SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Revolutionizing And Optimization Of Industrial Safety With A Deep Learning Framework For Predictive Risk Assessment And Mitigation

Dr. N. Indumathi[1*], Dr.K. Ramkumar[2], Dr. Tan Kuan Tak3, Dr.R.S. Shanmugasundaram[4]

[1*]Assistant Professor, Department of Computer Science and Engineering, Rajalakshmi Institute of Science and Technology, Chennai, Tamil Nadu, India. E-mail: induarivu27@gmail.com

[2]Professor, Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India. E-mail: ramkumar1975@gmail.com

[3]Associate Professor, Engineering Cluster, Singapore Institute of Technology, Singapore.
E-mail: tankunatak121@outlook.com

[4]Professor, Department of Computer Science and Engineering
Vinayaka Mission`s Kirupananda Variyar Engineering College, Salem, Tamil Nadu, India.
E-mail: shanmugasundaram@vmkvec.edu.in

*Corresponding author: Email: induarivu27@gmail.com

Abstract

In recent days, industry growth has been incredible in various sources by providing various technologies and frames in software and hardware, chemical, thermal, agriculture, and soon in multiple fields. Developing infrastructure, the environment, and the nature of work require safety and precautions for all employees. Analyzing the risk of various factors is important to improve safety and employee management by analyzing the risk assessment. At the existing levels, most techniques take features observation from employees to estimate risk factors and provide safety precautions. Analyzing improper features and labels creates more dimension, causing the precision false rate to degrade the prediction on risk assessment due to high er false positives. Addressing the problem and considering developing a new optimized deep learning system will improve prediction accuracy in risk assessment and improve industry safety measures. To address these issues, an Arima index feature analysis abased Deep learning using LSTM-gated RNN to predict the risk assessment and enhance the safety measures. The preprocessing is done by Min-max-Data Normalization Scaling Factor (DNSF). Then, an Automated Integrated Moving Average (ARIMA) is used to identify the variance scaling difference on feature levels. The adaptive XGboost feature section identifies the risk margins based on the stress impacts on reducing the non-relation features. The prediction is done by LSTM gated Recurrent neural network to determine the risk assessment. The proposed system improves the accuracy of prediction in risk assessment to improve safety measures and protect the employee. The accuracy attains higher precision, recall rate, sensitivity, specificity, f1 score, and reduced false rate and time complexity.

Keywords: Risk Assessment, Deep Learning, LSTM, ARIMA, Employee Safety, Feature Selection, XGBoost, Prediction Accuracy, Workplace Safety, False Positive Rate.

1. Introduction

In industrial environments, thermal conditions are essential in labor productivity and analysis. Employers can demonstrate a solid commitment to protecting employees, boosting productivity, and upholding a secure work environment by implementing effective measures to reduce thermal pollution [1]. Improper thermal conditions leading to occupational thermal stress can lower productivity and work performance in industrial safety, potentially raising the likelihood of illness and health issues. Thermal and cold stress assessments can be used to determine the cause of any deviations in the core body temperature between 36 and 37 °C. The human body can adjust its temperature to suit different work environments, allowing workers to function optimally.

Moreover, the importance of industrial safety requirements has grown in recent decades. Consequently, various engineering sectors can be evaluated using diverse strategies focused on industrial safety. Moreover, these advancements are especially relevant in complex systems with inadequate traditional methods [2]. In addition, artificial intelligence advancements can mitigate impacts on various aspects of society, including evaluation and engineering processes. Thus, robust reliability and risk analysis are essential to ensure operational safety and minimize risks from unforeseeable events. Similarly, various concepts and analyses are presented, including availability, resilience, vulnerability, accident modeling, and retrospective accident data.

Conducting a risk assessment is essential for change in the chemical processing industry to avoid serious accidents. The three main components of risk assessment involve predicting potential accident hazards, assessing the likelihood of risks, and evaluating the severity of those risks. Single-task learning analyses that measure the probability and severity of crash risks often need to pay more attention to the interconnectedness of tasks, hindering the sharing of feature information [3]. The three primary functions of risk assessment are to analyze the type, probability, and severity of accident hazards. Furthermore, they use calculated dynamic risks to plan optimal inspection and maintenance intervals to ensure safe and unsafe operations.

Since many target compounds lack safety-related properties that would restrict their employment, accurate predictive models are necessary to evaluate the compound's effects on industrial safety. The properties of organic molecules relevant to conservation are extensively used in chemical and environmental engineering [4]. Therefore, computer-aided asset forecasting is seen as an alternate approach to these issues from the perspective of process safety and risk management, and it expedites computer-aided product design for the safe operation of industrial processes.

Nevertheless, greater attention is given to developing and utilizing a risk assessment framework to evaluate these main risk categories comprehensively. While prior analyses have offered valuable perspectives on simultaneously evaluating and addressing occupational risks, there needs to be more in-depth reviews on frameworks and models covering various aspects of risk assessment [5]. One significant environmental concern is climate change, which leads to coastal land loss, rising sea levels, and ecological changes. The rapid growth in global energy demand has emphasized the need to find ways to meet the industrial sector's future thermal energy requirements.

This research outlines a novel approach to enhance prediction accuracy in risk assessment through an optimized deep learning framework, specifically utilizing Long Short-Term Memory (LSTM) gated Recurrent Neural Networks (RNNs) in conjunction with advanced preprocessing techniques and feature selection methods. By integrating these methodologies, the proposed system aims to significantly improve safety measures in industries significantly, thereby protecting employees from potential hazards.

The proposed framework leverages an innovative combination of techniques to address the limitations of existing risk assessment methodologies. At its core, it employs an LSTM-gated RNN, which is well-suited for sequential data analysis and can effectively capture temporal dependencies in risk-related data. The framework comprises several key components: data preprocessing using Min-Max Normalization Scaling Factor (DNSF), Automated Integrated Moving Average (ARIMA) for variance identification, and Adaptive XGBoost for feature selection.

Effective preprocessing of data is crucial for the success of any machine learning model. The Min-Max Normalization Scaling Factor (DNSF) is employed to scale the input features to a uniform range, typically between 0 and 1. This technique enhances the convergence speed of the LSTM model and ensures that all features contribute equally to the learning process. Normalizing the data can mitigate the influence of outliers and improve the model's overall performance. The next step involves ARIMA to identify the variance scaling across different feature levels. ARIMA, a well-established time series forecasting method, helps understand the underlying patterns and trends in the data. By analyzing historical risk data, ARIMA can reveal significant variations and assist in adjusting the model to account for these fluctuations. This step is critical in ensuring the LSTM model is trained on data that accurately reflects the real-world risk landscape.

To further refine the input data, Adaptive XGBoost is employed for feature selection. This technique identifies the most relevant features contributing to risk assessment while eliminating non-relevant or redundant features.

The model can reduce complexity and enhance interpretability by focusing on the most impactful variables. Moreover, this adaptive approach allows for identifying risk margins based on stress impacts, ensuring that the model remains sensitive to changes in operational conditions. The heart of the proposed framework is the LSTM-gated RNN, which excels in processing sequential data and capturing long-term dependencies. The LSTM model can effectively predict real-time risk assessments by training on the preprocessed and optimized feature set. The architecture of the LSTM allows it to retain information over extended periods, making it particularly effective for applications where historical data plays a crucial role in forecasting future risks. The effectiveness of the proposed system is evaluated using various performance metrics, including precision, recall rate, sensitivity, specificity, F1 score, and false positive rate. The goal is to achieve higher accuracy in risk assessment while simultaneously reducing time complexity. By improving these metrics, the framework aims to provide a reliable tool for industries to enhance their safety measures.

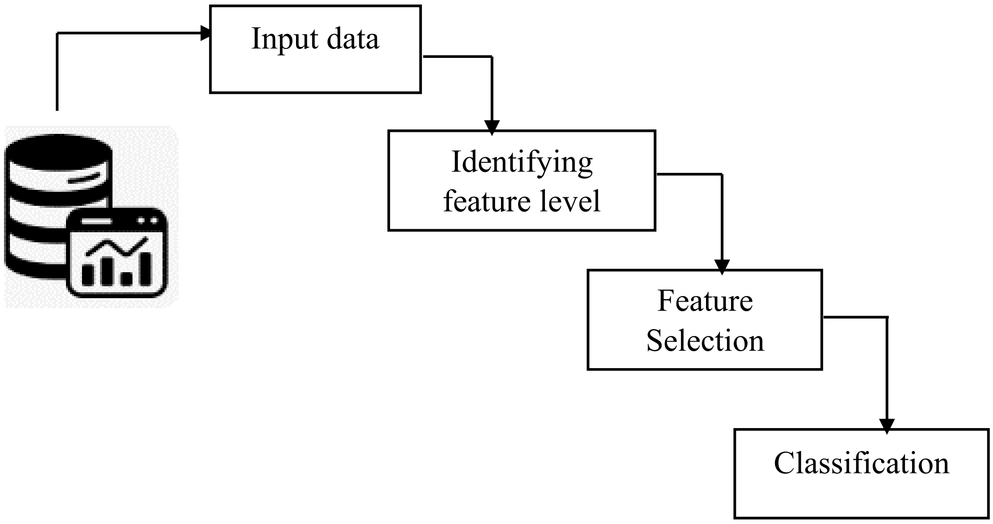

Figure 1. The Industrial Safety Architecture Diagram

The industrial safety architecture diagram illustrates essential ways to mitigate risks through data preprocessing, functional feature variance, feature selection, and classification. By systematically improving these processes, companies can maintain security against potential threats. In addition, by implementing specific security measures (as shown in Figure 1), the evaluation of these risk mitigation strategies can be significantly enhanced, ensuring a robust and resilient industrial security posture.

2. Literature Survey

Risk assessment is a critical component of industrial safety management. It involves identifying potential hazards, evaluating risks, and implementing appropriate safety measures to mitigate them. However, conventional risk assessment techniques often rely on static models that may not adequately capture the dynamic nature of industrial environments. Factors such as fluctuating operational conditions, human error, and equipment failures can lead to unforeseen risks. Therefore, there is a pressing need for a more robust and adaptive approach to risk prediction that can integrate real-time data and provide accurate assessments.

. Industrial Control Protocols (ICP) allow for identifying and mitigating critical security threats in industrial networks, thereby improving the safety and sustainability of IC Systems (ICS) in the digital era. However, these processes could be more labor-intensive, efficient, and time-consuming, posing significant challenges in ICP [8].

The mining industry incorporates an autonomous technology Artificial Intelligence (AI) model to analyze safe working environments using fully unmanned workers and mining robots. Additionally, risk assessment provides a valuable resource for analyzing the advancements of AI techniques in the mining industry [9].

Explores various aspects of strategic Safety-Critical Systems (SCS) and their reliability and security analysis by developing and evaluating rigorous procedures with experienced experts worldwide [10]. Nevertheless, achieving strict certification requirements and reliability and safety standards presents a significant barrier to the development of such systems.

The selection of safety production indicators for industrial plants is based on monitoring and managing preventive measures using production work data. However, controlling some chemical's complex and varying physical properties under different process conditions is one of the significant challenges [11].

They suggested using the Wasserstein Generative Adversarial Networks (WGAN) approach to create input data and concentrate on structure identification or analysis through manual reverse engineering [12]. On the other hand, conventional fuzzing testing may restrict generating test cases according to software specifications or the tested protocol, and there could be partial compatibility issues between software specifications and implementations.

Furthermore [13], by solving the resource allocation problem, the proposed model aims to reduce risks frequently encountered in resource-limited industrial environments. Accurate data from large-scale maintenance projects with the presented method in simulation experiments demonstrate its scalability in petrochemical plants.

Similarly, random fields are produced by a Bidirectional Long-Short-Term Memory (Bi-LSTM) algorithm to extract semantic features and classify phrases [14]. Moreover, it employs the proposed model for training and testing and evaluates its performance against two sophisticated protocol inversion tools.

The industrial safety study employs a Deep Reinforcement Learning (DRL) algorithm based on a transient control architecture to ensure adequate training in real-world settings. However, these approaches cannot directly apply to safety constraints [15]. Furthermore, the initial behavior in the environment is almost random due to multiple interactions.

Utilize business intelligence tools to derive insights from diverse data sources, enabling data visualization and Machine Learning (ML) algorithms. Additionally [16], a business intelligence framework for managing security data in an industrial setting must be developed, and the necessary procedures must be evaluated. However, reducing the frequency of events remains an ongoing and unresolved challenge in mitigating possibility severity.

Random Forest (RF) methods evaluate directional and detailed boundary divisions and discuss the corresponding underlying directional and structural systems [17]. Similarly, rapid technological advances make industrial processes more complex, interconnected, and accurate. Furthermore, the proposed model introduces new pitfalls requiring more work.

For most industrial safety procedures, characterization using Deep Neural Network (DNN) approaches assures the safety and dependability of monitoring and soft measures [18]. They also created a brand-new hybrid learning framework for supervised variational learning based on feature representation.

Non-traditional dynamic process modeling uses LSTM algorithms for soft sensor modeling of industrial process data, introducing fitted variables and complicating LSTM quality prediction models in industrial safety [19].

Visualization is carried out to enhance the prediction accuracy of a fine-tuned Safety-Mobile Net model based on experimental results [20]. The model exhibited high performance with a latency of 30ms and significant potential for utilization in construction sites.

By conducting a literature review, they identify performance indicators and collect their data by introducing a safety barrier classification system. Simultaneously, the role of safety barriers is broadened to contain the concept of stability [21]. Overcoming the challenges of enhancing design and management is imperative when analyzing safety barriers.

A Metric Learning-Based Fault Detection and Anomaly Detection (MLFDAD) approach is presented and continuously integrated into the end-to-end model to address the challenge of differentiating between unknown and known anomalies. The growing requirements for safety have been a daunting task for industrial systems [21].

The suggested approach employs the Self-Organizing Map (SOM) technique to create a database model containing various operating conditions. Moreover [22], its effectiveness is confirmed through comprehensive testing scenarios and compared against conventional methods.

Table 1: Deep Learning based on optimization of industrial safety in thermal pollution

Author	Year	Technique Used	Drawbacks
Golcarenarenji, G [23]	2022	Path Aggregation Network (PANet)	Operating a crane can threaten the surrounding environment since the operator may not be able to detect nearby workers.
Pham, H.T.T. L [24]	2021	Recurrent Neural Networks (RNN)	The fatality rates in the construction industry are, respectively, three and two times higher than those in other sectors.
Z. Huang [25]	2020	Gray-Fuzzy Evaluation Method (GFEM)	Hazards are identified when simulating the burning and spillage of dangerous chemicals.
Song, L [26]	2022	Convolutional Neural Network (CNN)	The analysis of chemical coal gangue is the most time-consuming and expensive technique.
B. Deon [27]	2022	Decision Support System (DSM)	A trade-off exists between reliability and cost, with catastrophic failure reducing power system reliability and causing economic losses.
Rama Mishra [28]	2024	Decision Tree (DT)	Safety-critical systems are becoming more complex, and modeling and evaluating these systems can be difficult and error-prone.
Mingzhu Wanga [29]	2019	Region-based CNN (R-CNN)	The vigorous activities of construction workers and equipment give rise to various safety hazards.
Y. J. Heo [30]	2019	Collision Detection Framework (CollisionNet)	Collusion detection is an essential component of collaborative robots for safety and dependability.
Zhu, X [31]	2024	LSTM	The evaluator's subjective influence can lead to misunderstandings and conflicts among the parties, posing a risk.
Alkaissy [32]	2023	Support Vector Machines (SVM)	The outcome was high compensation expenses and the scheduled halting of construction projects.

The literature analysis, as demonstrated in Table 1, highlighted the technologies and limitations of employing deep learning for enhancing the safety of thermal pollution industries, as recommended by the author.

Table 2: Industrial safety-based predictive risk assessment

Reference No	Year	Methodology	Performance Evaluation	Accuracy
33	2022	RF	Precision, True Positive Rate	94.09%
34	2022	Naïve Bayes (NB), SVM, RF	Higher accuracy and F1-score.	89%
35	2022	CNN, DNN	Precision, Recall, mean Average Precision (mAP)	90%
36	2023	Logistic Regression (LR), SVM	Area Under the Receiver Operating Curve (AUROC)	79.3%
37	2022	Graph Convolutional Network (GCN)	Accuracy	94%
38	2023	LightGBM	Mean absolute percentage error	93.2%
39	2018	ANN	Accuracy evaluation	91%
40	2020	ANN	Precision, Recall	83.6%
41	2023	Content Retrieval Method	precision, accuracy	86%
42	2023	Multi-Layer Fuzzy Logic (MFL)	real-time monitoring, improved risk assessment	73.4%
43	2021	Particle Swarm Optimization (PSO)	Mean Squared Error (MSE)	89%
44	2020	Fault Tree Analysis (FTA)	Sensitivity analysis	87.5%
45	2021	Genetic Algorithm (GA)	prediction effect, error distribution, time cost	60%

As shown in Table 2, significant methods for predictive risk assessment based on industrial safety can avoid thermal pollution accidents by improving their efficiency evaluation and accuracy.

The novel explores the significance of employing safety devices to mitigate risks and avoid workplace accidents, with an evaluation conducted through extensive discussions with seasoned workers and managers [46].

In addition, they utilized the Learning Vector Quantization Neural Network (LVQNN) algorithm and graphical analysis to evaluate the risk of hypertension in steelworkers. The classification results also enable us to measure the impact of varying sample sizes in different intervals [47].

The Dynamic Model Interpretation Guided Learning Scheme (DMI-LS), which integrates many learning techniques and uses data to assess and evaluate performance, is one method to enhance models using fragmented data. In unsafe conditions, the proposed method is crucial for preventing dangers, lowering the risk of injuries, and avoiding damage to facilities [48].

Furthermore, the author [49] mentioned that the proposed technology preserves the natural temperature conditions and habitats below the reservoirs by using a simulated optimization framework for practical testing to reduce thermal pollution. The significant impact of dams on downstream and upstream neighborhoods should be addressed.

The "51X" metric system is created by combining the possibility, severity, and dynamicity (PSD) using the offered method to forecast intrinsic, control, and disruptive risk factors [50]. Risk assessments must also be continually updated to incorporate new risk factors that address current and future challenges.

3. Methodology

The proposed methodology encompasses several key components: data preprocessing, feature selection, and predictive modeling. Each component is designed to optimize the overall performance of the risk assessment model, ensuring that its predictions yield high accuracy and reliability. Effective data preprocessing is crucial for the success of any machine learning model. In this methodology, we employ DNSF to standardize the dataset. This technique rescales the data to a fixed range, typically [0, 1], which helps mitigate feature scaling issues and ensures that all features contribute equally to the model's learning process. By normalizing the data, we can also enhance the convergence speed of the LSTM model during training, ultimately leading to improved prediction accuracy. The next step involves utilizing the ARIMA model for a thorough feature analysis. ARIMA is a powerful statistical tool well-suited for time-series data, allowing us to identify the underlying patterns and variances in the dataset. By applying ARIMA, we can assess the variance scaling at different feature levels, which helps us understand the temporal dynamics of risk factors. This analysis will inform the selection of relevant features, ensuring the model is trained on the most pertinent data points.

Following the ARIMA analysis, we implement Adaptive XGBoost for feature selection. XGBoost is an efficient and scalable implementation of gradient boosting, which excels in handling large datasets with complex feature interactions. By employing this technique, we can identify risk margins based on stress impacts while simultaneously reducing the inclusion of non-relevant features. This step is critical, as it streamlines the dataset, allowing the LSTM model to focus on the most influential variables contributing to risk assessment. The core of our proposed methodology lies in the predictive modeling phase, which utilizes an LSTM-gated recurrent neural network. LSTMs are particularly well-suited for time-series forecasting because they capture long-range dependencies and temporal patterns in sequential data. By training the LSTM model on the selected features, we aim to enhance its predictive accuracy in risk assessment.

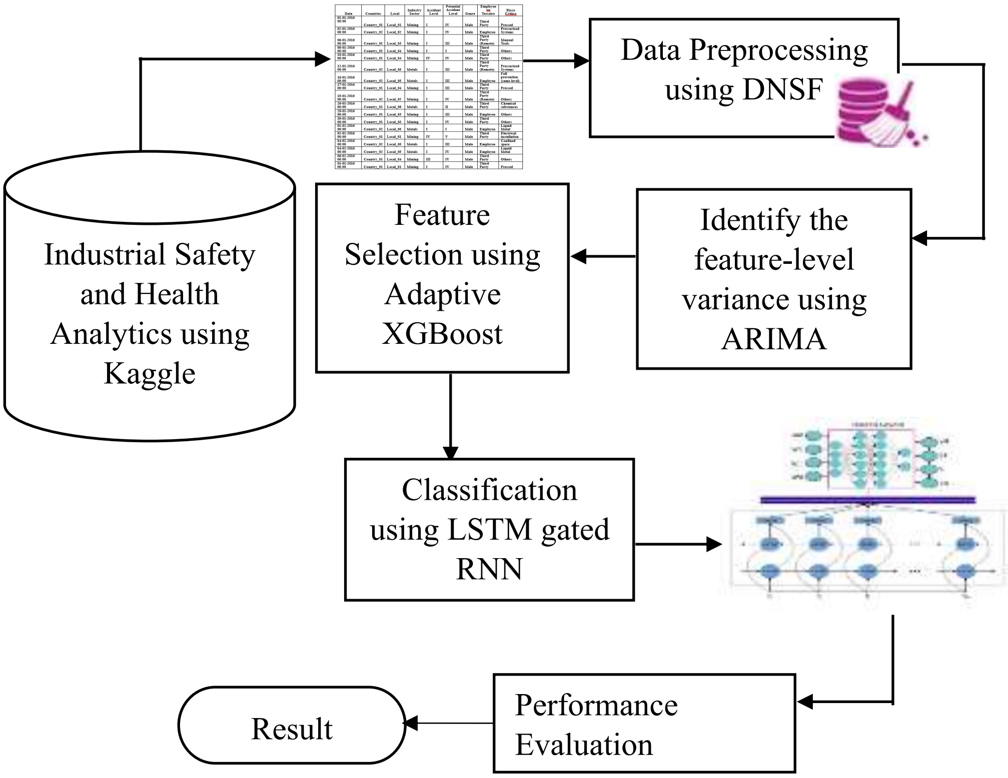

Figure 2: The proposed LSTM gated-RNN architecture diagram

In the illustration presented in Figure 2, the proposed LSTM-gated RNN can be utilized to predict risk assessment and optimize security measures. This means that risk assessments can be conducted based on analytics data related to occupational safety and health collected by Kaggle, leading to improved safety measures. Additionally, a preprocessing-based DNSF approach can be applied to normalize data. ARIMA techniques are used to estimate feature-level variance and identify measured variance. Furthermore, Adaptive XGBoost provides segmentation capabilities to identify risk margins based on stress effects, thereby enhancing the predictive accuracy of risk assessments and predicting measures to protect workers.

3.1 Dataset Collection

In this section, experimental analysis evaluation can be described by implementing an industrial safety and health analysis database obtained from Kaggle to improve industrial safety by providing risk assessment and reduction predictions. The database consists of accident records from 12 factories in three countries. Each input value in the industrial safety and health analysis dataset is analyzed to provide risk assessment predictions in an industrial environment. The total dataset is 440, test 323, and train 117, and various parameters are analyzed. In addition, https://www.kaggle.com/datasets/ihmstefanini/induscial-safety-and-health-analytics-database/data can be used to predict the accuracy and mitigation of risk assessment to improve industrial safety.

Data	Countries	Local	Industry Sector	Accident Level	Potential Accident Level	Genre	Employee ou Terceiro	Risco Critico
01-01-2016 00:00	Country_01	Local_01	Mining	I	IV	Male	Third Party	Pressed
02-01-2016 00:00	Country_02	Local_02	Mining	I	IV	Male	Employee	Pressurized Systems
06-01-2016 00:00	Country_01	Local_03	Mining	I	III	Male	Third Party (Remote)	Manual Tools
08-01-2016 00:00	Country_01	Local_04	Mining	I	I	Male	Third Party	Others
10-01-2016 00:00	Country_01	Local_04	Mining	IV	IV	Male	Third Party	Others
12-01-2016 00:00	Country_02	Local_05	Metals	I	III	Male	Third Party (Remote)	Pressurized Systems
16-01-2016 00:00	Country_02	Local_05	Metals	I	III	Male	Employee	Fall prevention (same level)
17-01-2016 00:00	Country_01	Local_04	Mining	I	III	Male	Third Party	Pressed
19-01-2016 00:00	Country_02	Local_02	Mining	I	IV	Male	Third Party (Remote)	Others
26-01-2016 00:00	Country_01	Local_06	Metals	I	II	Male	Third Party	Chemical substances
28-01-2016 00:00	Country_01	Local_03	Mining	I	III	Male	Employee	Others
30-01-2016 00:00	Country_01	Local_03	Mining	I	IV	Male	Third Party	Others
01-02-2016 00:00	Country_02	Local_05	Metals	I	I	Male	Employee	Liquid Metal
02-02-2016 00:00	Country_01	Local_01	Mining	IV	V	Male	Third Party	Electrical installation
04-02-2016 00:00	Country_02	Local_05	Metals	I	III	Male	Employee	Confined space
04-02-2016 00:00	Country_02	Local_05	Metals	I	IV	Male	Employee	Liquid Metal
06-02-2016 00:00	Country_01	Local_04	Mining	III	IV	Male	Third Party	Others
01-01-2016 00:00	Country_01	Local_01	Mining	I	IV	Male	Third Party	Pressed

Figure 3: Dataset feature collection

As illustrated in Figure 3, industrial safety can be improved by utilizing an industrial safety and health analysis database, which offers risk assessment and mitigation predictions. Moreover, the columns can present a range of dataset attributes, including incident timestamp, location, industry sector, incident level, potential incident level, type, employee or third-party involvement, critical hazard, and description.

3.1 Data Normalization Scaling Factor (DNSF)

Data preprocessing involves removing irrelevant and redundant data from a dataset and transforming raw data into a well-structured dataset for analysis. Data preprocessing involves removing irrelevant data to estimate missing data in the dataset using the DNSF approach to prevent the correlation of accidents to workers in the industry and analyzing the consistency of data representation. Additionally, normalization involves scaling features through a transformation process. In cases with a substantial gap between a feature's maximum and minimum values, normalizing the values involves scaling their magnitude toward the lower end of the scale. It is usually measured as the data in a numeric variable range between a value such as 0 and 1. Similarly, data transformation acts as a preprocessing step and parameter assessment. Moreover, the model performance can be improved by normalizing the data to the interval [0, 1]. Data preprocessing was analyzed to remove all irrelevant data and ensure consistency of data representation.

As shown in Equation 1, data preprocessing is used to calculate the value of the autocorrelation function and determine the correlation strength between features. Let's assume the v-time slot, q-time series value, $\bar{v}$ −mean values, L-autocorrelation coefficient, *and* I_L −linear correlation.

$$I_L = I(v_q, v_{q-L}) \frac{\sum_{q=L+1}^{G}(v_q-\bar{v})(v_{q-L}-\bar{v})}{\sum_{q=1}^{G}(v_q-\bar{v})^2} \quad (1)$$

Calculate the mutual information between two features using the data distribution by analyzing the digamma function of the number of points at a distance, as shown in Equation 2. Let's assume G_c −mutual information, s and t-features, Ψ −digamma function, H_{neigh} −number of nearest values, and i-number of points.

$$G_c(v,w) = \Psi(H_{neigh}) - \frac{1}{H_{neigh}} - \frac{1}{G}\sum_{c=1}^{G}[\Psi(G_s(c)) + \Psi(G_s(c))] + \Psi(g) \quad (2)$$

Variable pairwise correlation is calculated by multiplying the mean variable times the number of features by the mean correlation between each feature, as shown in Equation 3. Let's assume b-feature, $d\bar{\iota}_b$ −average correlation feature, $\bar{\iota}_{bb}$ −pairwise correlation variable.

$$M_{erits} = \frac{d\bar{\iota}_b}{\sqrt{d+(d-1)\bar{\iota}_{bb}}} \quad (3)$$

Equations 4, 5 and 6 show that attribute values are normalized by calculating the mean using standard deviation and Z-score normalization. Linear data can be transformed using Z-score normalization of minimum and maximum attribute values. Data are normalized on a scale of 0 to 1, and relationships between data values are calculated using minimum-maximum normalization. Let's assume the b-mean of the attribute, σ_b −standard deviation, s-data normalization, H_c −number of intervals, s_i −Data consumption interval, Ma_b −maximum attribute, Mi_b −minimum attribute, S_{Tra} −preserved minimum and maximum normalization data values transformation.

$$\bar{s}_b = \frac{b-\bar{b}}{\sigma_b} \quad (4)$$

$$S_{Tra} = \frac{b_c-Mi_b}{Ma_b-Mi_b}(Ma_b - Mi_b) + Mi_b \quad (5)$$

$$s_i = \sum_{L=1}^{H_c} Si_L \quad (6)$$

The correlation coefficients are averaged according to Equation 7 and then used to estimate the modulated signals in the discrete Fourier transform domain. Let's assume ζ_{0L} −coefficient correlation, s_r −data preserved, γ_{0L} −average correlation data, $u(h), v_{(k)}$ −time domain and signal transformed, H- input signal length.

$$\begin{gathered} S_r = Ma_L(sr_L), \quad L=1,2,\ldots,24 \\ \gamma_{0L} = \frac{1}{G}\sum_{q=1}^{G}\zeta_{0L}(q) \\ v_{(k)} = \sum_{h=0}^{H-1} u(h)a^{-c2\pi kh/H} \end{gathered} \quad (7)$$

The data are normalized as indicated by the gray coefficient values between the rows, and the correlation between the data is evaluated as described in Equation 8. Let's assume $\wedge_o^*(k), \wedge_c^*(k) between\ \lambda_o^*(k) - \lambda_c^*(k)$ −gray coefficient sequence, $\zeta dist$ −Distinguished coefficient.

$$\begin{gathered} \Gamma(\wedge_o^*(k), \wedge_c^*(k)) = \frac{\Delta_{mi}+\zeta dist\Delta_{ma}}{\Delta_{oc}(k)-\zeta dist\Delta_{ma}}, \zeta dist\epsilon(0,1) \\ \Delta_{oc}(k) = |\lambda_o^*(k) - \lambda_c^*(k)| \\ \Delta_{ma} = ma_{c,k}|\lambda_o^*(k) - \lambda_k^*(k)| \\ \Delta_{mi} = mi_{c,k}|\lambda_o^*(k) - \lambda_c^*(k)| \end{gathered} \quad (8)$$

Calculate the weight of each feature based on the data normalization as shown in Equation 9. Furthermore, the Euclidean distance value is estimated as shown in Equation 10. Let's assume j_n −actual load, ED-Euclidean distance, Z-weight, h-normalized data, r − Distinguished value, j_q^{24} − The actual power generation observed load time, MS −weight of natural gas index price.

$$r(u_c, u_L) = \sqrt{Z_1(u_{c1}, u_{L1})^2 + \cdots + Z_h(u_{ch}, u_{Lh})^2} \quad (9)$$

$$\|ED\| = \sqrt{(c_n - j_q^{24})^2 + (MS_h - MS_q^b)^2} \quad (10)$$

Furthermore, the correlation coefficient between the moving average and the input variable is calculated based on the correlation coefficient between the data. Calculate the proportion using the utility function to minimize missing values and irrelevant data, as shown in Equation 11. Let's assume ix_b − identify feature variance, q-time, d_r −residential load, E_i −available capacity.

$$ix_b = \frac{d_r(q)}{E_i(q)} * 100 \tag{11}$$

Data with missing or defective values are removed and transformed by a mean method to normalize the data during data preprocessing. Moreover, data are normalized between maximum and minimum values.

3.3 Automated Integrated Moving Average (ARIMA)

An ARIMA model can forecast total variance by analyzing aggregate values from industrial safety data. Additionally, it can predict total variance by applying an automatic moving average to smooth the time series data. When implemented using statistical properties, it uses an automated integrated moving average method to predict linear data patterns and provides predictive value in industrial safety risk assessment. In addition, the ARIMA model uses variance and power transformations to convert univariate time series to stationary time series and to estimate time series conditions, trends, and heterogeneity. The parameters are fully estimated after establishing the ARIMA model to minimize the cumulative error. In addition, ARIMA techniques can be developed to select constant values in the function.

Calculates the actual value of an integer and its random error over time, as illustrated in Equation 1. ARIMA models' fixed parameters are a significant obstacle in predicting labor demand in highly diverse industrial sectors. Additionally, using the ARIMA method, multiple observations are used to determine the best-fitting model for the data series. Autocorrelation model calculates the linear function of past values.

Estimating the average level of risk in an occupation through occupational safety management measures is essential by considering the cumulative impact of immediate changes over time. Time series of integrated processes can be evaluated with relatively small variations between observations. The assessed total variance ratio can be predicted at different time intervals to provide optimum protection in the industry. By analyzing the consistency of differences in these integrated processes, they can use statistical analysis of time series to predict risk estimates. The difference between two continuous values is expressed as a fixed-order integration process, as shown in equation 12. Let's assume q-time period, v_q −autual number, ε_q −random error, s and t- integer value, v_q −linear function, ε −random component, α_1 −regression coefficient.

$$v_q = ix_b \begin{cases} \alpha_0 + \alpha_1 v_{q-1} + \alpha_s v_{q-s} + \varepsilon - \beta_1 \varepsilon_{q-1} + \beta_t \varepsilon_{q-t} \\ \alpha_1 v_{q-1} + \varepsilon_q \\ v_{q-1} + \varepsilon_q \end{cases} \tag{12}$$

The perturbation to the current value is a linear combination of one or more previous perturbations and is an estimate of the moving average function. As shown in Equation 13, the moving average is calculated as a measure of the current value of the past time.

$$v_q = \varepsilon_q - \theta_1 \varepsilon_{q-1} \tag{13}$$

Non-linear transformations, such as difference and logistic functions, can convert non-stationary series into stationary ones. ARIMA employs a signal structure as a filter to distinguish the signal from past noise. Regression is performed using ordinal values to assess the impact of previous data. The calculation method of the autoregressive model is shown in Equation 14. Eliminate arbitrary information fluctuations and restore value to previously incorrect components. As illustrated in Equation 14, a moving average of the estimated forecast error is generated. Let's assume O-average moment, c and r-modelling variable, i-create different series, s and t-variable, sp-series.

$$i_O = \begin{cases} p + \sum_{L=1}^{s} \phi_L \, i_{O-L} + \epsilon_O \\ \mu_O + \sum_{c=1}^{s} \theta_L \, \epsilon_{O-L} + \epsilon_O \end{cases} \tag{14}$$

Equation 15 illustrates the development and estimate of ARIMA models that can be used with multiple series in industrial safety with different settings.

$$\hat{\imath}'_O = p + \sum_{L=1}^{s} \emptyset_L \, I'_{O-L} + \sum_{c=1}^{s} \emptyset_L \in_{O-L} + \in_O \qquad (15)$$

ARIMA models typically analyze stationary factors like seasonality, gradual and rapid changes, and periodic anomalies, calculating their total variances. Additionally, the timing of the first and second differences is identical. As shown in equation 16, predict the value of total variance in industrial safety. Let's assume $\tilde{\imath}_O$ —total variation predict value, O-time series moment, u-constant, $\in$ —noise term, t-moving average model, s-previous value of series, s, t and r − hyper parameter and order of model, α, β − weight of the coefficient vector.

$$\begin{gathered}\hat{\imath}_O = \sum_{c=1}^{t} \beta_c \in_{O-c} + \sum_{c=1}^{s} \alpha_c I_{O-c} + \in_O + U \\ \nabla^r \tilde{I}_O = \sum_{c=1}^{t} \beta_c \in_{O-c} + \sum_{c=1}^{s} \alpha_c I_{O-c} + \in_O + U \\ \tilde{I}_O = \nabla^r \tilde{I}_O + \sum_{c=0}^{r-1} \nabla^c I_{O-1}\end{gathered} \qquad (16)$$

ARIMA models that analyze multiple series can be applied in industrial safety to assess total variance across various time intervals by evaluating the actual value of an integer along with its random error over time. Furthermore, using the ARIMA method allows the identification of the best-fitting model for the data series.

3.4. Extreme Gradient Boosting (XGBoost)

This section discusses using the adaptive XGBoost algorithm to identify risk margins based on assessable stress effects. Extreme Gradient Boosting is an optimal method for gradient tree boosting, which generates hierarchical decision trees. The XGBoost algorithm is developed from a decision tree-based approach, in which graphical representations of alternative decision solutions are computed based on specific parameters. An ensemble meta-algorithm called stacking is subsequently created to combine the predictions from various decision trees employing a majority voting system. A bagging strategy can form a forest or a group of decision trees through random feature selection. Moreover, generating XGBoost models and minimizing their errors can enhance model performance. Gradient descent algorithms are also utilized to reduce the mistakes in sequence models. Additionally, the XGBoost algorithm is considered a valuable approach for enhancing gradient-boosting algorithms by addressing missing values and overfitting problems through parallel processing. The mean square error improves accuracy by analyzing the differentiable loss function's first and second-order terms. The regularization term enables control overfitting and balances the learned weights. Decision tree structure analysis can be utilized to estimate the total number of leaves in the tree and assign leaves and their weights. The minimum loss function is evaluated for leaf nodes when leaf or edge weights are node-disjoint. Furthermore, the structured decision tree is controlled by sequentially balancing the leaf weights, and the Taylor series equations and simplification can solve the desired objective function.

Calculate the function parameters using standard optimization methods in Euclidean space, as depicted in Equations 17 and 18. Applying a quadratic approximation can also enhance the loss function's linear and quadratic gradient features in fixed conditions. Let's assume j-loss function, c-instance, q-iteration, b-function. b_q —minimizing object, $m_c b_q(u_c) + \frac{1}{2} n_c b_q^2(u_c)$ —gradient statistics on the loss function.

$$j^{(q)} \simeq \tilde{I}_O \begin{cases} \sum_{c=1}^{h} l\left(v_c, \hat{v}_c{}^{q-1} + b_q(u_c) + \Omega(b_q)\right) \\ \sum_{c=1}^{h} \left[l(v_c, v_c{}^{q-1}) + m_c b_q(u_c) + \frac{1}{2} n_c b_q^2(u_c)\right] + \Omega(b_q) \end{cases} \qquad (17)$$

$$\tilde{j}^q = \sum_{c=1}^{h} \left[m_c b_q(u_c) + \frac{1}{2} n_c b_q^2(u_c)\right] + \Omega(b_q) \qquad (18)$$

Equation 19 shows the optimal weighting of the event set in the leaf-fixed structure. A scoring function can quantify the quality of the tree structure and calculate the decision tree score using various objective functions outlined in equations 20 and 21. Let's assume t-quality of tree structure, q-tree, L-leaf, z_L^* —optimal weight leaf.

$$\tilde{j}^{(q)} = \sum_{c=1}^{h} \begin{matrix} \left[m_c b_c(u_c) + \frac{1}{2} n_c b_q^2(u_c)\right] + \gamma^q + \frac{1}{2}\lambda \sum_{L=1}^{q} z_L^2 \\ \sum_{L=1}^{q} \left[\left(\sum_{c \epsilon C_c} m_c\right) z_L + \frac{1}{2}\left(\sum_{c \epsilon C_L} n_c + \lambda\right) z_L^2\right] + \gamma^q \end{matrix} \qquad (19)$$

$$z_L^* = -\frac{\sum_{c \in C_c} m_c}{\sum_{c \in C_c} n_c + \lambda} \qquad (20)$$

$$\tilde{j}^{(q)}(t) = -\frac{1}{2} \sum_{L=1}^{q} \frac{\left(\sum_{c \in C_c} m_c\right)^2}{\sum_{c \in C_c} n_c + \lambda} + \gamma^q \qquad (21)$$

Analyze the branches in a single-leaf spiral tree and evaluate the reduction in loss for the left and right node sets, as shown in Equation 22. Let's assume s-split,

$$j_p = \frac{1}{2}\left[\frac{\left(\sum_{c\in C_c} m_c\right)^2}{\sum_{c\in C_c} n_c+\lambda} + \frac{\left(\sum_{c\in C_c} m_c\right)^2}{\sum_{c\in C_c} n_c+\lambda} - \frac{\left(\sum_{c\in C_c} m_c\right)^2}{\sum_{c\in C_c} n_c+\lambda}\right] - \gamma \quad (22)$$

Calculate parallel multithreading by analyzing the node partitioning process, as shown in Equations 23 and 24. Utilize a decision tree structure depicted by leaves and their weights. Additionally, the tree's overall leaf count is measured through the tree shape's performance. Let's assume b-function, k-tree ensemble value, z-weight structure, w-tree shape, p-total number of trees, and D^p —regression size.

$$\hat{v_i} = \sum_{k=1}^{k} b_k\,(u_c), b_k \epsilon b \quad (23)$$

$$b = \{b(u) = z_w(u)^2\}(w: d^g \to p, z_w \in d^p) \quad (24)$$

As shown in Equations 25 and 26, the goal of the XGBoost model can be achieved by minimizing errors in the decision tree structure. Let's assume u-different target function, $v_c, \hat{v_i}^{d-1}$ —actual and predicted value, b_k —decision tree function, b_d-regression function, z-weight or leaves edges, λ $and \gamma$ —regularization function.

$$E^{(d)} = \sum_{c=1}^{h} x\left(\left(v_c, \hat{v_i}^{d-1}\right) + b_d(u_c)\right) + z(b_k) \quad (25)$$

$$z(b_k) = \gamma^p + \frac{1}{2}\lambda|z|^2 \quad (26)$$

Calculate loss data's first and second derivatives to identify risk margins based on stress effects, as shown in Equation 27. Let's assume s_c $b_d(u_c)$ —prevailed as the loss function and b determined the risk margin.

$$b = \sum_{c=1}^{g}\left[b_d(u_c)s_c + \frac{1}{2}b_d(u_c)^2 t_u\right] + \gamma^p + \frac{1}{2}\lambda\sum_{L=1}^{p} z_L^2 \quad (27)$$

The scoring function uses an objective function to analyze the score of the decision tree and can quantify the quality of the tree structure. Furthermore, XGBoost models can be implemented to reduce errors in the decision tree structure.

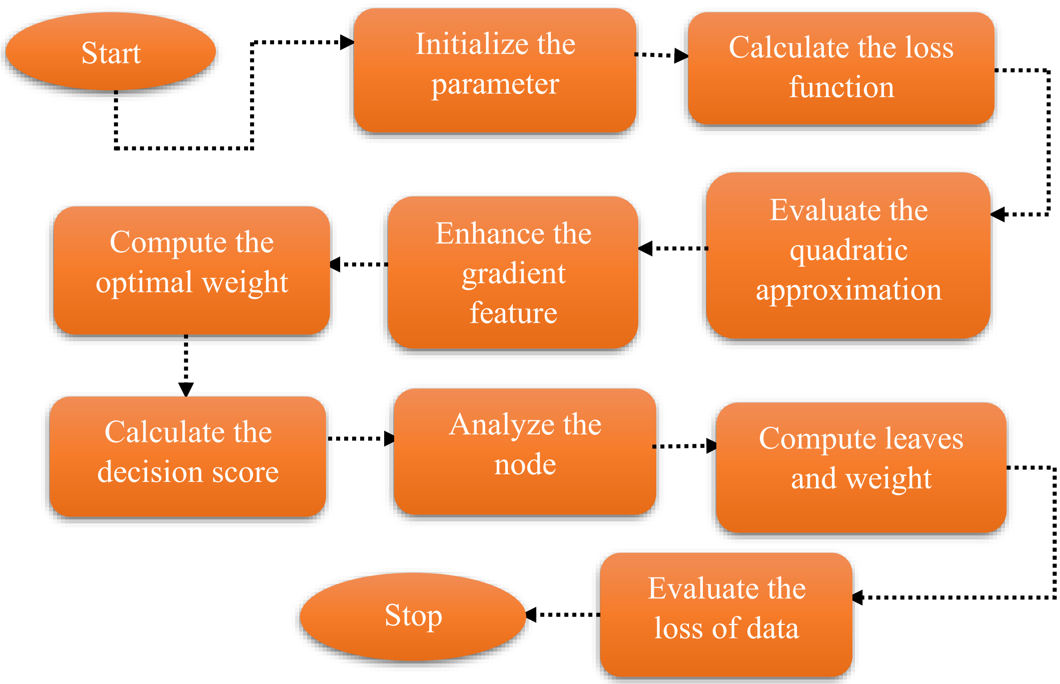

Figure 4: Extreme Gradient Boosting Flowchart Diagram

As shown in Figure 4, utilizing static optimization techniques in Euclidean space and computing function parameters through quadratic approximations can enhance the loss function's linear and quadratic gradient characteristics at a constant level. Different objective functions can be employed to assess decision tree scores and determine the optimal weights for a group of events in a leaf-level system. Furthermore, the efficiency of the tree shape gauges the total number of leaves on a tree. The XGBoost model aims to minimize errors in the structure of decision trees.

3.5 Long Short-Term Memory-Recurrent Neural Networks (LSTM-RNN)

This section uses DL-based LSTM cadet RNN to predict risk assessments to improve safety measures and protect workers by improving safety measures. RNN techniques are commonly used to analyze temporal dynamics in time series, enabling sequence processing to assess the current state based on past states and information. Changing process patterns can be identified using a health analytics database for industrial safety. RNNs utilize multilayer perceptrons to create a directed graph of neural connections for time-series data evaluated from prior layers. The connected nodes obtain input from current and past states via connections that restart according to the dependencies formed within the network. A real-valued time delay assesses the connection between the input and output layers. Input weights can be utilized to estimate the weight matrix connecting two layers. Additionally, the internal feedback weight vector from the input layer serves as the bias value for the output layer. The fundamental hidden units of the RNN are substituted with LSTM units, employing LSTM-RNN technology to analyze data related to industrial security. Health data can be preserved to mitigate issues associated with vanishing and exploding gradients when assessing long-term trends. The LSTM-RNN unit features three non-linear gates that selectively evaluate the current data stored in the memory cell and the industrial safety database.

Calculate the hyperbolic tangent of the activation function in the hidden layer and the specified linear function in the output layer, as illustrated in Equation 28. Let's assume e^J −output layer, J-layer, b^J −activation function, $h^J(q)$ −input signal, q-hyperbolic tangent.

$$e^J(q) = (b^J(h^J(q))) \tag{28}$$

Calculate the size of the feedback weight matrix for the input and output layers using the input weights from Equations 29 to 31. Evaluate the weight matrix between the input and output feedback matrices. Evaluate the input and output layer vectors with bias values at the specified time. Let's assume CZ −input weight, $JZ^{1,1}$ −internal feedback weight matrix, $JZ^{1,2}$ −external feedback weight matrix, $\underline{f}^1, \underline{f}^2, \underline{I}$ − bias values, $\mathcal{TDL}_{input}$ −tapped input delay value, $\mathcal{TDL}_{output}$ −tapped output delay value, q-time,

$$h^1(q) = CZ^{1,1}[s(q); s(q-1); \dots s]\left(q - \mathcal{TDL}_{input}\right) + JZ^{1,1}\left[e^1(q-1); \dots e^1\left(q - \mathcal{TDL}_{input}\right)\right] + JZ^{1,2}\left[e^2(q-1); \dots e^2\left(q - \mathcal{TDL}_{input}\right)\right] + JZ^{1,3}\left[e^3(q-1); \dots e^3\left(q - \mathcal{TDL}_{output}\right)\right] + \underline{f}^1 \tag{29}$$

$$h^2(q) = JZ^{2,1}e^2(q) + JZ^{2,2}\left[e^2(q-1); \dots e^2\left(q - \mathcal{TDL}_{input}\right)\right] + JZ^{2,3}\left[e^3(q-1); \dots e^3\left(q - \mathcal{TDL}_{input}\right)\right]\underline{f}^2 \tag{30}$$

$$h^3(q) = JZ^{2,2}e^2(q) + JZ^{3,3}\left[e^3(q-1); \dots e^3\left(q - \mathcal{TDL}_{input}\right)\right] + \underline{I} \tag{31}$$

The gates that control the reset of the cell state are assessed to integrate additional information into the cell state. Furthermore, by specifying the updates to the cell state, analysis can be conducted through an output gate that controls the quantity of cell state integrated into the hidden state. Calculate the position of the LSTM unit block as demonstrated in Equations 32 to 33. Let's assume b-forget gate, m-cell candidate, c-input gate, K-output gate, w-input weight, D-recurrent, f-biases, σ_m −sigmoid function, σ_I −hyperbolic tangent function.

$$\begin{gathered} b(q)=\sigma_m[z^b U(q)+D^b N(q-1)+f^b] \\ m(q)=\sigma_I[z^m U(q)+D^m N(q-1)+f^m] \\ C(q)=\sigma_C[z^m U(q)+D^C N(q-1)+f^C] \\ K(q)=\sigma_K[z^K U(q)+D^K N(q-1)+f^K] \end{gathered} \tag{32}$$

Update the cell and output state each time, as shown in Equation 6, using the point multiplication operator for the two vectors. Let's assume $I(q)$ −cell state, $n(q)$ −output state, $\odot$ − Hadamard product.

$$\begin{gathered} I(q)=b(q)\odot I(q-1)+C(q)\odot m(q) \\ n(q)=k(q)\sigma_I(I(q)) \end{gathered} \tag{33}$$

Standard LSTM-RNN methods minimize the weight and bias loss functions by updating the industrial parameters. Determine the contribution of gradient steps before the current iteration. Although computationally more efficient, LSTM-RNN performs well in industrial safety applications. As shown in Equation 34, the direction of the negative gradient of the loss is evaluated in each iteration. Let's assume the γ −gradient step in the current iteration, θ −weight, and bias industrial parameter, O-direction of negative gradient loss, and $A(\theta_O)$ −minimize loss function.

$$\theta_{O+1}\begin{cases} \alpha > 0 \\ \theta_O - \alpha\nabla A(\theta_O) \\ \theta_O - \alpha\nabla A(\theta_O) + \gamma(\theta_O - \theta_{O-1}) \end{cases} \tag{34}$$

The size of the feedback weight matrix for the input and output layers can be calculated using the input weights in a specific linear function. LSTM can update the unit each time by analyzing the state of the unit block and outputting the state. The proposed LSTM-RNN technique is effective for industrial safety applications.

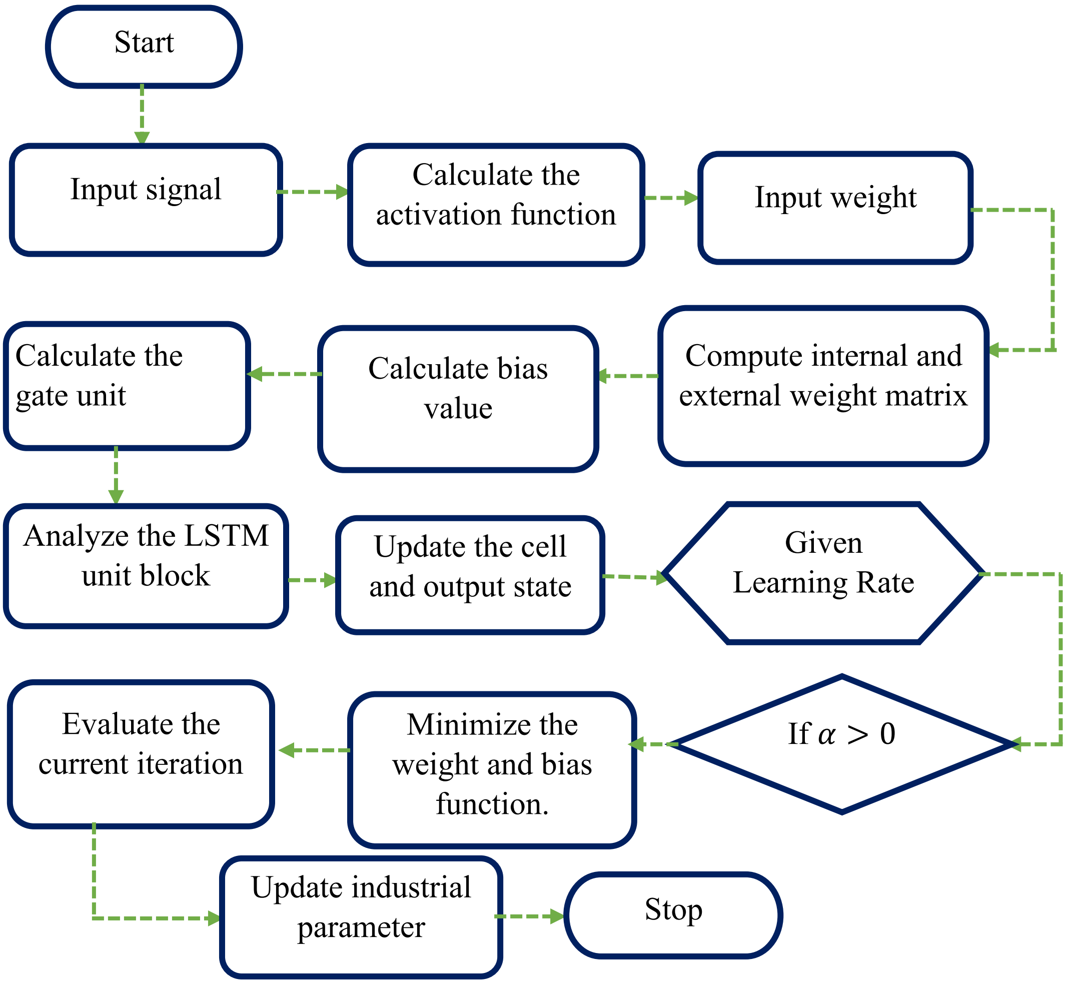

Figure 5: Flow chart diagram based on LSTM-RNN

In Figure 5, the flowchart for the proposed LSTM model shows that the networks update their units by evaluating the current state of the unit block. Furthermore, LSTM-RNN plays a significant role in controlling and improving industrial safety applications by efficiently processing continuous data.

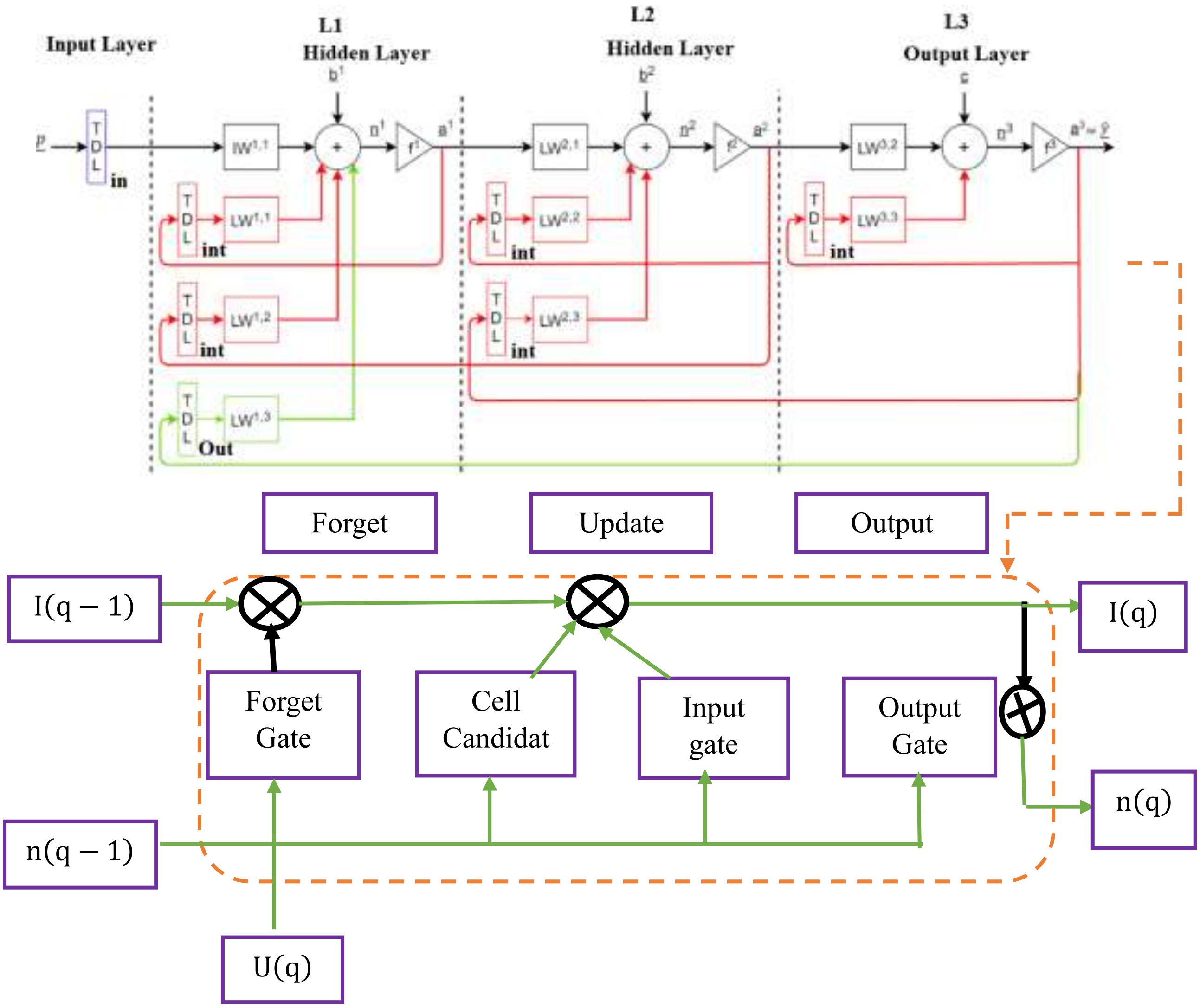

Figure 6. LSTM and RNN Architecture Diagram

The LSTM-RNN technique manages its gate's hidden state and industrial safety cells, as shown in the Figure 6 architecture diagram. A real-valued time delay connects a layer's output to its previous input. A continuous tapped delay is implemented through layer blocks, illustrating an RNN with input delays and hidden layers. Information can be added or removed at each time step to update the gate. Where t-time, n(q)-hidden state, I(q)-cell state, n(q−1) and i(q−1)-prior of time, U(q)-input state.

4. Result and Discussion

This section describes the experimental analysis results based on the performance evaluation using the proposed method for assessing industrial safety risks. Comparing the proposed technique with previous methods such as GFEM, CNN, and SVM, it achieved higher accuracy on both the test and training sets. The proposed approach can be implemented to predict risk assessment in industrial security and increase accuracy using standard metrics to evaluate performance. A performance measure can be chosen and analyzed using the model provided to predict risk assessment in industrial security. Furthermore, evaluation metrics such as precision, accuracy, Recall, F1 score, and error rate can be selected to predict risk assessment and improve industrial safety.

Table 3. Simulation parameter

Simulation	Value
Dataset Name	Industrial Safety and Health Analytics Database
Number of Epoch	440
Training	323
Testing	117
Language	Python
Tool	Jupyter

Table 3 describes the simulation parameters for analyzing industrial safety and database, epoch 440 data, training, and test 323 and 117 data using Python language and Jupyter Notebook.

Table 4. Various types of performance evaluation

Measure Name	Formula
Accuracy	$\frac{True_{Positive} + True_{Negative}}{True_{Positive} + False_{Positive} + False_{Negative} + True_{Negative}}$
Precision (PRE)	$\frac{True_{Positive}}{True_{Positive} + False_{Positive}}$
Recall (REC)	$\frac{True_{Positive}}{True_{Positive} + False_{Positive}}$
F1-score	$2\frac{PRE * REC}{PRE + REC}$
Error Rate	$\frac{\text{Number of errors}}{\text{Total number of task attempts}}$
Mean Average Precision (MAP)	$MAP = \frac{1}{N}\sum_{h=1}^{N} AP^{h}$
Mean Absolute Error (MAE)	$MAE = \frac{1}{N}\sum_{i=1}^{N} \lvert t_i - y_i \rvert$
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \lvert t_i - y_i \rvert^2}$

As demonstrated in Table 4, the confusion matrix is created by applying a DL predictive model within a framework that incorporates various performance evaluation methods to enhance industrial safety. Additionally, numerous performance metrics can be extracted from this confusion matrix. Furthermore, the proposed approach allows for an assessment of overall performance and an increase in accuracy by predicting risk levels by calculating the prediction accuracy ratio.

Table 5. Dataset description based on comparisons methods on training, testing and accuracy validated

Methods	Industrial Safety and Health Analytics Database			Industrial Dataset			Leading Industries Dataset			Parts Manufacturing - Industry Dataset			Severely Injured Workers		
	Tra	Tes	Acc	Tra	Tes	Acc	Tra	Tes	Acc	Tra	Tes	Acc	Tra	Tes	Acc
WGAN	336	104	77.1%	87653	1,235	79.6%	216	106	78.3%	353	148	77.6%	17896	3,683	84.5%
PSO	282	118	73.6%	7894	2,106	81.3%	247	75	79.4%	374	127	79.3%	17,010	4569	86.3%
DNN	333	107	79.2%	8946	1,054	83.4%	203	119	80.7%	327	174	81.6%	19876	1,703	89.2%
RNN	231	169	73.4%	8463	1,537	85.1%	201	121	82.4%	389	112	83.4%	16894	4,685	89.45%
GFEM	317	123	75.1%	7682	2,318	84.6%	213	109	84.1%	303	198	85.8%	15783	5,796	89.79%
GCN	320	120	77.8%	6487	3513	87.1%	225	97	86.3%	356	145	87.2%	18,332	3247	90.6%
MFC	251	189	84.6%	7984	2,016	83.7%	209	113	87.4%	378	123	88.3%	19,215	2364	92.4%
Bi-LSTM	334	106	87.9%	8,763	1237	87.6%	186	136	88.2%	354	147	89.1%	18,682	2897	92.89%
Pro	323	117	96.4%	7956	2,044	89.7%	215	107	89.32%	388	113	90.36%	18,393	3186	94.7%

Table 5 illustrates that various datasets can enhance industrial safety assessments through comparative methods involving training, testing, and accuracy, eventually assisting in risk prediction and mitigation. Furthermore, industrial safety can be supported by employing different datasets, including the Industrial Safety and Health Analysis Database, Industrial Dataset, Leading Industry Dataset, Parts Manufacturing - Industrial Dataset, and Severely Injured Workers Database. Additionally, by gathering a total of 440 samples from these datasets and analyzing 323 trains along with 117 data points in the test set, the proposed method was evaluated against earlier techniques such as WGAN, PSO, DNN, RNN, GFEM, GCN, MFC, and Bi-LSTM for predictive risk assessment and mitigation. This proposed approach yields a 96.4% accuracy rate, significantly enhancing industrial safety.

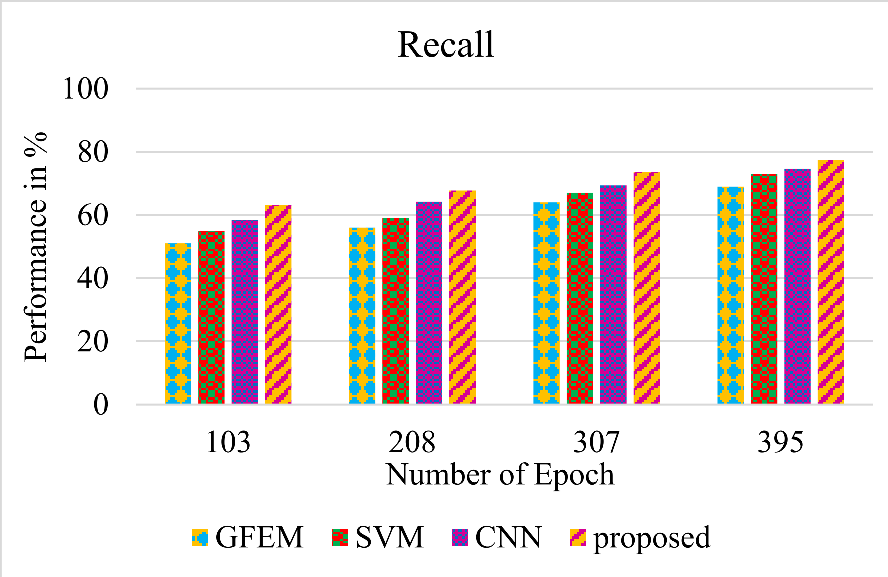

Figure 7: Analysis of Recall

As shown in Figure 7, the recall analysis can be further evaluated to provide risk assessment and mitigation predictions based on the DL framework. Additionally, the proposed method and traditional methods such as GFEM, SVM, and CNN can be used to improve industrial safety. Moreover, industrial safety and health analysis databases can be implemented to enhance occupational safety by providing risk assessment and predictions for mitigation. Furthermore, recall analysis comparing the proposed method with existing GFEM, SVM, and CNN techniques improves the accuracy of the presented method to 77.6%. It can increase the accuracy of approaches like GFEM, SVM, and CNN to 69%, 73%, and 74.6% from the previous methods.

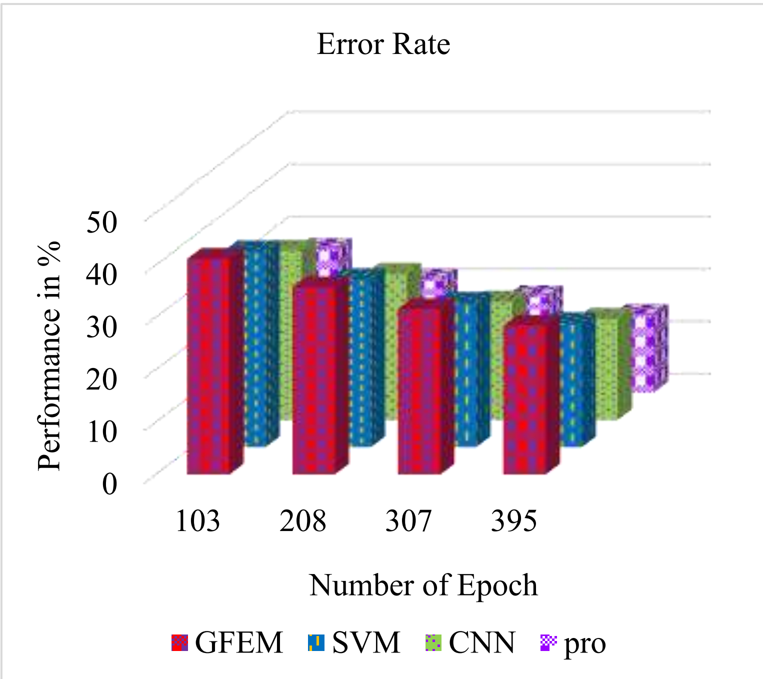

Figure 8: Analysis of error rate

Figure 8 presents an analysis of error rates, demonstrating the proposed method, which utilizes a DL architecture for risk assessment and mitigation predictions. The error rate is determined by comparing the performance of the proposed method and previous techniques, thereby assessing the accuracy of industrial security. Notably, the error rates of the proposed approach are significantly lower at 28.4%, 23.7%, and 19.3% compared to the previous GFEM, SVM, and CNN methods. Furthermore, the analysis error rate achieves a low of 14.8%, indicating that the proposed method effectively assesses industrial safety risks.

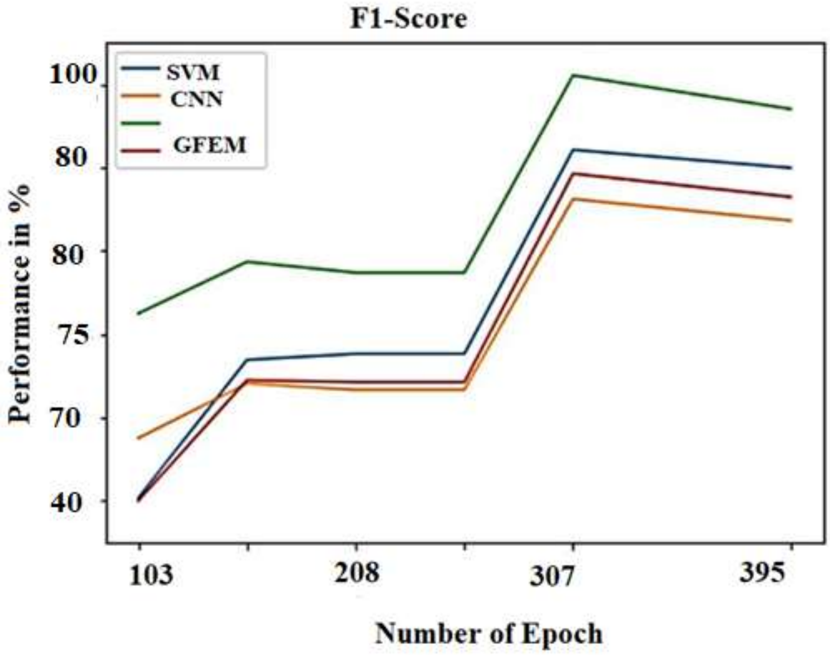

Figure 9: Analysis of F1-Score

Figure 9 illustrates the proposed approach using the safety technology DL framework for predictive risk assessment and mitigation through F1 score analysis. This score is calculated by comparing the performance of the proposed method with previous techniques to gauge industrial security accuracy. The F1 scores for the previous GFEM, SVM, and CNN methods are 73.2%, 76.8%, and 79.9%, respectively. In contrast, the assessed F1 score of the proposed method reached 83.4%, significantly enhancing the accuracy of risk assessment predictions in industrial safety.

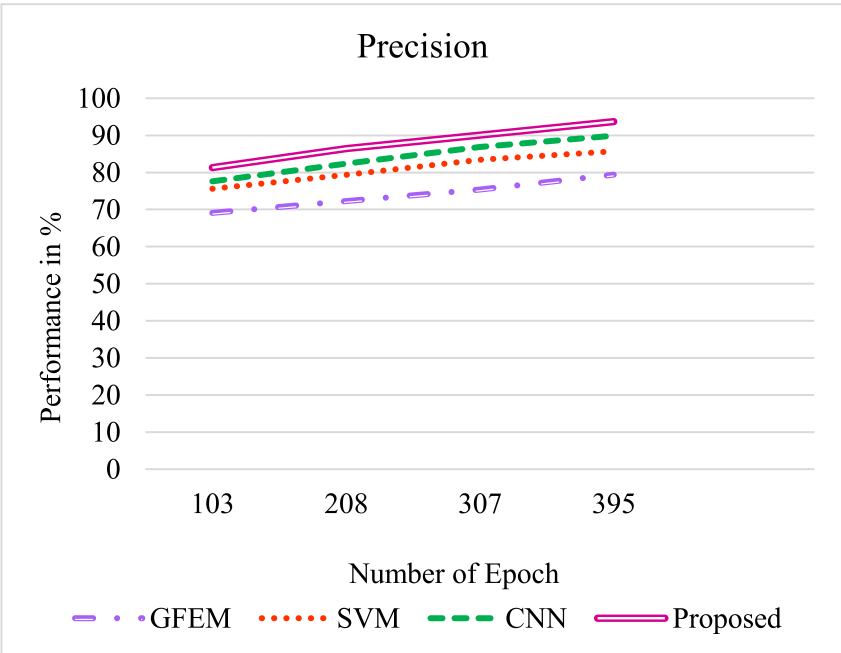

Figure 10. Analysis of Precision

As illustrated in Figure 10, precision analysis can be further evaluated to provide risk assessment and mitigation predictions based on DL approaches. Additionally, implementing an industrial safety and health analysis database can improve industrial safety by providing risk assessment and forecasts for mitigation. Furthermore, the presented technique and previous approaches, such as GFEM, SVM, and CNN, can be used to improve industrial safety. Similarly, the precision analysis comparing the offered method with the existing GFEM, SVM, and CNN techniques indicates that the accuracy of the proposed method has enhanced to 93.7%. The accuracy of methods like GFEM, SVM, and CNN can be improved to 79.4%, 85.7%, and 89.9% compared to previous approaches.

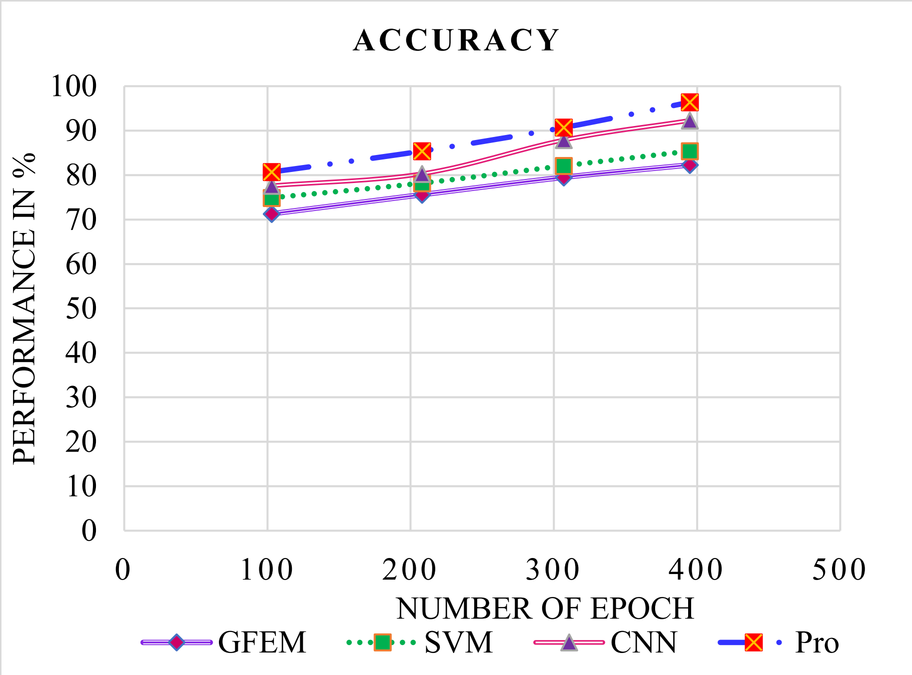

Figure 11. Analysis of accuracy

Figure 11 depicts the proposed approach of a DL framework aimed at improving security technologies to facilitate risk assessment and mitigation through industrial predictive capabilities. To demonstrate the effectiveness of this method, a comparative analysis using existing techniques such as GFEM, SVM, and CNN will be performed. Performance measurements show that the proposed framework achieves an accuracy of 82.3%, 85.4%, and 92.3%, respectively, compared to these previous methods. Moreover, the new approach achieved an impressive accuracy of 96.7%, representing a significant improvement in industrial safety, especially in risk assessment predictions.

5. Conclusion

In conclusion, developing an optimized deep-learning framework for risk assessment represents a significant advancement in enhancing safety measures within industrial environments. The proposed system integrates a comprehensive approach that combines ARIMA index feature analysis with LSTM-gated recurrent neural networks to predict and assess risks effectively. By employing DNSF for preprocessing, the model ensures that data is standardized, enhancing the accuracy of subsequent analyses. The utilization of ARIMA facilitates the identification of variance at different feature levels, allowing for a more nuanced understanding of risk factors. Furthermore, the Adaptive XGBoost feature selection process is crucial in pinpointing risk margins and eliminating non-relevant features, thereby streamlining the predictive model. The results demonstrate that the proposed deep learning approach not only improves prediction accuracy but also enhances critical performance metrics such as precision up to 96.6 %, recall 95.7 %, sensitivity 96.2 %, specificity 96.8 %, and F1 score 95.4 %, while minimizing false favorable rates and reducing time complexity. By providing a robust framework for risk assessment, this methodology contributes to developing safer industrial practices, ultimately protecting employees and fostering a safety culture. As industries continue to evolve, integrating advanced predictive analytics will be essential in addressing the complexities of risk management, ensuring that safety measures are effective and adaptable to changing environments.

References

1. Rinjea, C., Chivu, O. R., Darabont, D.-C., Feier, A. I., Borda, C., Gheorghe, M., & Nitoi, D. F. (2022). Influence of the thermal environment on occupational health and safety in automotive industry: A case study. International Journal of Environmental Research and Public Health, 19, 8572. https://doi.org/10.3390/ijerph19148572
2. Leoni, L., BahooToroody, A., Abaei, M. M., Cantini, A., BahooToroody, F., & De Carlo, F. (2024). Machine learning and deep learning for safety applications: Investigating the intellectual structure and the temporal evolution. Safety Science, 170, 106363. https://doi.org/10.1016/j.ssci.2023.106363
3. Guo, Y., Ai, X., & Luo, W. (2024). A multi-task learning risk assessment method for the chemical process industry. Process Safety and Environmental Protection, 186, 980–994. https://doi.org/10.1016/j.psep.2024.04.030
4. Wang, Z., Wen, H., Su, Y., Shen, W., Ren, J., Ma, Y., & Li, J. (2022). Insights into ensemble learning-based data-driven model for safety-related property of chemical substances. Chemical Engineering Science, 248(Part A), 117219. https://doi.org/10.1016/j.ces.2021.117219
5. Chenani, K. T., Zarei, E., Yazdi, M., Klockner, K., Alimohammadlou, M., & Kamalinia, M. (2024). A systematic review of the integration between occupational and process safety risk analysis methodologies. Journal of Loss Prevention in the Process Industries, 91, 105387. https://doi.org/10.1016/j.jlp.2024.105387
6. Kumar, L., Hasanuzzaman, M., & Rahim, N. A. (2019). Global advancement of solar thermal energy technologies for industrial process heat and its future prospects: A review. Energy Conversion and Management, 195, 885–908. https://doi.org/10.1016/j.enconman.2019.05.081
7. Zong, X., Luo, W., Ning, B., He, K., Lian, L., & Sun, Y. (2024). Diffusion fuzz: Fuzzing framework of industrial control protocols based on denoising diffusion probabilistic model. IEEE Access, 12, 67795–67808. https://doi.org/10.1109/ACCESS.2024.3399820
8. Matloob, S., Li, Y., & Khan, K. (2021). Safety measurements and risk assessment of coal mining industry using artificial intelligence and machine learning. Open Journal of Business and Management, 9, 1198–1209. https://doi.org/10.4236/ojbm.2021.93064
9. Singh, P., & Singh, L. K. (2021). Reliability and safety engineering for safety critical systems: An interview study with industry practitioners. IEEE Transactions on Reliability, 70(2), 643–653. https://doi.org/10.1109/TR.2021.3051635
10. Wang, Y. (2022). Safety production supervision of industrial enterprises based on deep learning and artificial intelligence. Computational Intelligence and Neuroscience, 2022, Article 1820082. https://doi.org/10.1155/2022/1820082
11. Li, Z., Zhao, H., Shi, J., Huang, Y., & Xiong, J. (2019). An intelligent fuzzing data generation method based on deep adversarial learning. IEEE Access, 7, 49327–49340. https://doi.org/10.1109/ACCESS.2019.2911121
12. Tewari, A., & Paiva, A. R. (2022). Modeling and mitigation of occupational safety risks in dynamic industrial environments (arXiv:2205.00894). arXiv. https://arxiv.org/abs/2205.00894
13. Ning, B., Zong, X., He, K., & Lian, L. (2023). PREIUD: An industrial control protocols reverse engineering tool based on unsupervised learning and deep neural network methods. Symmetry, 15, 706. https://doi.org/10.3390/sym15030706
14. Hu, B., & Li, J. (2021). Shifting deep reinforcement learning algorithm toward training directly in transient real-world environment: A case study in powertrain control. IEEE Transactions on Industrial Informatics, 17(12), 8198–8206. https://doi.org/10.1109/TII.2021.3063489
15. Nakhal, A. J., Patriarca, R., Di Gravio, G., Antonioni, G., & Paltrinieri, N. (2021). Investigating occupational and operational industrial safety data through business intelligence and machine learning. Journal of Loss Prevention in the Process Industries, 73, 104608. https://doi.org/10.1016/j.jlp.2021.104608
16. Wei, Z., Liu, H., Tao, X., Pan, K., Huang, R., Ji, W., & Wang, J. (2023). Insights into the application of machine learning in industrial risk assessment: A bibliometric mapping analysis. Sustainability, 15, 6965. https://doi.org/10.3390/su15086965
17. Ai, M., Xie, Y., Ding, S. X., Tang, Z., & Gui, W. (2023). Domain knowledge distillation and supervised contrastive learning for industrial process monitoring. IEEE Transactions on Industrial Electronics, 70(9), 9452–9462. https://doi.org/10.1109/TIE.2022.3206696
18. Yuan, X., Li, L., Shardt, Y. A. W., Wang, Y., & Yang, C. (2021). Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. IEEE Transactions on Industrial Electronics, 68(5), 4404–4414. https://doi.org/10.1109/TIE.2020.2984443

19. Park, J., Lee, H., & Kim, H. Y. (2022). Risk factor recognition for automatic safety management in construction sites using fast deep convolutional neural networks. Applied Sciences, 12, 694. https://doi.org/10.3390/app12020694
20. Yuan, S., Yang, M., Reniers, G., Chen, C., & Wu, J. (2022). Safety barriers in the chemical process industries: A state-of-the-art review on their classification, assessment, and management. Safety Science, 148, 105647. https://doi.org/10.1016/j.ssci.2021.105647
21. Huang, K., Wu, S., Sun, B., Yang, C., & Gui, W. (2024). Metric learning-based fault diagnosis and anomaly detection for industrial data with intraclass variance. IEEE Transactions on Neural Networks and Learning Systems, 35(1), 547–558. https://doi.org/10.1109/TNNLS.2022.3175888
22. Saucedo-Dorantes, J. J., Delgado-Prieto, M., Osornio-Rios, R. A., & Romero-Troncoso, R. D. J. (2020). Industrial data-driven monitoring based on incremental learning applied to the detection of novel faults. IEEE Transactions on Industrial Informatics, 16(9), 5985–5995. https://doi.org/10.1109/TII.2020.2973731
23. Golcarenarenji, G., Martinez-Alpiste, I., Wang, Q., et al. (2022). Machine-learning-based top-view safety monitoring of ground workforce on complex industrial sites. Neural Computing and Applications, 34, 4207–4220. https://doi.org/10.1007/s00521-021-06489-3
24. Pham, H. T. T. L., Rafieizonooz, M., Han, S., & Lee, D.-E. (2021). Current status and future directions of deep learning applications for safety management in construction. Sustainability, 13, 13579. https://doi.org/10.3390/su132413579
25. Huang, Z., et al. (2020). Safety assessment of emergency training for industrial accident scenarios based on analytic hierarchy process and gray-fuzzy comprehensive assessment. IEEE Access, 8, 144767–44777. https://doi.org/10.1109/ACCESS.2020.3013671
26. Song, L., Yu, Y., Yan, Z., Xiao, D., Sun, Y., Zhang, X., Li, X., Cheng, B., Gao, H., & Bai, D. (2022). Rapid analysis of composition of coal gangue based on deep learning and thermal infrared spectroscopy. Sustainability, 14(23), 16210. https://doi.org/10.3390/su142316210
27. Deon, B., Cotta, K. P., Silva, R. F. V., Batista, C. B., Justino, G. T., Freitas, G. C., Cordeiro, A. M., Barbosa, A. S., Loução, F. L., Simioni, T., Morais, A. M., Medeiros, I. E. A., Almeida, R. J. S., Araújo Jr., C. A. A., Soares, C., & Padoin, N. (2022). Digital twin and machine learning for decision support in thermal power plant with combustion engines. Knowledge-Based Systems, 253, 109578. https://doi.org/10.1016/j.knosys.2022.109578
28. Mishra, R., & Kumar, E. V. (2024). Industrial safety with artificial intelligence and algorithm. Journal of merging Technologies and Innovative Research, 11(3), JETIR2403689.
29. Wanga, M., Wong, P. K.-Y., Luoa, H., Kumar, S., Delhib, V. S. K., & Chenga, J. C. P. (2019). Predicting safety hazards among construction workers and equipment using computer vision and deep learning techniques. In Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC 2019).
30. Heo, Y. J., Kim, D., Lee, W., Kim, H., Park, J., & Chung, W. K. (2019). Collision detection for industrial collaborative robots: A deep learning approach. IEEE Robotics and Automation Letters, 4(2), 740–746.https://doi.org/10.1109/LRA.2019.2893400
31. Zhu, X., Wang, A., Zhang, K., & Hua, X. (2024). A deep learning method to mitigate the impact of subjective factors in risk estimation for machinery safety. Applied Sciences, 14(11), 4519. https://doi.org/10.3390/app14114519
32. Alkaissy, M., Mohammadi Golafshani, E., Arashpour, M., Hosseini, M. R., Khanmohammadi, S., Bai, Y., & Feng, H. (2023). Enhancing construction safety: Machine learning-based classification of injury types. afety Science, 162, 106102. https://doi.org/10.1016/j.ssci.2023.106102
33. Gondia, A., Ezzeldin, M., & El-Dakhakhni, W. (2022). Machine learning–based decision support framework for construction injury severity prediction and risk mitigation. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 8(3), 04022024.
34. Khairuddin, M. Z. F., Hui, P. L., Hasikin, K., Razak, N. A. A., Lai, K. W., Saudi, A. S. M., & Ibrahim, S. S. (2022). Occupational injury risk mitigation: Machine learning approach and feature optimization for smart workplace surveillance. International Journal of Environmental Research and Public Health, 19(21), 13962. https://doi.org/10.3390/ijerph192113962
35. Gallo, G., Di Rienzo, F., Garzelli, F., Ducange, P., & Vallati, C. (2022). A smart system for personal protective equipment detection in industrial environments based on deep learning at the edge. IEEE Access, 10, 110862–110878. https://doi.org/10.1109/ACCESS.2022.3215148
36. Alkaissy, M., Arashpour, M., Golafshani, E. M., Hosseini, M. R., Khanmohammadi, S., Bai, Y., & Feng, H. (2023). Enhancing construction safety: Machine learning-based classification of injury types. Safety Science, 162, 106102. https://doi.org/10.1016/j.ssci.2023.106102

37. Mostofi, F., Toğan, V., Ayözen, Y. E., & Tokdemir, O. B. (2022). Construction safety risk model with construction accident network: A graph convolutional network approach. Sustainability, 14(23), 15906.
38. Moghadasi, M., Ghadamian, H., Moghadasi, M., et al. (2023). Prediction of outlet air characteristics and thermal performance of a symmetrical solar air heater via machine learning to develop a model-based operational control scheme—An experimental study. Environmental Science and Pollution Research, 30, 27175–27190. https://doi.org/10.1007/s11356-022-24169-0
39. Guzman-Urbina, A., Aoyama, A., & Choi, E. (2018). A polynomial neural network approach for improving risk assessment and industrial safety. ICIC Express Letters, 12, 97–107.
40. Ma, H., Aviv, D., Guo, H., & Braham, W. W. (2021). Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality. Renewable and Sustainable Energy Reviews, 135, 10436. https://doi.org/10.1016/j.rser.2020.110436
41. Pedro, A., Baik, S., Jo, J., Lee, D., Hussain, R., & Park, C. (2023). A linked data and ontology-based framework for enhanced sharing of safety training materials in the construction industry. IEEE Access,11, 105410–105426. https://doi.org/10.1109/ACCESS.2023.3319090
42. Xu, R., Kim, B. W., Moe, S. J. S., Khan, A. N., Kim, K., & Kim, D. H. (2023). Predictive worker safety assessment through on-site correspondence using multi-layer fuzzy logic in outdoor construction environments. Heliyon, 9(9), e19408. https://doi.org/10.1016/j.heliyon.2023.e19408
43. Wang, Q., Pan, L., Lee, K. Y., et al. (2021). Deep-learning modeling and control optimization framework for intelligent thermal power plants: A practice on superheated steam temperature. Korean Journal of hemical Engineering, 38, 1983–2002. https://doi.org/10.1007/s11814-021-0865-6
44. Li, C., et al. (2020). Machine learning for text mining based on prediction of occupational accidents and safety risk calculation. Australian Journal of Engineering and Applied Science, 13, 11–17.
45. You, M., Li, S., Li, D., & Xu, S. (2021). Applications of artificial intelligence for coal mine gas risk assessment. Safety Science, 143, 105420. https://doi.org/10.1016/j.ssci.2021.105420
46. Mursid, F., & Herawati, S. (2023). Risk mitigation analysis and safety equipment against work accident prevention. ARRUS Journal of Social Sciences and Humanities, 3(4), 459–468. ttps://doi.org/10.35877/soshum1930
47. Wu, J.-H., et al. (2019). Risk assessment of hypertension in steel workers based on LVQ and Fisher-SVM deep excavation. IEEE Access, 7, 23109–23119. https://doi.org/10.1109/ACCESS.2019.2899625
48. He, X., & Liu, Z. (2024). Dynamic model interpretation-guided online active learning scheme for real-time safety assessment. IEEE Transactions on Cybernetics, 54(5), 2734–2745. https://doi.org/10.1109/TCYB.2023.3339242
49. Sedighkia, M., Datta, B., & Razavi, S. (2022). A simulation–optimization framework for reducing thermal pollution downstream of reservoirs. Water Quality Research Journal, 57(4), 291–303. https://doi.org/10.2166/wqrj.2022.018
50. Luo, C., Zhao, Y., & Xu, K. (2024). A risk assessment method considering risk attributes and work safety informational needs and its application. Chinese Journal of Chemical Engineering, 68, 253–262. https://doi.org/10.1016/j.cjche.2023.12.014.