Review Article

New Perspectives on Occupational Safety Science: A Multidimensional Review of Machine Learning Applications in Occupational Accident Analysis

Orçun Nefes[1], Burcu Yılmaz Kaya[2]*

[1] Directorate General for EU and Foreign Affairs, Ministry of Industry and Technology, Çankaya, Ankara, Türkiye

[2]Industrial Engineering Department, Faculty of Engineering, Gazi University, Maltepe, Çankaya, Ankara, Türkiye

Article Info

Article history

Received:17/10/2025
Revised 1:10/11/2025
Accepted: 26/11/2025

Keywords:
Ergonomics
Human factors engineering
Occupational accident analysis
Data analysis
Machine learning
Artificial intelligence

ABSTRACT

Occupational accidents pose a major threat to economic enterprises in many different industries, reducing labor productivity and causing substantial economic losses, particularly in highly industrialized regions. According to ILO data, losses due to occupational accidents in Türkiye reached to 1 trillion 51 billion TL in 2023, highlighting the need for evidence-based analytical approaches to strengthen prevention strategies. Effective occupational accident analysis is essential not only for occupational safety and health (OHS) but also for sustainable development. Recent studies increasingly leverage data-driven methods, although existing findings remain fragmented and lack methodological synthesis. Alongside traditional statistical methods, modern machine learning (ML) techniques offer enhanced analytical capabilities for accident analysis. Machine learning methods, including prediction, classification, and clustering algorithms, are increasingly used to support risk identification, pattern analysis, and data-driven safety insights by learning from historical data and modeling complex accident patterns. This study systematically reviews existing research on ML applications in occupational accident analysis, examining investigated sectors, employed algorithms, data sources, and analytical trends. The review also identifies key research gaps and outlines future directions for ML-driven OHS studies.

1. Introduction

Today, occupational accidents remain one of the most pressing challenges in occupational health and safety (OHS). Occupational accidents, particularly in highly industrialized regions, cause serious losses both in terms of human life and economically. According to research conducted by the International Labour Organization (ILO), approximately three million people worldwide lose their lives each year due to occupational accidents. Furthermore, based on estimates for 2019, approximately 395 million non-fatal occupational accidents occur, and a large proportion of these results in temporary disability [1]. In Türkiye, 46.4% of occupational accidents occur in the mining, construction and metal industries and 41.1% of related deaths are caused by occupational accidents in these industries [2]

Occupational accidents lead to permanent loss of labor and loss of life, as well as economic damage and lost workdays [3, 4]. It is estimated that occupational accidents caused by insufficient and inappropriate OHS measures cost the global economy 3.94% of Gross Domestic Product annually [5]. This shows that losses due to occupational accidents in Türkiye in 2024 amount to 1 trillion 710 billion liras. In developing countries such as Türkiye, the prevention and management of occupational accidents have become one of the fundamental elements of sustainable economic development.

A wide range of methodological approaches have been employed in occupational accident research, including surveys, statistical modeling, data mining, and multi-criteria decision-making techniques. These studies span both sector-specific investigations and cross-industry analyses, addressing topics such as accident causation, classification, and predictive modeling. However, traditional statistical methods may face limitations in capturing complex patterns and producing robust predictive and classification performance, while also requiring substantial time and manual effort [6].

In the OHS literature, machine learning is widely applied to detect accident patterns, identify risk factors, and support data-driven decision-making through predictive, classification, and clustering models [7, 8, 9]. Machine learning (ML) is the ability of a computer to learn the characteristics of an event through experience learned from past events and to make decisions or produce solutions for similar events that have never been encountered before [7, 8]. The goal of ML technology is to understand data structures and transform this data into predictions that can be used by humans in various applications, making it suitable for specific models [9]. ML methods are increasingly used in occupational safety research to identify risk factors, detect accident patterns, and support data-driven decision-making. However, existing findings remain fragmented and lack systematic synthesis, particularly in the context of Türkiye.

***Corresponding author:** Burcu Yılmaz Kaya
*E-mail address:burcuyilmaz@gazi.edu.tr https://doi.org/10.56158/jpte.2025.142.4.02

This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

This study provides a systematic literature review that synthesizes existing research on machine learning applications in occupational accident analysis, with a focus on algorithms, data characteristics, industrial domains, and research trends. The review evaluates how literature positions the strengths and limitations of machine learning methods in comparison to traditional approaches in occupational accident research. Four research questions (RQs) were developed to perform a multidimensional analysis of the potential improvements that could be achieved by using ML algorithms instead of traditional methods. Concisely, while ML has achieved widespread success in traditional analytical fields, its potential in OHS domains remains underexplored. Conducting targeted research and developing application-specific frameworks for ML use in safety engineering could substantially enrich the literature. Such contributions would not only fill a critical academic void but also enable practitioners to leverage ML techniques more effectively for improving human well-being, system safety, and organizational efficiency.

The research findings were categorized based on various parameters such as the methods used, the industries researched, the types of problems, and the advantages and disadvantages compared to traditional methods. The aim was to provide a valuable resource that would contribute to the development of OHS strategies specific to Türkiye and also for international scientific researchers and field practitioners and provide strategic guidance to decision-makers.

The remainder of this study was organized as, Section 2 provides information on the concepts of occupational accidents and ML. Section 3 presents the method and application performed in the study, i.e. ML employment in occupational accident analysis, the methods used, and its potential in occupational accident analysis through a systematic literature review method. In the fourth section, the results obtained will be discussed within the framework of the developed research questions. In the final section, the main findings of the study will be summarized, and recommendations for future research will be made.

2. The Basic Arguments and Theoretical Background

The concept of occupational accident has been defined by various international organizations and national legislations. According to the Turkish Occupational Health and Safety Law No. 6331, an occupational accident is defined as "*an event that occurs at the workplace or due to the performance of work, resulting in death or physical or mental disability*" [10]. Similarly, the Turkish Social Insurance and General Health Insurance Law No. 5510 details the circumstances under which an accident qualifies as work-related, such as incidents occurring at the workplace, during official duties, while commuting in employer-provided transport, or during breastfeeding breaks, if they result in physical or mental harm [11].

Internationally, the World Health Organization (WHO) defines an occupational accident as "*an unplanned event leading to personal injury, property damage, or interruption of work*" [12]. The International Labour Organization (ILO) similarly describes it as "*an event occurring during work that results in fatal or non-fatal injury*," while defining an occupational disease as "*a condition resulting from exposure to workplace hazards*" [13]. Within the European Statistics on Accidents at Work (ESAW) framework, an occupational accident is defined as "*an individual event occurring during work that causes physical or mental harm*," where "during work" refers to time spent performing professional duties [14].

In Türkiye, occupational accident data are officially recorded by the Turkish Social Security Institution (SGK). Before 2012, statistics only covered cases for which compensation payments had been finalized. Since 2013, however, in alignment with EU standards, all reported occupational accidents involving insured employees have been digitally registered [11].

2.1. Causes of Occupational Accidents

Occupational accidents arise from multiple underlying causes that may vary across regions, countries, industries, and organizational settings. Commonly reported factors include industrial structures, inadequate safety inspections, limited accident data and research, insufficient safety awareness, and employee qualifications, although country-specific variations may also exist [15].

Evidence shows that occupational accidents are rarely the result of a single factor but rather a sequence of interconnected causes. Previous studies estimate that approximately 80% of workplace accidents are associated with human-related factors, 18% with unsafe physical or mechanical conditions, and 2% with unexpected events. This indicates that nearly 98% of occupational accidents could be prevented through appropriate control measures [9].

While numerous elements—including technology, work environment, and sociological or psychological influences—may contribute to accident occurrence, the literature generally classifies root causes under two main categories: unsafe conditions and unsafe behaviors [12].

Table 1. Causes of Occupational Accidents [12]

Unsafe Conditions	Unsafe Behaviors
Unguarded Machinery and Machine Tools	Performing the Work Recklessly
Unsafe Working Methods	Absent-Mindedness and Carelessness
Unsafe and Unhealthy Environmental Conditions	Removing Machine Guards
Ungrounded Electrical Machines	Working at Dangerous Speeds
Unsuitable Hand Tools	Performing Work Outside of Duty
Uncontrolled and Untested Pressure Vessels and Lifting Machines	Failure to Comply with Work Discipline
Stacking at Dangerous Heights	Failure to Use Appropriate Machinery and Tools for the Job
Unsealed Gaps	Being in a Dangerous Area Without Authorization or Permission
Workplace Disorder	Failure to Use Personal Protective Equipment
	Driving Without a License and at Dangerous Speeds, etc.

2.2. Occupational Accident Statistics

Similar to their causes, occupational accident statistics differ across regions, countries, and industries, and are collected by various institutions worldwide. In Türkiye, these statistics are recorded and regularly published by the Turkish Social Security Institution (SGK).

At the European level, the European Statistics on Accidents at Work (ESAW) project was launched in 1990 to standardize occupational accident data across countries. The European Statistics on Accidents at Work – Methodology report, published in 2001 by Eurostat and the Directorate-General for Employment and Social Affairs, defines the data collection and classification framework, which remains in use today. According to ESAW standards, all accidents resulting in more than three days of work incapacity are included in official statistics. The methodology enables cross-country comparison by harmonizing detailed variables such as injury type, deviation events, and causal factors [9].

The ESAW framework provides a standardized dataset covering the following key dimensions [14]:

- Characteristics of the injured person,
- Characteristics of the injury, including severity (number of days lost),
- Characteristics of the business, including economic activity,
- Characteristics of the workplace,
- Characteristics of the accident, including the sequence of events that characterized the causes and conditions of the accident.

2.3. Machine Learning and Artificial Intelligence

ML applications are widely used to analyze large and complex datasets, often referred to as big data. ML is a fundamental subfield of artificial intelligence (AI) and forms the basis of many AI-driven technologies [7, 8]. AI refers to computational systems capable of learning from data and performing tasks typically associated with intelligent behavior [16, 17]. Modern AI technologies integrate methods such as machine learning and deep learning to address complex problems, automate processes, and enhance efficiency in industrial and practical applications [18].

Big data comprises datasets too large and intricate for traditional information systems to process effectively [19, 20], and is now widely used across sectors ranging from commerce and administration to scientific research.

ML algorithms are computational methods capable of deriving insights from data autonomously, learning patterns, and improving their accuracy through accumulated experience, and, are broadly categorized into supervised and unsupervised learning, depending on the type of data and learning process used [7, 8, 9]. Supervised learning is commonly applied in classification and regression tasks using labeled datasets [7, 9, 17], while unsupervised learning is used for clustering, anomaly detection, and dimensionality reduction [7, 9].

ML algorithms make predictions or decisions based on data patterns obtained from the training set, rather than following step-by-step instructions like traditional optimization methods [17]. Among the most widely adopted supervised learning algorithms are Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), Classification and Regression Trees (CART), and the Decision Tree (DT) algorithm [7, 8, 9, 17, 21, 22]. These algorithms are preferred for their interpretability, scalability, and strong performance in diverse problem domains. Regression-based models, including linear, non-linear, and logistic regression (LR), are also frequently used for predictive analytics and classification [7, 9]. Despite their broad applicability, selecting the appropriate algorithm depends on factors such as data structure, problem complexity, interpretability, and computational constraints, necessitating careful research to ensure efficiency and reliable results.

ML offers several advantages over traditional statistical techniques, including its ability to detect nonlinear patterns, learn autonomously over time, maintain high predictive accuracy even with missing or imperfect data, and reduce human intervention in modeling processes [9, 22, 23]. These strengths make ML well suited for domains where datasets are complex, heterogeneous, and influenced by multiple interacting variables.

While the existing literature on ML provides extensive guidance for application areas like economics, finance, demand forecasting, and pricing, there remains a notable gap concerning its applications in human factors engineering. In this field, where the interaction between humans and systems is studied to improve safety, performance, and ergonomics, the integration of ML has the potential to revolutionize analysis and decision-making processes. ML can support ergonomics and safety research by identifying relationships between environmental factors, human behavior, and system performance, and by analyzing sensor-based data to assess workload, fatigue, and risky postures.

Despite such promising opportunities, systematic studies providing methodological frameworks or comparative evaluations of algorithms tailored for ergonomics-based problems are scarce. The lack of comprehensive, domain-specific analyses limits the accessibility and usability of ML methods for ergonomics and safety researchers. Developing structured guides or reference frameworks for selecting algorithms in these contexts would help bridge this gap. Hence, researches aiming to establish practical guidelines for applying ML in these intact fields—detailing which algorithms perform best under certain data conditions, how to preprocess and label human-centered data effectively, and how to validate model outcomes for real-world usability, exploring how hybrid or ensemble approaches—combining multiple algorithms could enhance model robustness and result accuracy in an area deserving of focused attention.

3. Methodology

A Systematic Literature Review (SLR) is a research method that systematically, transparently, and reproducibly analyzes existing studies in the literature to answer a specific research question [24].

A SLR was conducted to examine studies on occupational accidents and machine learning (ML) published between 2010 and 2024. The literature search was performed in the YÖK National Thesis Database and the Web of Science (WoS) using a combination of keywords, including “occupational accident,” “machine learning,” and “data analysis.” In the YÖK database,

an initial search for “machine learning” yielded 1,302 graduate theses. After applying “accident” as an additional screening keyword, 11 theses were identified, of which 3 met the inclusion criteria for ML-based occupational accident research and were included in the review. In Web of Science, a combined search using “machine learning” and “accident” returned 4,522 publications. These records were filtered to retain peer-reviewed studies in the fields of science and engineering that simultaneously included “occupational accident,” “machine learning,” and “data analysis” in the topic or abstract, and provided full-text access. Following duplicate removal and eligibility screening, 25 studies were included in the final review. The screening and selection steps are summarized in the SLR flow diagram (Fig. 1). The descriptive characteristics of the included studies are presented in Table 2.

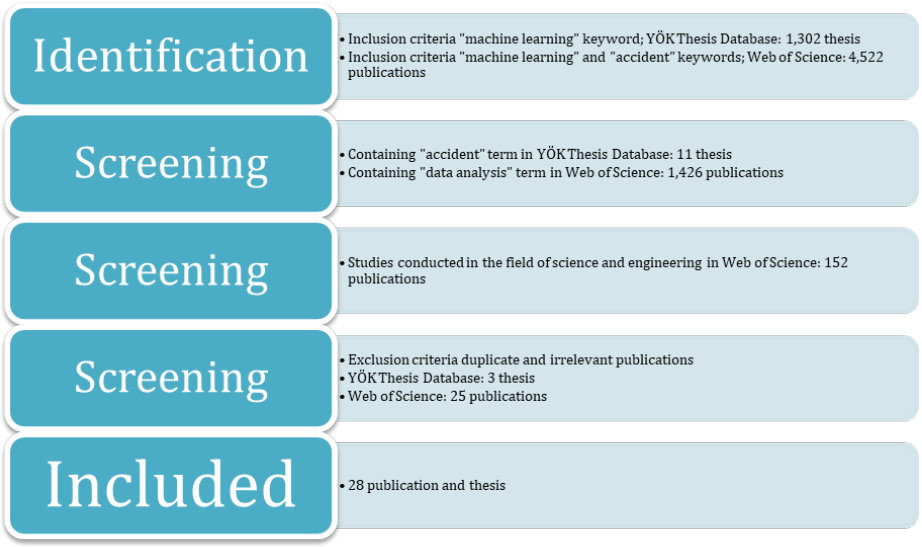

Fig. 1. Application SLR Flowchart

Table 2. Publication List

Publication	***Industry***	***Country***
Yang et al. [23]	*Construction*	*USA*
Goh & Ubeynarayana [24]	*Construction*	*USA*
Goh & Ubeynarayana [25]	*Construction*	*Singapore*
Kang & Ryu [26]	*Construction*	*South Korea*
Sarkar et al. [27]	*Metal*	*India*
Akşehir et al. [28]	*Construction*	*Türkiye*
Ayanoğlu and Kurt [29]	*Metal*	*Türkiye*
Choi et al. [30]	*Construction*	*South Korea*
Koç, Ekmekçioğlu and Gürgün [31]	*Construction*	*Türkiye*
Oyedele et al. [32]	*Construction*	*UK*
Zhu et al. [33]	*Construction*	*China*
Akaner [34]	*General*	*Türkiye*
Fuentes et al. [35]	*Cleaning*	*Spain*
Kang, Koo and Ryu [36]	*Construction*	*South Korea*
Koç, Ekmekçioğlu and Gürgün [37]	*General*	*Türkiye*
Özdemir Aydın [38]	*Mining*	*Türkiye*
Pan et al. [39]	*Construction*	*USA*
Recal [40]	*General*	*Türkiye*
Alkaissy et al. [41]	*Construction*	*Australia*
Çamlıyurt et al. [42]	*Maritime*	*South Korea*
Gholamizadeh et al. [43]	*General*	*Iran*
Karakaya Özkan [7]	*Metal*	*Türkiye*
Lee, Lee and Cho [44]	*General*	*South Korea*
Luo et al. [45]	*Construction*	*China*
Mammadov et al. [46]	*Pipeline construction*	-
Zermane et al. [47]	*Construction*	*Malaysia*
Das, Khanwelkar and Maiti [20]	*General*	*USA*
Karakaya Özkan and Ulaş [48]	*Metal*	*Türkiye*
Tunçman [49]	*Textile*	*Türkiye*

The research questions (RQs) were developed based on the thematic patterns and preliminary findings observed during the initial scoping phase of the systematic literature review. As part of the SLR process, an exploratory examination of existing studies revealed several recurring gaps, including inconsistent categorization of ML algorithms, limited sector-specific comparisons, unclear distinctions between the problem types addressed by ML models, and insufficient discussion on how ML methods are positioned relative to traditional analytical approaches. These observed deficiencies guided the structuring of the review and informed the formulation of the RQs. Accordingly, four RQs were designed to capture the main analytical dimensions emerging from the literature—algorithmic trends, sectoral distribution, problem-type classifications, and the reported advantages and limitations of ML-based approaches in occupational accident research. Together, these RQs establish a coherent framework for synthesizing the current state of knowledge.

The RQs developed for this study are as follows:

RQ.1: Which ML methods are most commonly used in the analysis of occupational accidents?

This question aims to identify algorithmic trends and examine which ML models are most frequently applied in the literature.

RQ.2: In which industries are ML algorithms more commonly used in the analysis of occupational accidents?

This question seeks to determine the industrial domains where ML-based safety analytics are concentrated on. The secondary aim is to examine the characteristics of occupational accidents in these industries and thereby develop safety measures on an industry-by-industry basis.

RQ.3: In which problem types are ML algorithms more commonly used in the analysis of occupational accidents?

This question aims to classify the analytical purposes of ML applications in the reviewed studies. By examining trends and development areas receiving more attention could be addressed and hence the research potential is aimed to be predicted.

RQ.4: How does the existing literature evaluate the strengths and limitations of ML algorithms compared to traditional analytical approaches in occupational accident research?

This question focuses on summarizing how previous studies position ML relative to conventional methods, and, investigates if the proven benefits of ML employment in case of OHS applications were derived.

4. Findings and Discussion

The findings obtained using the SLR method in this study are presented in this section. First, a general assessment of the number of occupational accidents and deaths in Türkiye based on SGK data was made, followed by an examination of the research questions. The number of occupational accidents that occurred between 2014 and 2023 according to SGK data is shown in Fig. 2.

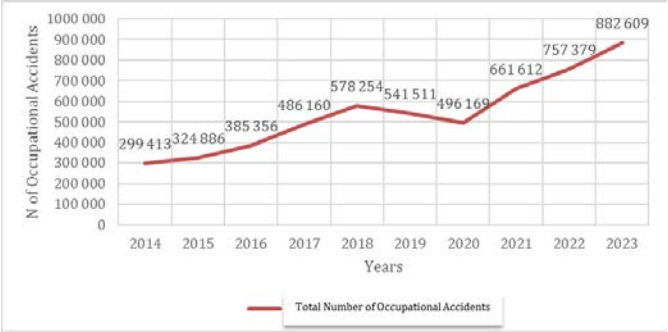

Fig. 2. Number of Occupational Accidents in Türkiye Between 2014 and 2023

Currently, efforts to prevent occupational accidents are mostly based on post-accident analysis. However, thanks to the opportunities offered by technological developments, it has become possible to predict accidents in advance and take preventive measures based on these predictions. In the industry-based analysis of the general occupational accident data presented in Fig. 2, not only the number of occupational accidents but also the incidence (frequency) and mortality (death) rates are taken into account. The incidence rate (Equation 1) is calculated by dividing the number of occupational accidents in the year under review by the number of employees in the industry for which the calculation is made in the same year and multiplying by 100,000 [9].

$$Incidance\ Rate = \frac{Number\ of\ Occupational\ Accidents}{Number\ of\ Employees}\ x\ 100\ 000 \quad (1)$$

As a result of this assessment, the highest incidence rate between 2014 and 2023 was observed in the "*Coal and Lignite Mining*" industry with NACE code "05" at 25.01%. This industry was followed by the "*Basic Metal Industry*" industry with NACE code "24" at 10.14%. During the same period, the industry with the highest number of occupational accidents was "*Building Construction*" (NACE code "41"), with a total of 288,706 accidents. Furthermore, the 2,622 fatal occupational accidents that occurred in this industry made the construction industry the area with the highest number of fatal accidents. The mining industry is one of the fundamental industries that carries out the process of finding and extracting minerals from underground in an economically efficient manner while minimizing environmental impacts [52]. With its broad scope of activities, the metal industry is at the forefront of other industries and contributes directly to many industries. It plays a fundamental role in many areas, from automotive to white goods, heavy industrial products to electrical and electronic devices. Various production and processing methods such as smelting, casting, forging, pressing, welding, cutting, and turning are widely used in this industry. The high production capacity and complexity of the processes make the industry risky in terms of occupational accidents. Due to the nature of manufacturing processes, continuous monitoring and attention are required at every stage of the workflow. In this respect, the metal industry is of critical importance in terms of OHS [53]. The construction industry is one of the most dynamic and important industries, both in terms of created employment and the added value to the economy. Operating largely with domestic capital, this industry encompasses many occupational groups and contributes directly to the production process. However, it is also the area where

occupational accidents occur most frequently. This situation leads not only to physical losses but also to significant psychological, social, and economic problems [54]. Based on real-life data obtained from Turkish Social Security Institution (SGK), the number of accidents, incidence rate, and mortality value in the mining, metal, and construction industries, which are the sub-industries with the highest rates, are presented in Fig. 3, while the number of deaths resulting from occupational accidents are presented in Fig. 4, within the framework of occupational safety and health practices and sectoral analyses of the occupational accident structure in Türkiye.

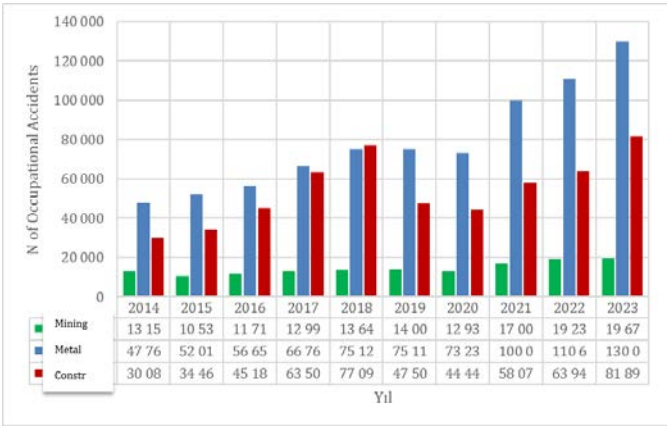

Fig. 3. Number of Occupational Accidents by Industry Between 2014 and 2023

While the number of occupational accidents in mining industry shows an upward trend over the years, the rate of increase in occupational accidents is lower compared to other industries. The number of occupational accidents in the metal industry increased by approximately 170% during the reviewed period, with a significant jump especially after 2020. This increase makes the metal industry one of the most perilous industries in terms of OHS conditions and occupational accidents. When construction industry data is examined, it is seen that the number of accidents followed a fluctuating trend; increasing until 2018, peaking that year, and then declining in 2019 and 2020 most likely due to pandemic catastrophe, before starting to increase again (Fig. 3).

Mortality data analysis by industry indicates that real-life data fluctuates due to the discontinuous and high-severity nature of accidents in the mining industry. According to data from the metal industry, the trend has generally been stable; however, data from the construction industry shows that this industry has the highest number of fatal accidents compared to other industries, and real-life mortality data shows a continuous upward trend (Fig. 4).

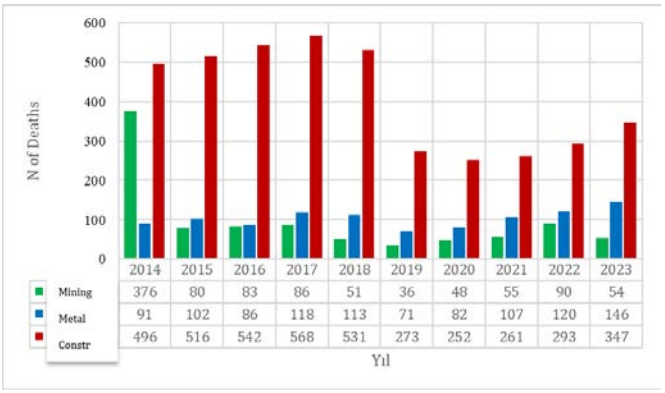

Fig. 4. Number of Deaths by Industry Between 2014 and 2023

Overall, the descriptive sectoral analysis reveals substantial variability in accident frequency and severity across industries, underscoring the importance of data-driven approaches for sector-specific risk assessment. Following the sectoral analysis and problem definition based on real-life data obtained from the Turkish Social Security Institution (SGK), these sectoral patterns highlight the need for analytical tools capable of capturing complex accident dynamics, especially in high-risk industries such as mining, metal, and construction. Traditional descriptive statistics provide useful insights but remain limited in modeling nonlinear relationships or predicting high-severity events. Therefore, the following sections examine how machine learning methods have been employed in the literature to address these challenges and how they respond to the research questions defined in this study.

4.1. RQ.1 Findings

The distribution of ML algorithms used in occupational accident analysis is presented in Fig. 5. The review indicates that RF is the most frequently applied algorithm across the examined studies. RF's dominance can be attributed to its strong performance on heterogeneous, nonlinear, and noise-prone datasets—characteristics commonly observed in occupational accident records.

SVM follows RF in frequency of use. In the broader ML literature, SVM is widely recognized as one of the most commonly employed algorithms for prediction tasks, particularly in studies involving small or medium-sized datasets and complex decision boundaries. Its consistent presence in occupational accident research therefore aligns with its established role in general ML applications.

In addition to RF and SVM, several other algorithms—including ANN, Logistic Regression (LR), DT, XGBoost, Naive Bayes, and KNN—have also been utilized. These algorithms represent a diverse set of methodological families, such as tree-based learning, probabilistic classification, and distance-based modeling. Their combined use illustrates the growing methodological variety in ML-driven occupational accident research and highlights the field's transition toward more sophisticated, data-driven analytical approaches.

In addition to their frequency of use, the reviewed studies indicate that algorithm selection is closely tied to dataset characteristics. RF and XGBoost are preferred in studies involving heterogeneous or imbalanced data due to their robustness and ability to capture nonlinear interactions. SVM generally performs well in smaller datasets with complex class

boundaries, while ANN-based models are used in problems requiring higher nonlinear approximation capacity. These patterns show that ML model choice in occupational accident research is shaped by both data structure and analytical objectives. These observations are consistent with prior research showing that RF and SVM frequently outperform traditional models in accident prediction tasks due to their robustness to noise and nonlinear structures [26, 37, 43].

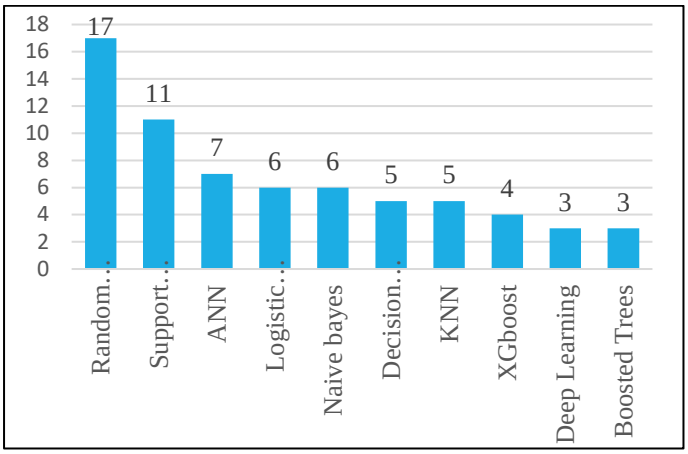

Fig. 5. Types of ML Algorithms Used in Occupational Accident Analyses

4.2. RQ.2 Findings

The sectoral distribution of studies applying ML algorithms to occupational accident analysis is presented in Fig. 6. The findings indicate that the construction industry is the most frequently studied sector in the existing literature. This predominance aligns with global and national accident statistics, as construction is consistently recognized as one of the highest-risk industries worldwide. As shown in Figs. 3 and 4, the Turkish Social Security Institution (SGK) data reveal that the "41 – Building Construction" sector experienced the highest number of occupational accidents in Türkiye between 2014 and 2023, with 288,706 reported incidents and 2,622 fatalities. The temporary and dynamic nature of construction worksites, together with the presence of numerous concurrent hazards, creates constantly shifting risk profiles, making the sector particularly suitable for ML-based predictive and diagnostic modeling [55].

The findings also show that a substantial number of ML-based studies adopt a general, industry-independent approach rather than focusing on a single sector. Such studies typically utilize aggregated datasets to investigate accident causation, pattern recognition, or predictive modeling across multiple industries simultaneously. Following general studies, the metal industry appears as another sector with notable ML adoption. This is consistent with the sector's high production volume, complex workflows, and increased accident rates, which create opportunities for the use of data-driven analytical tools.

Together, these results demonstrate that ML applications in occupational safety tend to emerge primarily in sectors characterized by high accident frequency, complex operational processes, and variable risk conditions—features that make traditional analytical methods less effective in capturing dynamic and nonlinear patterns.

In addition, several studies emphasize that sector-specific challenges, such as data fragmentation in construction, process complexity in metal manufacturing, and hazard variability in mining, directly influence the choice and performance of ML models. This suggests that effective ML implementation requires aligning algorithm capabilities with sectoral risk profiles and data structures. This sectoral concentration is consistent with previous research indicating that mining, metal, and construction industries inherently involve high-risk, complex, and variable operational conditions, making them frequent targets of ML-based accident analyses [52,53,54].

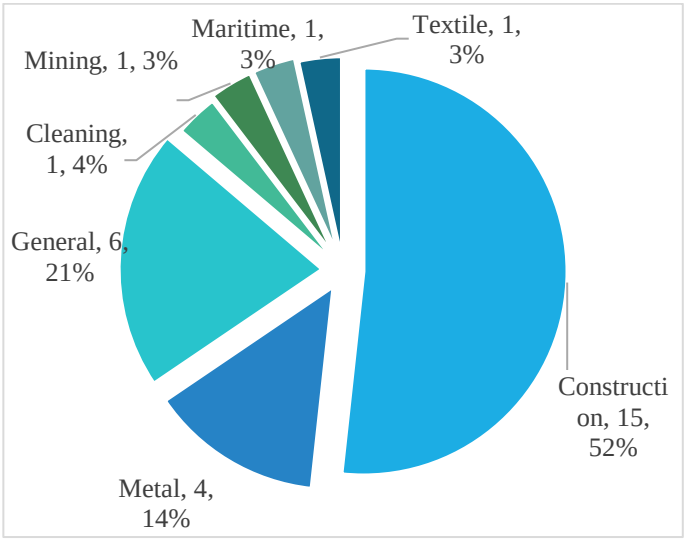

Fig. 6. Industries Using ML in Occupational Accident Analysis

4.3. RQ.3 Findings

The review conducted under RQ.3 examines the analytical purposes for which ML algorithms have been employed in occupational accident research. The findings reveal that prediction-oriented studies constitute the dominant trend in the literature. Specifically, 22 out of the 28 included studies (approximately 79%) focus on predicting the likelihood, frequency, type, severity, or potential consequences of occupational accidents. This emphasis on prediction reflects the growing interest in proactive safety management, as predictive models enable organizations to identify emerging risks before accidents occur and to support data-driven decision-making.

The remaining six studies apply ML algorithms primarily for classification purposes. These studies typically aim to categorize accidents or workers based on features such as injury type, hazard category, or demographic characteristics (e.g., gender, age, or occupation). Classification approaches are often used when the objective is to group similar accident patterns or to understand factors that differentiate one class of events from another.

Overall, the distribution of analytical purposes demonstrates that ML is predominantly used as a predictive tool in occupational safety research, reflecting the field's increasing shift toward anticipatory and prevention-focused risk assessment. The results of RQ.3 are summarized in Fig. 7, which presents the proportion of prediction and classification studies and highlights the methodological inclinations within the existing literature.

In addition, several studies emphasize that prediction-oriented models benefit from the ability of ML algorithms to capture

temporal trends, seasonal patterns, and multi-factor accident precursors, which traditional statistical models often overlook. Classification-focused studies, by contrast, tend to highlight the interpretability of ML outputs, particularly when identifying worker groups or hazard categories requiring targeted interventions. These findings are also consistent with previous research showing that ML methods are predominantly used for prediction tasks, such as forecasting unsafe behaviors, injury outcomes, or accident severity, with classification models applied less frequently and often limited to identifying injury types or worker subgroups [27, 37, 43, 48].

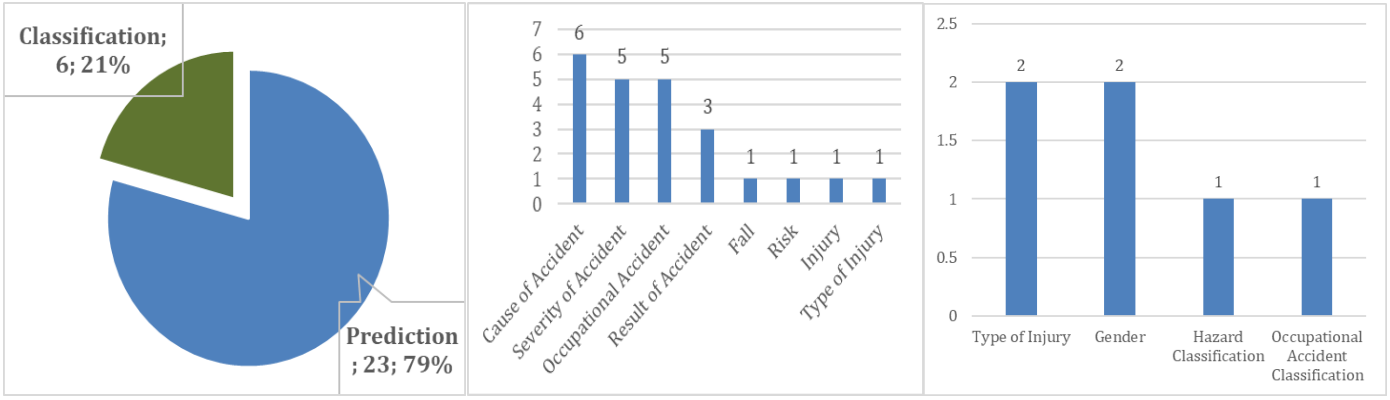

Fig. 7. Application Areas of ML Algorithms in Occupational Accident Analysis According to Problem Type

4.4. RQ.4 Findings

In the context of RQ.4, the review focuses on studies that report comparable performance metrics for both ML models and traditional analytical methods, as these allow for a meaningful assessment of the relative strengths of ML approaches. The comparative review of studies employing both machine learning (ML) algorithms and traditional analytical methods shows a consistent pattern in favor of ML-based approaches. Across the examined literature, ML models demonstrate a stronger capacity to capture complex, nonlinear, and multidimensional relationships inherent in occupational accident data. Unlike conventional statistical techniques, which often depend on predefined functional forms and simplified assumptions, ML algorithms can process large datasets efficiently and uncover hidden patterns that are typically overlooked in classical analyses. This capability becomes particularly valuable in industries such as construction, mining, and metal manufacturing, where accident mechanisms are shaped by interacting factors and rapidly changing operational conditions.

Another important advantage highlighted in the reviewed studies is the adaptive nature of ML models. By learning from historical accident records and updating their internal representations as new data become available, ML algorithms enable more dynamic and responsive safety strategies. Traditional models, in contrast, are generally static and less suited for environments characterized by temporal variability and fluctuating risk profiles.

The literature further indicates that ML methods consistently outperform traditional techniques in both predictive and classification tasks. For example, Goh and Ubeynarayana [25] compared six ML algorithms with logistic regression in predicting unsafe worker behaviors on a construction site and reported that all ML models achieved superior results. Fuentes et al. [35] similarly found that the SVM algorithm yielded more accurate predictions of injury occurrence and work incapacity than MARS models in the cleaning industry. In another study, Alkaissy et al. [41] applied XGBoost, RF, SVM, and logistic regression to predict injury types in the Australian construction sector, demonstrating that RF produced the highest performance across all evaluation metrics, while logistic regression showed the weakest results. Likewise, Mammadov et al. [46], who tested eleven different algorithms to predict accident outcomes in pipeline construction projects, concluded that deep learning methods provided the most reliable predictions.

Overall, the findings associated with RQ4 suggest that ML algorithms provide methodological and practical advantages that extend beyond improved accuracy. Their ability to process high-dimensional data, model nonlinear relationships, adapt to evolving conditions, and deliver more robust predictive insights positions them as increasingly valuable tools in occupational accident research. These advantages are consistently reflected across studies from multiple industries, indicating a clear shift toward data-driven approaches in modern OHS analytics.

5. Conclusion

This study presented a systematic review of ML applications in occupational accident research, synthesizing findings across industries, algorithms, and analytical purposes. The review shows that ML techniques have been increasingly adopted between 2010 and 2024 to analyze complex, high-dimensional accident datasets that traditional statistical methods often cannot capture effectively.

Across the four research questions, several consistent patterns emerged. RF and SVM were the most frequently used algorithms due to their ability to model nonlinear relationships, handle missing or imbalanced data, and achieve high predictive accuracy. Other methods such as ANN, DT, Naive Bayes, and XGBoost were also applied but less commonly. The construction industry was the most studied sector, reflecting its high global accident rates and inherent risks, while mining and metal industries followed similar trends. Prediction-focused studies dominated literature, indicating a shift toward proactive and preventive safety management.

Overall, the findings show that ML methods outperform traditional analytical approaches by better capturing complex accident dynamics and providing more reliable predictive

insights. However, important gaps remain, including limited sector-specific validation, small sample sizes, and insufficient algorithm benchmarking. The integration of ML with ergonomics and human factors also remains underexplored.

This review contributes to the field by mapping the methodological landscape of ML-based occupational accident research and emphasizing its potential to support data-driven OHS strategies. Future work should expand dataset diversity, enhance cross-industry comparisons, adopt explainable ML techniques, and incorporate multidisciplinary, human-centered perspectives to strengthen practical applications and promote safer workplaces.

Acknowledgement

This study is a research article based on an abstract proceeding study titled “Artificial Intelligence and Occupational Health and Safety Applications: Use of Machine Learning in Work Accident Analysis” which was presented orally at the 30th National Ergonomics Congress (2024, Karabük, Turkey) and whose only abstract was published.

Authorship Contributions

Authors contributed equally to this work

Declaration of conflicting interests

There are no conflicts of interest related to the manuscript.

Funding

No funds were received in support of this study.

Ethics

There are no ethical issues with the publication of this manuscript.

References

1. International Labour Organization. (2023). *World statistics.* Retrieved April 24, 2025, from https://www.ilo.org/global/topics/safety-and-health-at-work/lang--en/index.htm
2. Delibalta, M.S. (2020). The effects of logistics management on occupational health and safety in the mining sector. *Niğde Ömer Halisdemir University Journal of Engineering Sciences*, 9(1), 480–487.
3. Dündar, S., Bilim, A., & Bilim, N. (2024). Sectoral distribution and analysis of injuries from work accidents in Turkey. *Niğde Ömer Halisdemir University Journal of Engineering Sciences*, 13(1), 028–033.
4. Nefes, O., & Yılmaz Kaya, B. (2024). Yapay zeka ve iş sağlığı ve güvenliği uygulamaları: İş kazası analizinde makine öğrenmesi kullanımı. *30th National Ergonomics Congress*, University of Karabük, 10–12 October 2024, Karabük, Türkiye.
5. International Labour Organization. (2023). *Safety and health at work.* Retrieved April 24, 2025, from https://www.ilo.org/global/topics/safety-and-health-at-work/lang--en/index.htm
6. Nefes, O. (2025). *Analysis of work accidents using machine learning within the framework of regional and sectoral distributions* (master’s thesis). Gazi University Institute of Science, Ankara.
7. Akdoğan, H., & Yılmaz Kaya, B. (2024). Increasing production planning and control efficiency by using machine learning methods in standard time calculation. *V. International Applied Statistics Congress (UYIK - 2024)*, İstanbul, Türkiye.
8. Hayta, S., & Yılmaz Kaya, B. (2024). Work measurement in the new era: Literature review of standard time estimation with ANN. *The XIX International Symposium of Organizational Sciences – SymOrg 2024*, Zlatibor, Serbia.
9. Karakaya Özkan, E. (2023). *Statistical analysis of occupational accidents in the metal industry and prediction studies using machine learning algorithms* (Doctoral thesis). Gazi University Institute of Science, Ankara.
10. Official Gazette (No. 28339; June 20, 2012). (2024). *Law No. 6331 on Occupational Health and Safety.* Retrieved June 13, 2024, from https://www.mevzuat.gov.tr/mevzuatmetin/1.5.6331.pdf
11. Official Gazette (No. 26200; June 16, 2006). (2024). *Social Insurance and General Health Insurance Law No. 5510.* Retrieved June 13, 2024, from https://www.mevzuat.gov.tr/MevzuatMetin/1.5.5510.pdf
12. Turkish Chamber of Engineers and Architects Mechanical Engineers Chamber. (2015). *Occupational Health and Safety.* TMMOB MMO/617, Ankara, Turkey. Retrieved June 13, 2024, from https://mmo.org.tr/sites/default/files/gonderi_dosya_ekleri/c819fd7f01dd438_ek_0.pdf
13. International Labour Organisation. (2015). *Investigation of occupational accidents and diseases: A practical guide for labor inspectors.* Retrieved February 20, 2025, from https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@ed_dialogue/@lab_admin/documents/publication/wcms_346714.pdf
14. Eurostat. (2012). *European statistics on accidents at work (ESAW) summary methodology.* Retrieved September 23, 2024, from https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/ks-ra-12-002
15. Camkurt, M.Z. (2007). The effect of workplace working systems and physical factors on workplace accidents. *TUHİS Journal of Labor Law and Economics*, 20(6), 80–106.
16. Encyclopedia Britannica. (2024). *Work accident; causes of work accidents.* Retrieved May 29, 2024, from https://www.britannica.com/technology/artificial-intelligence
17. Yılmaz, A. (2022). *Artificial Intelligence.* Kodlab Publishing Distribution Software and Education Services, Istanbul.
18. Russell, S.J., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach.* Prentice Hall Series, USA.
19. Razbonyalı, C. (2016). *A study on big data and its impact on traditional data storage and processing methods* (Doctoral thesis). Okan University Institute of Science, Istanbul.

20. Aktan, E. (2018). Big data: Application areas, analytics, and security dimension. *Journal of Information Management*, 1(1), 1–22.
21. Nilsson, J.N. (2009). *The quest for artificial intelligence.* Cambridge University Press, USA.
22. Das, S., Khanwelkar, D.R., & Maiti, J. (2024). A semi-automated coding scheme for occupational injury data: An approach using Bayesian decision support system. *Expert Systems with Applications*, 237, 121610.
23. Petekci, A.R. (2021). *Theoretical Machine Learning.* KDY, Turkey.
24. Kitchenham, B. (2004). *Procedures for performing systematic reviews.* Keele University Technical Report: TR/SE-0401, Staffordshire, England.
25. Yang, K., Ahn, C.R., Vuran, M.C., & Aria, S.S. (2016). Semi-supervised near-miss fall detection for ironworkers with a wearable inertial measurement unit. *Automation in Construction*, 68, 194–202.
26. Goh, Y.M., & Ubeynarayana, C.U. (2017). Construction accident narrative classification: An evaluation of text mining techniques. *Accident Analysis and Prevention*, 108, 122–130.
27. Goh, Y.M., Ubeynarayana, C.U., Wong, K.L.X., & Guo, B.H.W. (2018). Factors influencing unsafe behaviors: A supervised learning approach. *Accident Analysis and Prevention*, 118, 77–85.
28. Kang, K., & Ryu, H. (2019). Predicting types of occupational accidents at construction sites in Korea using random forest model. *Safety Science*, 120, 226–236.
29. Sarkar, S., Vinay, S., Raj, R., Maiti, J., & Mitra, P. (2019). Application of optimized machine learning techniques for prediction of occupational accidents. *Computers and Operations Research*, 106, 210–224.
30. Akşehir, Z.D., Pekel, E., Akleylek, S., & Kılıç, E. (2019). Application of Bayesian networks in the occupational health and safety sector. *Turkish Informatics Foundation Journal of Computer Science and Engineering*, 12(1), 47–59.
31. Ayanoğlu, C.C., & Kurt, M. (2019). A work accident prediction model proposal using data mining methods in the metal industry. *Ergonomics*, 2(2), 78–87.
32. Choi, J., Gu, B., Chin, S., & Lee, J. (2020). Machine learning predictive model based on national data for fatal accidents of construction workers. *Automation in Construction*, 110, 102974.
33. Koç, K., Ekmekcioğlu, Ö., & Gurgun, A.P. (2021). Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability status of construction workers. *Automation in Construction*, 131, 103896.
34. Oyedele, A., Ajayi, A., Oyedele, L.O., Delgado, J.M.D., Akanbi, L., Akinade, O., Owolabi, H., & Bilal, M. (2021). Deep learning and boosted trees for injuries prediction in power infrastructure projects. *Applied Soft Computing*, 110, 107587.
35. Zhu, R., Hu, X., Hou, J., & Li, X. (2021). Application of machine learning techniques for predicting the consequences of construction accidents in China. *Process Safety and Environmental Protection*, 145, 293–302.
36. Akaner, M. (2022). *A neural network-based model proposal for determining hazard classes in activity areas* (Doctoral thesis). Gazi University Institute of Science, Ankara.
37. Fuentes, A.G., Serrano, N.M.B., Lasheras, F.S., Valverde, G.F., & Sanchez, A.S. (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. *Applied Ergonomics*, 105, 103847.
38. Kang, K., Koo, C., & Ryu, H. (2022). An interpretable machine learning approach for evaluating the feature importance affecting lost workdays at construction sites. *Journal of Building Engineering*, 53, 104534.
39. Koç, K., Ekmekcioğlu, Ö., & Gurgun, A.P. (2022). Accident prediction in construction using hybrid wavelet-machine learning. *Automation in Construction*, 133, 103987.
40. Özdemir Aydın, M. (2022). *Accident severity prediction of Zonguldak district underground coal mines by machine learning techniques* (Doctoral thesis). Middle East Technical University Institute of Science, Ankara.
41. Pan, X., Zhong, B., Wang, Y., & Shen, L. (2022). Identification of accident-injury type and body part factors from construction accident reports: A graph-based deep learning framework. *Advanced Engineering Informatics*, 54, 101752.
42. Recal, F. (2022). *Machine learning approach in creating a decision support system from national work accident data* (Doctoral thesis) Yıldız Technical University Institute of Science, Istanbul.
43. 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.
44. Camliyurt, G., Park, Y., Kim, D., Kang, W.S., & Park, S. (2023). Machine learning with multi-source data to predict and explain marine pilot occupational accidents. *Journal of Marine Science and Technology*, 31, 348–364.
45. Gholamizadeh, K., Zarei, E., Yazdi, M., Rodrigues, M.A., Shirmohammadi-Khorram, N., & Mohammadfam, I. (2023). An integration of intelligent approaches and economic criteria for predictive analytics of occupational accidents. *Decision Analytics Journal*, 9, 100357.
46. Lee, J., Lee, W., & Cho, S. (2023). Characteristics of fatal occupational injuries in migrant workers in South Korea: A machine learning study. *Heliyon*, 9, e20138.
47. Luo, X., Li, X., Goh, Y.M., Song, X., & Liu, Q. (2023). Application of machine learning technology for occupational accident severity prediction in the case of construction collapse accidents. *Safety Science*, 163, 106138.
48. Mammadov, A., Kazar, G., Koc, K., & Tokdemir, O.B. (2023). Predicting accident outcomes in cross-border pipeline construction projects using machine learning algorithms. *Arabian Journal for Science and Engineering*, 48, 13771–13789.
49. Zermane, A., Tohir, M.Z.M., Zermane, H., Baharudin, M.R., & Yusoff, H.M. (2023). Predicting fatal fall from

heights accidents using random forest classification machine learning model. *Safety Science*, 159, 106023.
50. Karakaya Özkan, E., & Ulaş, H.B. (2024). Comparison of four machine learning methods for occupational accidents based on national data on the metal sector in Turkey. *Safety Science*, 174, 106468.
51. Tunçman, B. (2024). *Evaluation of risk factors causing occupational accidents in the textile industry using data mining methods* (Master's thesis) Bursa Uludağ University Institute of Science, Bursa.
52. Demiroğlu, O. (2023). *Statistical analysis of occupational accidents and diseases in the mining sector* (Master's thesis).Graduate Education Institute, Hasan Kalyoncu University, Gaziantep.
53. Şahiner, E. (2024). Industry 4.0 and occupational health and safety: A study on the metal sector.
54. Çınar, M.C. (2018). *Turkish construction sector and its effects on the national economy* (Master's thesis) Institute of Social Sciences, Nevşehir Hacı Bektaş University, Nevşehir.
55. Bilim, A., & Çelik, O.N. (2018). General assessment of work accidents occurring in the construction sector in Turkey. *Ömer Halisdemir University Journal of Engineering Sciences*, 7(2), 725–731.