Predictive Analytics Methods for Early Detection of Potential Occupational Accidents

Roid Abbadfitrah[1]**, Doni Hikmat Ramdhan**[1]
[1]Department of Occupational Safety and Health, Faculty of Public Health, Universitas Indonesia, Depok, Indonesia
Correspondence: **Roid Abbadfitrah**: Kampus UI Depok, Indonesia; roid.abbadfitrah@gmail.com

ABSTRACT
Occupational safety and health is a critical component in protecting workers and sustaining organizational operations, particularly in high-risk sectors such as energy, construction, and manufacturing. Conventional reactive approaches to occupational safety and health are often insufficient to prevent workplace accidents. Digital transformation has enabled the adoption of predictive analytics as a proactive strategy, utilizing historical data and machine learning algorithms to identify risks before incidents occur. This study aimed to systematically review the application of predictive analytics for the early detection of occupational accidents, mapping the methods, algorithms, and implementation outcomes across various industries. A scoping review was conducted following the PRISMA-ScR framework. Articles published between 2015 and 2025 were retrieved from IEEE Xplore, ScienceDirect, and PubMed. From 264 identified records, 9 relevant empirical studies were included in the final analysis. The review indicates that commonly applied algorithms, including Random Forest and Support Vector Machine, demonstrate high predictive performance in identifying accident risks, ergonomic hazards, and worker compliance, with accuracy levels reaching 90% or higher. The implementation of predictive analytics was associated with a reduction in workplace accidents of up to 25% and contributed to improved safety management through early warning systems. This systematic review underscores the significance of predictive analytics in transforming occupational safety and health management from a reactive to a proactive approach. The integration of big data, Internet of Things, and artificial intelligence supports the development of data-driven systems for effective accident prevention and provides practical recommendations for researchers and occupational safety and health practitioners to adopt predictive models within sustainable prevention strategies.
Keywords: predictive analytics; early detection; occupational accidents; machine learning

INTRODUCTION
Occupational Safety and Health plays an essential role in ensuring business continuity while protecting workers, particularly in high-risk sectors such as manufacturing, oil and gas, mining, and construction [1]. To date, Occupational Safety and Health management practices have largely been reactive, with interventions implemented only after incidents occur. Such an approach has often proven insufficient in reducing both the severity and frequency of workplace accidents. The rapid advancement of digital technologies offers an opportunity to shift this paradigm toward a preventive approach through the application of predictive analytics. This method integrates statistical techniques, predictive algorithms, and machine learning to identify patterns and estimate potential risks before incidents actually occur [2].

Recent developments in data analytics have expanded the potential for applying predictive analytics in Occupational Safety and Health. Previously, these methods have been widely utilized in the financial, transportation, and healthcare sectors to support data-driven decision-making [3]. In the context of Occupational Safety and Health, predictive analytics enables the identification of risk patterns and the generation of early warnings regarding potential incidents, including unsafe behaviors, hazardous working conditions, worker fatigue, and equipment reliability issues [4]. This can be achieved by leveraging historical data through data mining techniques, Internet of Things–based sensors, and machine learning algorithms [5]. Such an approach aligns with the proactive safety management paradigm, which emphasizes early detection and timely intervention prior to incident occurrence, thereby enabling preventive and corrective actions to reduce accident likelihood.

Despite its considerable potential, the literature indicates that the application of predictive analytics in Occupational Safety and Health remains relatively novel and insufficiently explored. Most existing studies focus primarily on mathematical or algorithmic model development, while methodological aspects, key performance indicators, and industry-level best practices have not been comprehensively examined. This situation reveals a knowledge gap that must be addressed to facilitate more effective integration of predictive analytics into Occupational Safety and Health management systems.

The utilization of Internet of Things technologies, smart sensors, big data, and artificial intelligence also enables organizations to collect large volumes of real-time data related to worker behavior, equipment conditions, environmental exposure, and ergonomic factors [6]. These data can then be processed using predictive analytics to generate actionable insights for early accident prevention. Therefore, predictive analytics can be regarded as a critical component in the development of technology-driven safety systems, consistent with the broader trend of digital transformation in global industries [7].

In the Indonesian context, research on the use of predictive analytics in Occupational Safety and Health is increasingly important, given that reported workplace accident rates remain at concerning levels. This underscores the need for more proactive and data-driven risk mitigation strategies. The implementation of predictive analytics in Occupational Safety and Health management is expected to reduce accident potential and strengthen a more adaptive and modern safety culture [8], while aligning with international standards such as ISO 45001:2018 and national regulations, including Government Regulation No. 50 of 2012 concerning the Occupational Safety and Health Management System.

Furthermore, existing literature demonstrates that studies on predictive analytics in Occupational Safety and Health are dispersed across various sectors without comprehensive synthesis, exhibiting substantial variation in approaches and outcomes [9]. This condition highlights a clear research gap that necessitates systematic mapping of analytical methods, modeling approaches, and key findings across industries. Accordingly, this study seeks to provide a structured and comprehensive synthesis of the current state of predictive analytics applications in Occupational Safety and Health, identify existing research gaps, and formulate both conceptual and practical recommendations for researchers and practitioners.

Ultimately, this study explicitly aims to systematically review and synthesize the existing evidence on the application of predictive analytics for early detection of occupational accidents, with the objective of clarifying methodological trends, identifying implementation challenges, and supporting the development of proactive, data-driven Occupational Safety and Health management systems.

METHODS
This study employed a descriptive scoping review method using the PRISMA-ScR framework, as it provides a systematic and transparent process for literature searching, selection, and analysis [10]. This approach was selected because the body of knowledge concerning predictive analytics for the early detection of potential occupational accidents is still evolving, thereby requiring an initial mapping of research trends, applied approaches, and key variables addressed in previous studies.

The search utilized combinations of the following keywords: "predictive analytics" OR "machine learning" OR "data mining" AND "occupational safety" OR "workplace accident" OR "accident prevention" AND "implementation" OR "deployment" OR "case study." The inclusion criteria comprised journal articles published between 2015 and 2025, classified as empirical studies (such as observational studies, case studies, or model validation studies), and employing historical data with predictive analytics approaches, including machine learning, data mining, deep learning, or predictive statistical models, to predict or enable early detection of occupational accidents or near-miss incidents within the context of occupational safety.

RESULTS

The Scoping Review stages applied in this study included identification, eligibility screening, and final article selection [10]. The literature search was conducted across several major databases, namely IEEE Xplore, ScienceDirect, and PubMed. Data collection employed a documentation method, and articles were accessed through the Full Open Access Library of Universitas Indonesia, as illustrated in Figure 1. In the initial PRISMA stage, 264 scientific articles were identified. Following preliminary screening based on the inclusion criterion of publication year, 141 articles remained. Subsequent title and abstract screening yielded 11 articles. Of these, one article was not an empirical study and another did not align with the objective of this study, resulting in a final inclusion of 9 scientific articles that met the inclusion criteria, as presented in Table 1.

Overall, the literature mapping indicates that the application of predictive analytics in occupational safety continues to expand across multiple sectors, particularly construction, energy, and mining [8]. The most dominant approaches involve machine learning and deep learning methods, including Convolutional Neural Networks, Support Vector Machines, Random Forest, Logistic Regression, and Extreme Gradient Boosting, which are employed to classify incidents, identify risk behaviors, or predict accident potential [11–15].

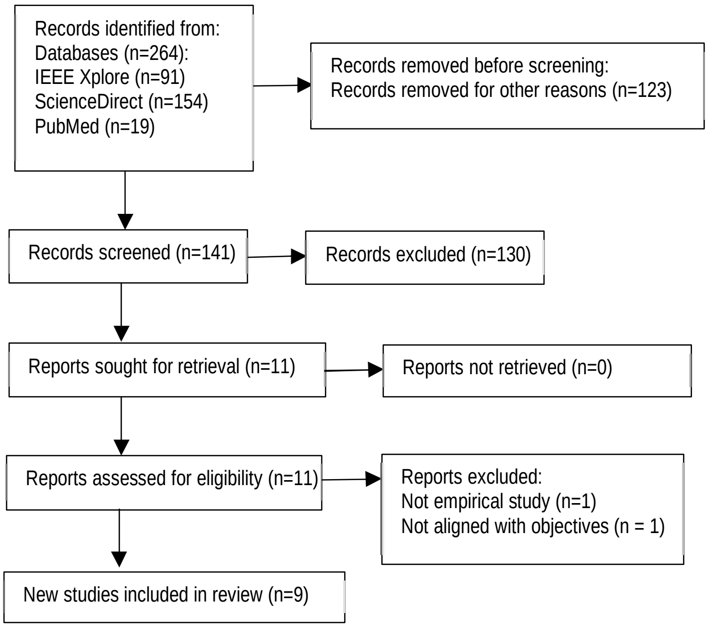

Figure 1. Scoping review stages using the PRISMA-ScR approach

Table 1. Results of the scoping review findings

No	Author and year	Article title	Predictive analytics method	Type of data used	Results
1	Laihua Fang, Lijun Wei, Guangwang Yi, Liu Jiang 2017	Research of potential safety hazard investigation and risk control system for mine enterprise	Prediction and prewarning model	Historical accident & hidden hazard (5 years), real-time sensor data, field inspection data	Improved safety management efficiency and predict and provide real-time warnings of potential accidents through IoT & data mining [11]
2	Hajar Ait Lamkademe, Ahmed Naddami, Karim Choukri. 2023	Deployment of safety predictive analytics to prevent workplace incidents and promote event reduction: a machine learning approach towards a data-driven safety system	Machine learning (ML)	Historical health, safety, and environment data from each business unit project	The implementation of safety predictive analytics offers a transformative approach to workplace safety [12]
3	Kawtar Benderouach, Khalifa Mansouri, Idriss Bennis, Ali Siadat, Abdelouahad Bellat. 2025	Classification of industrial accidents in the energy sector using machine learning models	Support Vector Machine (SVM), Logistic Regression (LR), Naïve Bayes (NB), Random Forest (RF)	Industrial accident dataset containing 4,739 workplace accidents in the energy sector (2015–2017), including event summaries, severity levels, and human and environmental factors	SVM and LR achieved 97% accuracy. SVM excelled in fatal accident recall (98%), while LR demonstrated more balanced overall performance. [13]
4	Aroa González Fuentes et al., 2022	Work-related overexertion injuries in cleaning occupations: An exploration of the factors to predict the days of absence by means of machine learning methodologies	Multivariate Adaptive Regression Splines (MARS), Support Vector Machines (SVM)	Microdata from Work Accident Statistics of the Spanish Ministry of Labor and Social Economy (2009–2019)	The SVM model demonstrated superior predictive performance compared to MARS, with lower RMSE, MAE, and MAPE values and higher R-squared coefficients [14]
5	Kanesan Muthusamy et al., 2021	Analysis of potential project work accidents: a case study of a construction project in Malaysia	Decision tree analysis using ID3, CART, and GINI index through python programming	Construction project incident reports (age, job type, working hours, nationality, duration of employment, etc.)	Accidents were most likely to occur at night due to long working hours (>10 hours), involving foreign workers (particularly Pakistani) aged ≤34 years who were classified as beginners [15]
6	Mohammad Z. Shanti et al., 2021	A novel implementation of an AI-based smart construction safety inspection protocol in the UAE	Convolutional Neural Network (CNN) using YOLOv3 algorithm	Image and video datasets from construction sites and the internet, including workers with/without safety equipment	The object detection model achieved 91.26% accuracy, 99% precision, and 90.2% recall, effectively detecting safety violations on construction sites [16]
7	Praveena G et al., 2025	Design and implementation of a smart AI-based assistant for intelligent tracking and monitoring	CNN for emotion recognition, RNN for task allocation, reinforcement learning, RTLS, predictive algorithms for maintenance	Real-time location data (GPS, RFID, Bluetooth), wearable physiological signals, facial expressions, historical worker performance, equipment maintenance data	The system reduced workplace accidents by 25%, improved worker efficiency by 20%, and predicted equipment failure with 92% accuracy. Emotion recognition achieved 87% accuracy to prevent fatigue and errors [17]
8	H. Vora et al., 2024	Musculoskeletal risk modeling using sensors and machine learning	Classification algorithm: Extreme Gradient Boosting (XGBoost)	Time-series data from wearable Electromyography (EMG) sensors, specifically surface EMG	The XGBoost model demonstrated high capability in classifying and predicting risk levels associated with manual lifting tasks based on sEMG signal data [18]
9	Mohamad Al Omari et al., 2025	Comparison of ML methods for analyzing scaffolding behavior: a case study on safety and multimodal results	Neural Networks (NN), Random Forest (RF), Support Vector Machine (SVM), Linear Regression (LR)	Dataset of 30,000 simulation scenarios derived from Finite Element Modeling (FEM)	Neural networks were the most accurate and efficient method for predicting scaffolding behavior, outperforming Random Forest, SVM, and Linear Regression [19]

Visual-based approaches such as YOLOv3 and Neural Networks are widely utilized in the highly dynamic construction sector, while classification algorithms such as Support Vector Machines, Logistic Regression, and Extreme Gradient Boosting are commonly applied to historical data analysis, ergonomic risk assessment, and accident severity prediction [16–19]. A recurring pattern shows that most studies utilize either real-time or structured historical data, ranging from CCTV recordings, workplace photographs and videos, wearable biometric signals, to Health, Safety, and Environment records and national statistical data [11–19]. The dominance of the construction sector reflects the high variability of risks; however, the energy and mining sectors also demonstrate increasing integration of Internet of Things technologies, sensors, and data mining techniques to strengthen early warning systems. Collectively, these studies confirm that predictive analytics enhances hazard identification accuracy, supports data-driven decision-making, and holds substantial potential for reducing accident rates.

DISCUSSION

Overall, the mapping of all reviewed studies demonstrates a consistent pattern in which predictive analytics acts as a catalyst for shifting safety management from a reactive approach toward a fully proactive system [11–19]. Nearly all studies employ machine learning and deep learning techniques such as Convolutional Neural Networks, YOLOv3, Support Vector Machines, Logistic Regression, Random Forest, and Extreme Gradient Boosting to predict incidents, identify hazards, or monitor risk-related behaviors. A fundamental similarity across these studies lies in the utilization of both historical and real-time data as the analytical foundation, including video recordings, photographs, physiological sensor signals, location data, and large-scale accident datasets [11–19]. Collectively, these studies consistently report high levels of predictive accuracy and demonstrate the capability of models to uncover behavioral and environmental patterns that were previously difficult to detect using conventional methods.

Despite sharing similar computational foundations, each study exhibits distinct characteristics in terms of data type, analytical objectives, and selected algorithms. Studies within the construction sector predominantly rely on computer vision techniques to detect unsafe behaviors and non-compliance with safety protocols [15,16,19], whereas research in the energy and mining industries emphasizes accident classification and the development of data-driven early warning systems based on extensive historical datasets [11,13,17]. In the field of ergonomics, the approach becomes more specialized through the use of wearable surface electromyography sensors, a method not commonly identified in other sectors [14,18].

Several important patterns emerge from these findings. First, the integration of artificial intelligence, Internet of Things technologies, and sensor systems represents a primary direction in the evolution of predictive safety frameworks [12,17,18]. Second, classification models are more dominant than regression models, reflecting a strong research focus on categorizing risk levels, such as safe versus unsafe or fatal versus non-fatal outcomes [13,16,18]. Third, the construction and energy sectors appear to be the most active adopters of predictive analytics, suggesting that high-risk industries serve as pioneers in safety digitalization efforts [12,13,16]. Moreover, nearly all studies demonstrate a significant transition from traditional leading indicators toward data-driven predictive indicators.

Three principal conclusions can be drawn. First, predictive analytics enables hazard identification prior to incident occurrence rather than merely assessing events retrospectively [12–14,16,18]. Second, various machine learning and artificial intelligence approaches show consistent and reliable performance, whether in terms of classification accuracy, prediction stability, or risk modeling capability, thereby reinforcing their role as credible decision-support tools in Health, Safety, and Environment management [11,12,15,16,19]. Third, the integration of real-time data emerges as a critical success factor, as it allows early warning systems to function automatically and continuously while facilitating faster and more precise interventions to prevent occupational accidents [11,12,16–19].

Although many predictive models demonstrate strong performance, numerous studies are conducted within limited contexts, relying on location-specific datasets or simulated data [8,9,20]. This restricts cross-industry generalizability. Furthermore, a gap remains between technical accuracy and real-world implementation readiness, particularly regarding organizational capacity, sensor deployment costs, workforce digital competence, and integration with existing Health, Safety, and Environment systems [9,20–23]. Several studies also focus primarily on algorithmic performance without adequately addressing critical issues such as data bias, overfitting risks, and ethical or privacy concerns related to biometric sensors or CCTV surveillance [9,16,18–20]. Additional critiques highlight the limited explanation of how predictive models can be directly operationalized in day-to-day decision-making processes [13,18]. This is reflected in the predominance of conceptual, evaluative, or descriptive discussions of occupational risk and safety, often without sufficient clarification of how recommendations are translated into operational practice [20,24]. Many studies do not systematically describe field-level implementation mechanisms, decision-maker roles and responsibilities, or the integration of predictive analytics into existing organizational safety management workflows [8,25]. Consequently, a substantial gap persists between the predictive capabilities demonstrated by models and their practical application in everyday work environments.

This study also presents several limitations that should be acknowledged to ensure balanced interpretation of the findings. First, although a scoping review approach was employed to ensure systematic identification and transparent source selection, this study did not conduct a formal methodological quality appraisal of each included article. As a result, the strength of evidence across publications cannot be critically assessed in the manner typical of systematic reviews applying standardized quality evaluation criteria [9]. The absence of such appraisal limits the ability to determine the internal and external validity of the overall findings. Second, the considerable heterogeneity in analytical approaches, algorithms, and data types across prior studies restricts synthesis to a descriptive and narrative level. This lack of homogeneity precludes the application of meta-analysis or other standardized quantitative synthesis methods [9,21], thereby limiting generalizability when comparing the effectiveness of different predictive modeling techniques. Third, the potential presence of publication bias must be considered. Most sources included in this review originate from peer-reviewed journals or reputable conference proceedings, which tend to emphasize positive findings, particularly models demonstrating high predictive accuracy. Studies yielding neutral or negative results may be underrepresented, potentially affecting the overall objectivity of the synthesis.

Limitations in generalizability also warrant critical attention. Many reviewed studies were conducted within highly specific industrial and geographical contexts, characterized by diverse operational conditions and data characteristics. Such variability poses challenges in directly transferring findings to Occupational Safety and Health management practices in Indonesia or other sectors with differing contextual conditions [12]. Considering these limitations, it can be concluded that while this study provides a broad initial overview of predictive analytics implementation in Occupational Safety and Health, further research employing more standardized methodological approaches and incorporating local contextual considerations is necessary. Such efforts are expected to strengthen the scientific foundation and contribute meaningfully to the development of more effective, data-driven early detection systems for occupational accidents.

CONCLUSION

The review demonstrates that predictive analytics has become a key approach in occupational accident prevention, supporting the shift from reactive to proactive, data-driven safety management through machine learning, deep learning, and sensor integration. Algorithms such as Support Vector Machines, Random Forest, Logistic Regression, Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks, and Extreme

Gradient Boosting show high predictive accuracy in identifying hazards, fatigue, unsafe behaviors, and potential equipment failures. Overall, predictive analytics enhances risk assessment, strengthens decision-making, improves resource allocation, and shows strong relevance across construction, energy, mining, and ergonomics sectors.

Organizations are encouraged to gradually integrate real-time monitoring technologies into Health, Safety, and Environment systems, supported by workforce capacity development and pilot implementation in high-risk areas. Future studies should expand cross-sectoral datasets, compare algorithms, and address issues of bias, privacy, and organizational readiness. Policymakers should also consider incorporating predictive indicators into occupational safety regulations to complement traditional lagging indicators.

Ethical consideration, competing interest and source of funding

-As this study was a systematic literature review from previously published articles, ethical approval was not required.
-There is no conflict of interest related to this study.
-Source of funding is authors.

REFERENCES

1. Simbolon RR, Harramain FP, Sonjaya MRP. Occupational safety and health (OSH) implementation as a determinant of work productivity optimization. PAJAMKEU: Pajak dan Manajemen Keuangan. 2024;1(3):17–31.
2. Pishgar M, Issa SF, Sietsema M, Pratap P, Darabi H. REDECA: A novel framework to review artificial intelligence and its applications in occupational safety and health. Int J Environ Res Public Health. 2021;18:6705.
3. Shmueli G, Koppius OR. Predictive analytics in information system research. MIS Quarterly. 2011;35(3):553–672.
4. Khairuddin MZF, Sankaranarayanan S, Hasikin K, Razak NAA, Omar R. Contextualizing injury severity from occupational accident reports using an optimized deep learning prediction model. PeerJ Computer Science; 2024;8(2):122-128.
5. Ajayi A, Oyedele LO, Delgado JMD, Akanbi L, Bilal M, Akinadé O, Olawale O. big data platform for health and safety accident prediction. World Journal of Science Technology and Sustainable Development. 2019;16(1):2–21.
6. Adikwu FE, Esiri AE, Aderamo AT, Akano OA, Erhueh OV. Leveraging digital technologies for health, safety, and environmental (HSE) management in industrial operations. International Journal of Applied Research in Social Sciences. 2024;6(11):2712–26.
7. Ankamah-Lomotey SA. The role of artificial intelligence in occupational safety and health: Investigating the potential applications and limitations of AI in improving workplace safety and health. International Journal of Research in Education Humanities and Commerce. 2025;6(5):69–82.
8. Saxena V. Predictive analytics in occupational health and safety. Popular Science, BARC Newsletter; 2024.
9. Wolniak R, Grebski W. Work safety and health – The business analytics usage in industry 4.0 conditions. Scientific Papers of Silesian University of Technology Organization and Management Series; 2023.
10. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package and shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. Campbell Systematic Reviews. 2022;18(2):18-28.
11. Fang L, Wei L, Yi G, Jiang L. Research of potential safety hazard investigation and risk control system for mine enterprise. 2nd IEEE International Conference on Computational Intelligence and Applications. 2017;1(1):112-118.
12. Hajar AL, Ahmed N, Karim C. Deployment of safety predictive analytics to prevent workplace incidents and promote event reduction: A machine learning approach towards a data-driven safety system. Proceedings - 2023 3rd International Conference on Digital Data Processing. 2023;3(1):175–179.
13. Benderouach K, Bennis I, Bellat A, Mansouri K, Siadat A. Classification of industrial accidents in the energy sector using machine learning models. 5th International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET. 2025;5(1):222-228.
14. González Fuentes A, Busto Serrano NM, Sánchez Lasheras F, Fidalgo Valverde G, Suárez Sánchez A. Work-related overexertion injuries in cleaning occupations: An exploration of the factors to predict the days of absence by means of machine learning methodologies. Appl Ergon. 2022;8(2):105-112.
15. Muthusamy K, Gunasegaran HR, Natarajan E, Renganathan K. Analysis of potential project work accidents: a case study of a construction project in malaysia. IEEE European Technology and Engineering Management Summit, E-TEMS. 2021;1(1):21–26.
16. Shanti MZ, Cho CS, Byon YJ, Yeun CY, Kim TY, Kim SK. A novel implementation of an AI-based smart construction safety inspection protocol in the UAE. IEEE Access. 2021;18(2):182-192.
17. Praveena G, Mirnalini P, Rajesh Kannan T, Ranjithkumar M, Rajaselvam M, Kavinkumar V. Design and implementation of a smart AI-based assistant for intelligent tracking and monitoring. Proceedings - 3rd International Conference on Artificial Intelligence and Machine Learning Applications. 2025;3(1):222-232.
18. Vora H, Davoudi Kakhki F, Moghadam A, Kyrarini M. Musculoskeletal risk modeling using sensors and machine learning. IEEE Computer Society; 2024;8(8):232–236.
19. Al Omari M, Eslahi M, Ammar A, El Meouche R, Ducoulombier L, Guillaumat L. Comparison of ML methods for analyzing scaffolding behavior: A case study on safety and multimodal results. Proceeding-12th International Conference on Information Technology. 2025;12(1): 451–457.
20. Luo X, Li X, Goh YM, Song X, Liu Q. Application of machine learning technology for occupational accident severity prediction in the case of construction collapse accidents. Saf Sci. 2023;12(8):163-172.
21. Saxena RR. Intelligent approaches to predictive analytics in occupational health and safety in India. arXiv Preprint. 2024 Dec 201(1):1-8.
22. Koklonis K, Sarafidis M, Vastardi M, Koutsouris D. Utilization of machine learning in supporting occupational safety and health decisions in hospital workplace. Technology & Applied Science Research. 2021;11(3):7262-7272.
23. Arthur AA, Annankra JA, Yakin Z. Examining the role of AI and machine learning in improving hazard detection and predictive analytics for accident prevention in mining operations. World Journal of Advanced Engineering Technology and Sciences. 2025;15(3):640–646.
24. Khosrowabadi N, Ghousi R. Decision support approach to occupational safety using data mining. International Journal of Industrial Engineering and Production Research. 2019;30(2):149–164.
25. Shehadeh A, Alshboul O. Enhancing occupational safety in construction: Predictive analytics using advanced ensemble machine learning algorithms. Engineering Application of Artificial Intelligence. 2025;18(12):159-168.
26. Sarkar S, Vinay S, Raj R, Maiti J, Mitra P. Application of optimized machine learning techniques for prediction of occupational accidents. Computer and Operations Research. 2019;106(1):210–224.