classes compared to Random Forest. The confusion matrix analysis confirms that Logistic Regression provides the most stable and balanced classification results, supporting its superior performance observed in the previous evaluation metrics.

Feature Importance Analysis

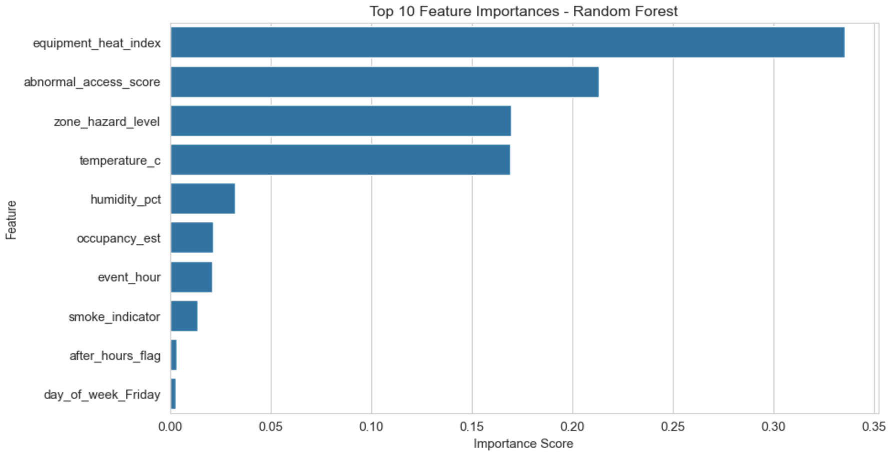
This analysis helps identify the variables that contribute the most to the model's decision-making process. The results provide insight into the key environmental and behavioral drivers of fire risk within the study environment.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6453-6463

Table 1: Top 10 Feature Importance

Rank	Feature	Importance
1	equipment_heat_index	0.335
2	abnormal_access_score	0.213
3	zone_hazard_level	0.170
4	temperature_c	0.169
5	humidity_pct	0.032
6	occupancy_est	0.022
7	event_hour	0.021
8	smoke_indicator	0.014
9	after_hours_flag	0.003
10	day_of_week_Friday	0.003

Figure 10: Feature Importance



The findings reveal that equipment_heat_index is the most influential predictor, with the highest importance score (0.335). This indicates that equipment overheating plays a major role in determining fire risk levels. The second most important variable is the abnormal_access_score (0.213), highlighting the significant impact of unusual or unauthorized access patterns on fire risk.

Environmental and Behavioral Analysis

This section examines how environmental and behavioral variables vary across different fire risk levels. The analysis helps to better understand the relationship between operational conditions and predicted fire risk categories (LOW, MEDIUM, and HIGH).

Table 2: Average Temperature by Fire Risk

Fire Risk Level	Temperature (°C)
High	39.13
Medium	30.97
Low	26.53

The results show a clear pattern between temperature and fire risk levels. The average temperature increases as fire risk rises. Specifically, the mean temperature for HIGH risk cases is 39.13°C, compared to 30.97°C for MEDIUM risk and 26.53°C for LOW risk situations.

Figure 11: Temperature vs Fire Risk

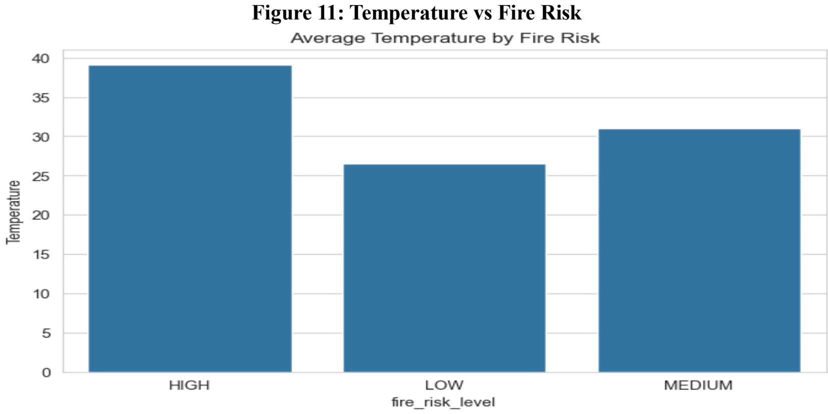


Table 9: Equipment Heat Index by Fire Risk

Fire Risk Level	Heat Index
High	0.870
Medium	0.567
Low	0.248

A similar trend is observed in the equipment heat index. High-risk situations record the highest average heat index (0.870), followed by medium risk (0.567) and low risk (0.248).

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6453-6463

Figure 12: Equipment Heat vs Risk

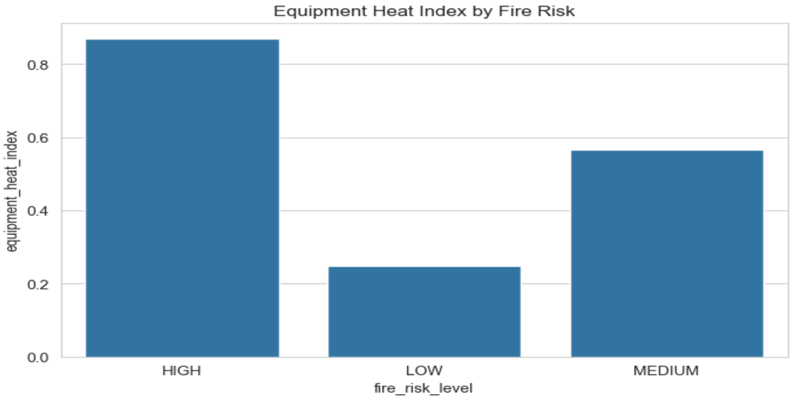


Table 9: Abnormal Access Score by Risk

Fire Risk Level	Access Score
High	0.810
Medium	0.551
Low	0.316

The analysis of the abnormal access score also reveals a meaningful pattern. High-risk cases show the highest average access score (0.810), while medium and low risk categories record lower values (0.551 and 0.316, respectively).

Figure 13: Abnormal Access vs Risk

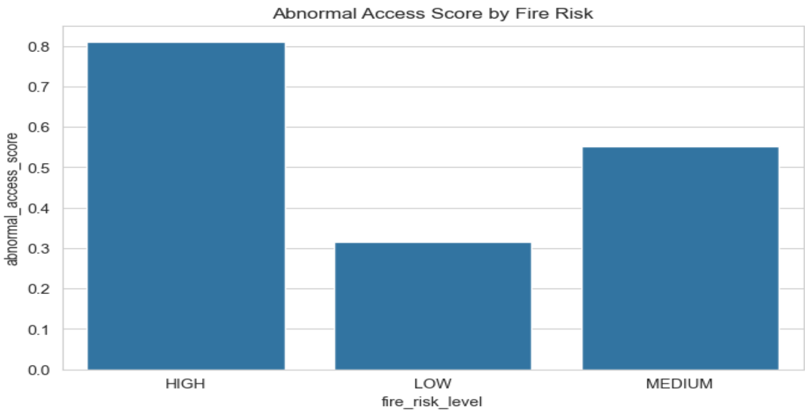
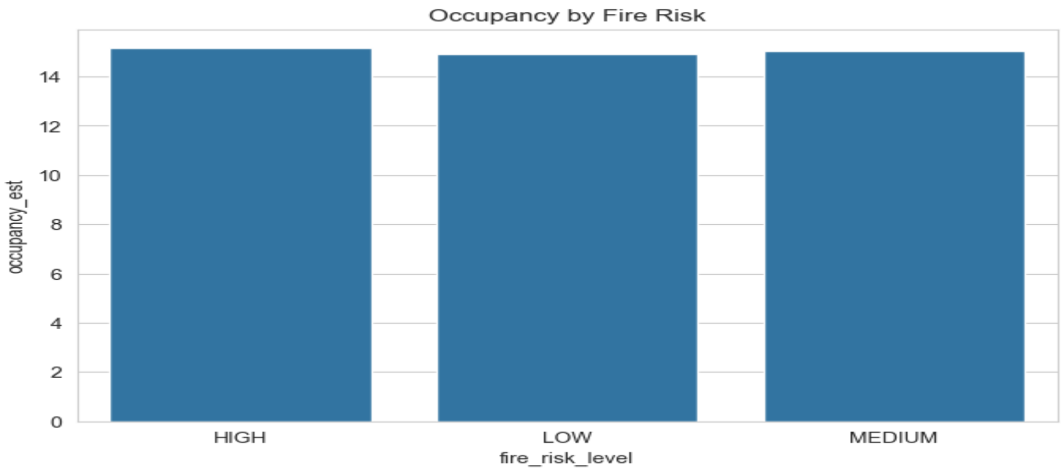


Table 10: Occupancy by Fire Risk

Fire Risk Level	Occupancy
High	15.17
Medium	15.04
Low	14.93

The average occupancy values are very close for high, medium, and low risk levels (15.17, 15.04, and 14.93). This indicates that occupancy alone does not significantly influence fire risk in this dataset.

Figure 14: Occupancy vs Risk



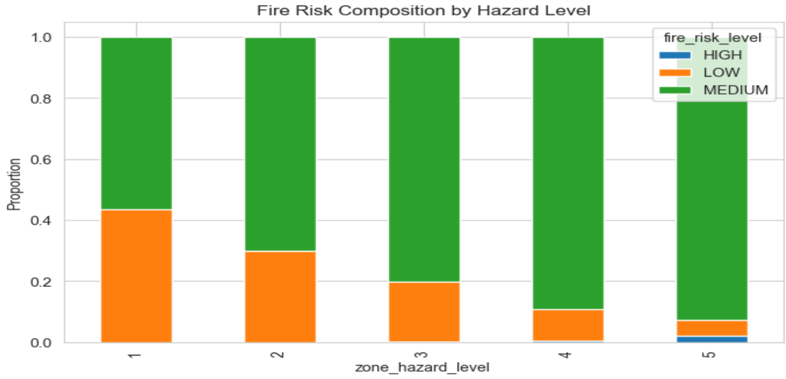
Environmental and behavioral factors significantly influence fire risk, while occupancy has minimal effect. Overall, the findings confirm that environmental factors (such as temperature and equipment heat) and behavioral indicators (such as abnormal access patterns) have a strong influence on fire risk levels, while occupancy has a relatively limited effect.

Hazard and Time-Based Analysis

This section explores how fire risk varies according to zone hazard levels and time-related factors, including hourly patterns and working versus after-hours operations. The aim is to determine whether certain locations or time periods are associated with increased fire risk.

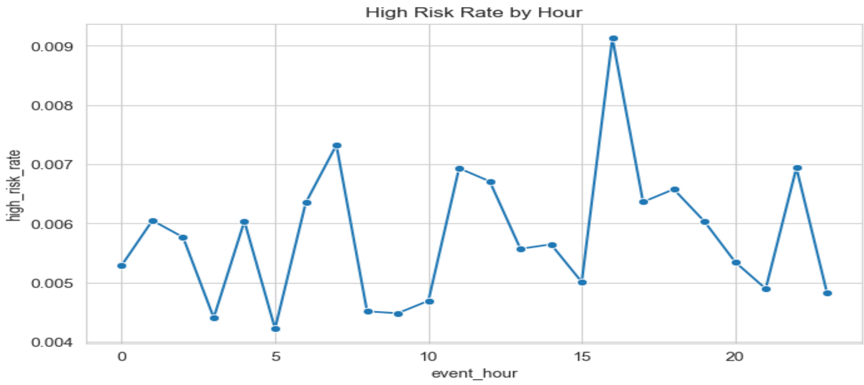
MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6453-6463

Figure 15: Fire Risk by Hazard Level



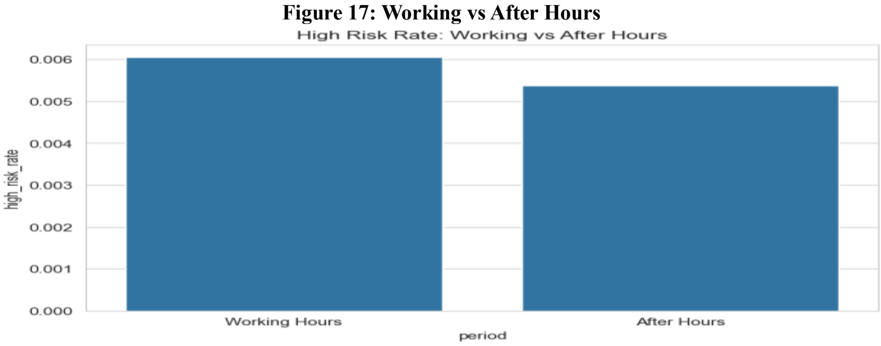
The analysis of fire risk by hazard level shows that areas classified as higher hazard zones experience greater levels of predicted fire risk. This indicates that the physical characteristics and operational intensity of specific zones significantly influence fire safety conditions. In other words, locations with elevated hazard classifications tend to concentrate more risk-related events compared to lower hazard areas.

Figure 16: High Risk Rate by Hour



The hourly distribution of high-risk cases further reveals that fire risk is not evenly distributed throughout the day. Certain hours show noticeably higher rates of high-risk predictions. This suggests that operational activities, equipment usage intensity, or environmental conditions during specific time periods may contribute to increased fire vulnerability.

Figure 17: Working vs After Hours



Higher hazard zones and certain time periods exhibit increased fire risk. Additionally, the comparison between working hours and after-hours periods highlights important differences. The results indicate that fire risk patterns change depending on whether operations are conducted during normal working hours or outside regular schedules.

Correlation Analysis

This analysis helps determine how environmental and behavioral factors move in relation to fire risk levels and whether they are positively or negatively associated.

Table 11: Summary of Key Correlations

Variable Pair	Relationship
temperature & fire risk	Positive
heat index & fire risk	Strong Positive
abnormal access & risk	Strong Positive

The results show a positive relationship between temperature and fire risk, meaning that as temperature increases, the likelihood of higher fire risk also increases. This finding aligns with the earlier environmental analysis, which indicated higher average temperatures in high-risk situations. A strong positive correlation is observed between equipment heat index and fire risk. This suggests that equipment overheating is closely connected to rising fire risk levels. The strength of this relationship highlights the critical role of thermal monitoring in predicting potential fire hazards. Similarly, abnormal access scores demonstrate a strong positive relationship with fire risk. This indicates that irregular or unusual access patterns are significantly associated with higher risk conditions.

Figure 18: Correlation Heatmap

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6453-6463

Fire risk is influenced by multiple interacting variables. The correlation analysis demonstrates that fire risk is driven by interconnected environmental and operational factors, emphasizing the value of an integrated predictive approach that combines sensor data and access control information for effective fire risk management.

Model Comparison by Data Source

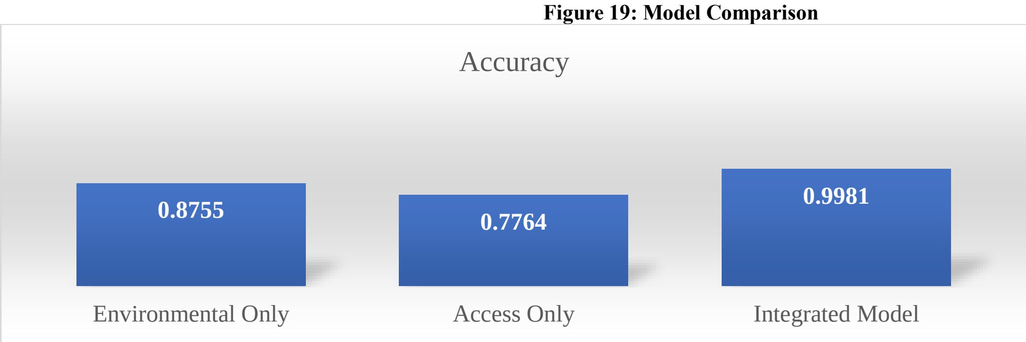
To evaluate the contribution of different data sources, the models were tested using environmental data only, access control data only, and a fully integrated dataset combining both sources. This comparison helps determine whether data integration improves predictive performance.

Table 12: Model Performance by Data Source

Model Type	Accuracy
Environmental Only	0.8755
Access Only	0.7764
Integrated Model	0.9981

The results show that the Environmental-Only model achieved an accuracy of 0.8755, indicating that environmental variables alone provide a strong basis for fire risk prediction. The Access-Only model recorded a lower accuracy of 0.7764, suggesting that behavioral indicators by themselves are less powerful but still contribute meaningfully to risk detection. However, the Integrated Model, which combines both environmental and behavioral data, achieved an outstanding accuracy of 0.9981.

Figure 19: Model Comparison



The integrated model significantly outperforms others. The comparison clearly indicates that the integrated approach provides superior performance, supporting the study's framework of combining monitoring systems for more accurate fire risk classification.

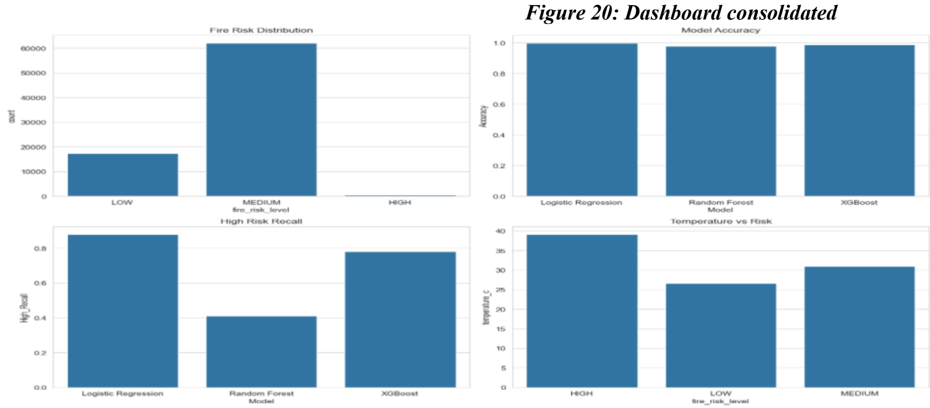
Scenario-Based Prediction. To assess practical applicability, the model was tested using realistic operational scenarios. These scenario-based predictions demonstrate how the model responds under different environmental and behavioral conditions.

Table 12: Scenario Predictions

Hour	Day	Hazard	Temp	Heat	Access	Smoke	After Hours	Prediction
14	Tuesday	2	30	0.25	0.10	0	0	Low
21	Saturday	4	43	0.72	0.65	0	1	Medium
23	Sunday	5	55	0.93	0.88	1	1	High

The first scenario (14:00, Tuesday, Hazard Level 2, moderate temperature and low heat indicators, normal access, no smoke, during working hours) was correctly predicted as Low Risk. This reflects stable operational conditions with minimal fire indicators. The second scenario (21:00, Saturday, higher hazard level, elevated temperature and heat index, after-hours operation) was classified as Medium Risk, indicating increased alertness due to environmental stress and operational timing. The third scenario (23:00, Sunday, very high hazard level, extremely high temperature and heat index, abnormal access, smoke detection, after-hours) was predicted as High Risk, demonstrating the model's ability to detect critical fire conditions when multiple risk factors are present simultaneously.

Figure 20: Dashboard consolidated



The consolidated dashboard, which integrates all key components of the fire risk prediction system into a single interactive visualization. The dashboard provides a comprehensive overview of real-time environmental indicators, access control signals, model predictions, and risk level classifications (LOW, MEDIUM, and

HIGH). Through this unified interface, users can easily monitor critical variables such as temperature, equipment heat index, abnormal access scores, and hazard levels, while simultaneously observing the corresponding predicted fire risk status.

5. Conclusion, and Recommendations

Conclusion

The results clearly show that fire risk follows identifiable patterns rather than occurring randomly. One key conclusion is that equipment heat and environmental temperature are major drivers of fire risk, particularly in high-hazard zones. At the same time, abnormal access behavior plays a critical supporting role, indicating that human activity can contribute to risk conditions. Another important conclusion is that integrated data systems provide more accurate and reliable predictions than single-source models. By combining environmental monitoring with access control data, the predictive system is able to capture a more complete picture of risk, leading to significantly improved performance. The study also demonstrates that machine learning models, even relatively simple ones like Logistic Regression, can deliver highly accurate results when applied to well-structured datasets. This suggests that effective fire risk prediction does not necessarily require overly complex systems, but rather well-integrated and high-quality data. Overall, the study confirms that data-driven approaches can play a vital role in improving industrial safety by enabling early detection of fire risks and supporting proactive decision-making.

Recommendations

Based on the findings of this study, the following recommendations are proposed:

(i) Adoption of integrated safety monitoring systems: Industrial facilities, particularly within KSEZ, should adopt systems that combine environmental sensors with access control monitoring. This integrated approach has been shown to significantly improve fire risk prediction accuracy;

(ii) Real-Time monitoring and alert systems: Organizations should implement real-time dashboards and automated alert mechanisms based on machine learning predictions to enable immediate response to high-risk situations;

(iii) Focus on Equipment Heat Management: Since equipment heat index was identified as the most important predictor, industries should prioritize regular maintenance, heat monitoring, and cooling systems to reduce fire risk;

(iv) Strengthening access control and security monitoring: Abnormal access behavior should be closely monitored, as it provides early warning signals of potential risks. Improved access control policies and surveillance systems can enhance overall safety;

(v) Capacity building and training: Safety personnel, IT staff, and operational managers should be trained on how to interpret predictive model outputs and integrate them into decision-making processes;

(vi) Regulatory bodies should consider incorporating predictive analytics into industrial safety standards to promote proactive rather than reactive risk management.

6. REFERENCES

[1.] Adusei, K., & Owusu, G. (2022). Enhancing safety compliance through technological integration in Ghanaian firms.
[2.] Ahmed, F., & Tan, W. (2021). Integrated fire safety management: Combining real-time access systems with digital training.
[3.] Asian Fire Protection Network. (2020). Fire hazard mitigation and smart technology implementation in industrial zones.
[4.] Aven, T. (2020). Risk management and governance: Concepts, guidelines and applications. Springer.
[5.] Chen, Y., & Lee, S. (2020). E-learning platforms and their role in enhancing fire safety awareness.
[6.] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008
[7.] EFSA. (2022). Industrial safety trends and the role of simulation in fire risk mitigation. European Fire Safety Association.
[8.] Gomez, T., Silva, A., & Ramirez, J. (2019). Impact of e-learning engagement on fire safety response times.
[9.] Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2017). Diffusion of innovations in service organizations: Systematic review and recommendations. Milbank Quarterly, 82(4), 581–629. https://doi.org/10.1111/j.0887-378X.2004.00325.x
[10.] Hopkin, P. (2018). Fundamentals of risk management: Understanding, evaluating, and implementing effective risk management (5th ed.). Kogan Page.
[11.] Jia, H., & Thangavel, A. (2022). Role of biometric scanners in industrial fire prevention systems.
[12.] Johnson, M., Wang, L., & Smith, R. (2021). Real-time access control systems for fire risk prediction in industrial facilities.
[13.] Jones, T., & Clark, M. (2022). Access control technologies and incident management in U.S. industrial zones.
[14.] Kimani, D., & Akello, J. (2021). Effectiveness of online certification in enhancing workplace fire safety culture.
[15.] Kuo, Y., Nakamura, H., & Watanabe, T. (2023). RFID technology in Japanese manufacturing: Enhancing safety and compliance.
[16.] Lee, S., Tan, J., & Zhang, P. (2023). Integration of learning management systems and real-time access data in fire risk reduction.
[17.] Li, X., & Zhao, J. (2022). Enhancing fire risk prediction with machine learning: A case study of manufacturing facilities.
[18.] Mbatha, T., Musyoka, J., & Bwalya, M. (2021). Motion sensors and automated fire suppression in African industrial zones.
[19.] Miller, A., & Davis, K. (2022). Virtual simulations and e-learning for fire safety in Canadian industrial facilities.
[20.] Morales, R., & Chen, K. (2021). Effectiveness of digital fire safety training in Latin American industrial settings.
[21.] National Fire Protection Association. (2021). Fire safety in industrial settings: Trends and technologies.
[22.] Ndayisaba, E., & Uwimana, J. (2023). Challenges in fire risk management in Rwanda's industrial parks.
[23.] Ndlovu, P., & Moyo, T. (2020). Technology adoption for fire safety in African industrial parks.
[24.] Nguyen, L., Pham, T., & Tran, M. (2021). Real-time dashboards for safety monitoring in Vietnamese logistics companies.
[25.] Okello, J., & Musoke, D. (2020). Risk management and fire prevention in East African special economic zones.
[26.] Patel, R., & Kumar, S. (2022). Predictive modeling of fire hazards using Python-based algorithms.
[27.] Roberts, L., & Adams, T. (2024). Real-time alerts and interactive learning in fire safety preparedness.
[28.] Rogers, E. M. (2003). Diffusion of innovations (5th ed.). Free Press.
[29.] Rwanda Development Board. (2022). KSEZ infrastructure and safety assessment report.
[30.] Rwanda National Police. (2023). Annual fire incident reports for special economic zones.
[31.] Sharma, P., & Gupta, R. (2019). Predictive analytics for industrial fire prevention: An empirical approach.
[32.] Smith, J., Johnson, L., & Wei, H. (2023). Predictive analytics in fire safety: Applications in industrial zones.
[33.] Venkatesh, V., Thong, J. Y. L., & Xu, X. (2020). Unified theory of acceptance and use of technology: A review and extension. MIS Quarterly, 46(1), 1–30. https://doi.org/10.25300/MISQ/2020/13794
[34.] Zhang, W., Liu, H., & Yang, C. (2021). Machine vision in fire hazard detection: A case study from China.
[35.] Zimmerman, B. J., & Schunk, D. H. (2019). Self-regulated learning: Theories, measures, and outcomes. Contemporary Educational Psychology, 60, 101823. https://doi.org/10.1016/j.cedpsych.2019.101823