Safety Science 184 (2025) 106764

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

Project characteristics-based predicting the likelihood of occupational accidents in public school maintenances using a topological approach

Uğur Yiğit [a,*], Gökhan Kazar [b,c]

[a] Dept. of Mathematics, Istanbul Medeniyet University, 34700 Üsküdar, Istanbul, Turkey
[b] Dept. of Civil Engineering, Istanbul Medeniyet University, 34700 Üsküdar, Istanbul, Turkey
[c] Dept. of Civil Engineering, Istanbul Gelisim University, 34310 Avcılar, Istanbul, Turkey

ARTICLE INFO

Keywords:
Rough topology
Feature selection
Occupational accidents
Machine learning
Maintenance project
School buildings
Project characteristics

ABSTRACT

Occupational accidents are common in construction projects. Although several previous studies have focused on this complex issue from different perspectives, such as predicting accidents for the general construction process, few studies have focused on the impact of project characteristics on the likelihood of accidents in building maintenance projects. Artificial intelligence-based predictive models for workplace accidents typically use ready-made algorithms or traditional methods like trial and error. Therefore, the main objective of this study is to predict the likelihood of occupational accidents in building maintenance projects by following a new feature selection process based on the topological approach. The information on the 1807 public school maintenance project was included in this study to test the proposed mathematical approach. Commonly used 7 different machine learning algorithms and a combination of these algorithms called a hybrid model was selected to apply the topological approach to the feature selection process. The results show that 5 out of 7 algorithms such as Stochastic Gradient Boosting (SGB), Extreme Gradient Boosting (EGB), Linear Discriminant Analysis (LDA), Gaussian Naive Bayes (GNB), and Hybrid (HYB) models show better performance after applying the topological technique. The main predictors of the likelihood of workplace accidents in these algorithms are site delivery (T2), cost breakdown ratio (F1), total duration (T1), and contractor size (P1). Using this approach, construction professionals can develop and implement more effective AI-based proactive safety management systems for maintenance projects.

1. Introduction

Building maintenance is a critical step in maintaining the lifecycle of existing assets. Preventive maintenance protects against significant deterioration, reducing the need for expensive repairs and ensuring that the building remains a valuable asset. Regular maintenance ensures that buildings are safe for occupants. It helps identify and rectify structural problems, electrical faults, and other hazards that could lead to accidents or health problems, such as mold or poor air quality (Hauashdh et al., 2021). Maintenance also ensures that all building systems (such as HVAC, plumbing, and electrical) operate efficiently. This helps to reduce energy costs, minimize downtime, and extend the life of these systems (Kadhim and Altaie, 2023). Proper maintenance practices contribute to environmental sustainability by reducing waste, energy consumption, and the need for new materials and construction (Wang and Xia, 2015).

However, there are some challenges and occupational risks associated with building maintenance projects. For example, older buildings may have obsolete materials or systems that are difficult to repair or replace. Maintenance tasks often involve significant risks for workers, such as working at heights, handling electrical systems, and dealing with hazardous materials (Kadhim and Altaie, 2023). Proper planning and risk assessment are essential to identify these hazards and ensure that safety measures are in place. Workers must be trained in safety protocols, such as the use of personal protective equipment (PPE), the safe handling of tools and materials, and procedures for working in confined spaces or with hazardous substances (Chandra Dahal and Raj Dahal, 2020; Hauashdh et al., 2022). Maintenance projects require careful planning to incorporate these protocols and minimize the risk of accidents. Maintenance activities may also involve exposure to hazardous materials such as asbestos, lead paint, chemicals, or mold. These substances can pose health risks, including respiratory problems, skin irritation, or long-term illness (Zolkafli et al., 2019). Ensuring proper

* Corresponding author.
E-mail addresses: ugur.yigit@medeniyet.edu.tr (U. Yiğit), gokhan.kazar@medeniyet.edu.tr, gkazar@gelisim.edu.tr (G. Kazar).

https://doi.org/10.1016/j.ssci.2024.106764
Received 4 October 2024; Received in revised form 10 November 2024; Accepted 18 December 2024
Available online 24 December 2024
0925-7535/

procedures are in place to identify, remove, and dispose of these materials is crucial for occupational safety during maintenance projects. Workers may also have to perform maintenance in confined spaces such as ducts, tanks, basements, or crawl spaces where limited ventilation and access pose risks of suffocation, toxic exposure, or entrapment. Wet or uneven surfaces, loose cables, or debris can cause slips, trips, and falls, leading to various injuries such as broken bones or concussions during maintenance projects (Hauashdh et al., 2021; Ogunbayo et al., 2022). Maintenance work can be physically demanding and stressful, especially when working under pressure, in uncomfortable conditions, or during emergencies (De Marco et al., 2010).

For the reasons mentioned above, maintenance projects may result in accidents at work or safety violations. These cases are highly dependent on the characteristics of the existing buildings being repaired (Liu and Ramadhan, 2017; Shohet, 2003). Understanding the reasons for occupational accidents and predicting the likelihood of occupational cases is crucial to providing effective proactive systems in maintenance projects. At this point, data-driven approaches are essential to establish proactive systems for safety management. To establish efficient proactive systems, machine learning (ML) algorithms have gained importance in recent years. To date, many studies have focused on predicting workplace accidents in the construction industry using various ML techniques with conventional methods (Ayhan and Tokdemir, 2019; Baker et al., 2020; Choi et al., 2020; Kang and Ryu, 2019; Kang et al., 2022; Kim et al., 2019; Koc et al., 2023, 2022, 2021; Koc and Gurgun, 2022; Sarkar et al., 2020; Tixier et al., 2016; Xia et al., 2018). However, none of these studies attempted to predict the likelihood of occupational accidents in maintenance projects, as most of the predictions were heavily focused on the construction process of new buildings (Kang et al., 2022; Koc et al., 2023, 2021; Koc and Gurgun, 2022; Tixier et al., 2016). Furthermore, these studies followed traditional ML methods or used ready-made algorithms. Some of them also utilized common optimization algorithms to improve the accuracy rates (Ayhan and Tokdemir, 2019; Baker et al., 2020; Wang and Xia, 2015). The primary objective of this study is to improve the accuracy of predicting the likelihood of occupational accidents in building maintenance projects by implementing a novel feature selection process based on a topological approach. This approach allows us to identify and prioritize the key factors contributing to accident risk by analyzing the interconnected relationships between different project and environmental variables. By using this topological framework, we can capture complex dependencies and underlying structures that traditional feature selection methods may miss, ultimately leading to a more robust and accurate predictive model. This study aims not only to improve accident prediction in maintenance projects, but also to lay the groundwork for integrating advanced predictive techniques into proactive safety management systems in the construction industry.

Using this novel technique, we can evaluate the key characteristics associated with workplace accidents and assess how these characteristics affect the predictive accuracy of maintenance projects. This approach allows for a more nuanced understanding of risk factors such as project type, building age, equipment condition, worker experience and environmental hazards. Identifying the relative impact of these features could provide critical insights into the causal mechanisms behind workplace accidents. Such knowledge is an important step towards the development of a proactive health and safety monitoring system specifically tailored to building maintenance projects, where conditions are often dynamic and vary significantly from site to site. This predictive capability has significant implications for real-time risk management, enabling project managers to anticipate potential hazards and take preventative action before accidents occur. By integrating this model into the maintenance workflow, safety protocols can be adapted to the specific characteristics of each project and the existing conditions of the building. As a result, occupational incidents can be predicted and mitigated during the ongoing maintenance process, promoting a safer working environment. In addition, this proactive approach can improve resource allocation, reduce project delays due to safety concerns and ultimately support the goal of achieving zero accidents in the construction and maintenance industry.

2. Literature review

2.1. Occupational accidents in building projects

Workplace accidents on building projects are a significant problem in the construction industry, characterized by high injury and fatality rates. The construction sector is notorious for its hazardous working environment, with statistics suggesting that it accounts for around 30 % of all workplace accidents worldwide (Damayanti et al., 2023). The International Labour Organization reports that construction workers face a 1 in 300 chance of dying on the job, with falls being the leading cause of death (Izudi et al., 2017). This alarming trend requires a comprehensive understanding of the factors contributing to these accidents, and the implementation of effective safety measurements.

One of the main factors influencing workplace accidents in the construction industry is a lack of awareness of workplace hazards. Research shows that workers who are uninformed about potential risks are more likely to engage in unsafe practices, resulting in higher injury rates (Berhanu et al., 2019). Furthermore, the presence of social networks among workers can increase awareness of workplace hazards, suggesting that promoting communication and safety culture within teams can mitigate risks (Santiago et al., 2020). The importance of training and education in promoting safety cannot be overstated; studies have shown that health and safety training significantly reduces injury rates among construction workers (Comu et al., 2021; Gao et al., 2019; Hire et al., 2021; Karimi and Taghaddos, 2019; Kazar and Comu, 2021).

Workplace accidents in building maintenance projects are a major concern because of their potential to cause serious injuries and fatalities (Cheng et al., 2013; Hon et al., 2010). The construction industry is widely recognized as one of the most dangerous sectors, with a high incidence of accidents attributed to several factors, including inadequate safety measures, lack of training, and poor risk management practices.

A comprehensive understanding of the causes of occupational accidents in building maintenance can be enhanced through the application of risk assessment methodologies. For example, Carrillo-Castrillo et al., (2015) emphasize the importance of prioritizing protective measures when quantitative risk assessments are limited, suggesting that medium-risk tasks should be addressed with appropriate preventive barriers. Furthermore, (Zákányiné Mészáros et al., 2022) highlight that a robust risk assessment framework not only mitigates accidents but also leads to improved employee motivation and reduced costs associated with workplace injuries. This link between safety practices and organizational performance is critical to fostering a culture of safety on construction projects.

Human factors play a key role in occupational safety, particularly in maintenance operations. One study discusses how understanding human behavior and its influence on maintenance practices can significantly reduce the likelihood of accidents (Reiman, 2020). Inadequate supervision and disregard for safety regulations are often cited as the main causes of accidents in the construction industry, with studies showing that unsafe acts account for a significant percentage of incidents (Kamalinia et al., 2012). Therefore, improving supervision and ensuring compliance with safety protocols are essential steps in accident prevention.

There is also a relationship between building characteristics and the likelihood of occupational accidents during maintenance projects in construction, which are influenced by various factors (Carrillo-Castrillo et al., 2015; Hauashdh et al., 2020; Hon et al., 2010). Older or poorly maintained buildings might have weakened structures, increasing the risk of collapses or accidents during maintenance. Complex architectural designs with intricate features, such as unusual angles, narrow spaces, or overhangs, can create difficult working conditions that increase the likelihood of accidents. The method used to construct the building (e.g., prefabricated vs. traditional) can influence the risk during maintenance operations (Hon et al., 2010). Prefabricated buildings may have different joining techniques or weak points that pose unique risks. Buildings with poor or limited access can make it difficult for workers to safely reach maintenance areas. This can lead to unsafe practices, such as using ladders inappropriately or failing to secure work platforms. Dragone et al. (2021) argue that traditional maintenance management methods are becoming outdated and innovative approaches need to be adopted to ensure safety and functionality. This is particularly relevant as the failure to perform timely maintenance can lead to hazardous conditions that increase the likelihood of accidents.

The role of management and organizational practices in preventing accidents is also crucial. Effective safety management systems, such as OHSAS 18001, have been shown to significantly reduce accident rates by promoting a culture of safety and compliance with regulations (Rajaprasad and Chalapathi, 2015). However, barriers to effective safety communication persist, often exacerbated by the transient nature of the workforce in the construction industry (Pandit et al., 2019). This highlights the need for ongoing engagement, digital systems, and training to ensure that all workers, regardless of employment status, are equipped with the knowledge and tools to work safely. Addressing occupational accidents in building maintenance projects requires a multifaceted approach that encompasses risk assessment, human factors analysis, economic considerations, and the adoption of modern maintenance practices. By prioritizing safety and implementing comprehensive OHS strategies, the construction industry can significantly reduce the incidence of workplace accidents, thereby safeguarding the well-being of workers and enhancing overall project performance.

2.2. Prediction studies on occupational accidents in buildings

Prediction studies on occupational accidents in buildings have gained significant attention due to the high incidence of occupational accidents in the construction industry. Various methodologies, including machine learning and statistical analysis, have been employed to understand the key reasons for accidents and predict occupational cases. These methodologies are crucial for developing effective preventive strategies and improving workplace safety.

The application of machine learning (ML) to the analysis of workplace accidents in the construction industry has received considerable attention in recent years. Several studies have investigated different ML techniques and their effectiveness in predicting and mitigating workplace accidents (Ayhan and Tokdemir, 2019; Baker et al., 2020; Choi et al., 2020; Kang and Ryu, 2019; Kang et al., 2022; Kim et al., 2019; Koc et al., 2023, 2021; Koc and Gurgun, 2022; Sarkar et al., 2020; Tixier et al., 2016; Xia et al., 2018). This approach presents key findings from the literature, focusing on the methodologies used, the types of accidents analyzed, and the implications for occupational safety.

A systematic review by Khairuddin et al. (2022) highlights the usefulness of text-based analysis in identifying causal factors of occupational injuries, which can be critical in developing predictive models. This approach is consistent with the findings of Tixier et al. (2016) who used random forest (RF) and stochastic gradient tree boosting (SGTB) models to predict safety outcomes based on a dataset derived from textual injury reports. These methods highlight the importance of data-driven approaches in understanding the complexities of workplace accidents.

In the construction sector, Lee et al. (2020) point out the importance of data preprocessing and accident prediction modeling, noting that historical accident data provides a valuable foundation for ML model development. This view is echoed by Koç et al. (2022) who argue that the integration of resampling techniques with ML methods can effectively deal with unbalanced datasets, which are common in accident datasets. Their research suggests that predictive modeling is a promising way to improve the safety performance of civil engineering works.

Furthermore, (Toptanci et al., 2023) provided insights into the factors influencing the severity of occupational injuries, identifying critical variables such as age, education, and work environment through logistic regression models. This is in line with the findings of Mohammadfam et al. (2015) who used Artificial Neural Networks (ANNs) to analyze factors affecting occupational accidents and found that individual and organizational factors significantly contribute to the severity of accidents. Such findings are crucial for tailoring safety training and preventive interventions in the workplace.

The role of machine learning in occupational safety goes beyond prediction to include risk assessment and mitigation strategies. For example, a study by Gondia et al. (2022) presents a decision-support framework for predicting injury severity and implementing risk mitigation measures. This framework demonstrates the practical applications of ML in real-world scenarios, emphasizing the need for actionable insights derived from predictive analytics.

Furthermore, the challenges associated with the implementation of ML in occupational safety are highlighted by Sarkar et al. (2020) They discuss the need to optimize ML techniques for effective accident prediction (Sarkar et al., 2020). They note that the integration of different ML algorithms can increase prediction accuracy, thereby improving safety outcomes in construction environments. This is particularly relevant in light of the findings of Choi et al. (2020) who developed a national data-based prediction model for construction worker fatalities, demonstrating the potential for large-scale applications of ML in safety analysis.

2.3. Research gaps and aims

Existing studies mainly focus on common machine learning models, without a comprehensive evaluation of their comparative effectiveness in different contexts. For example, Lee et al. (2020) and Sarkar et al. (2020) highlight the application of optimized ML techniques to predict occupational accidents but do not systematically compare the performance of these models with different methods. This lack of comparative analysis limits the understanding of which ML techniques provide the most reliable predictions in different occupational settings. Furthermore, while Koç (2023) provides insights into the prediction of accident susceptibility, it does not address the integration of these predictive models into a coherent framework that can be universally applied across different industries, particularly in construction.

Furthermore, while some research has begun to explore the socio-technical aspects of occupational safety, as highlighted by Elkhoukhi et al. (2022), there is a lack of empirical studies that incorporate project characteristics into ML models. The 'new view' of safety emphasizes the importance of worker adaptability and resilience, yet current ML applications often overlook these critical elements, focusing instead on quantitative data. A more holistic approach, combining quantitative ML techniques with qualitative insights, could lead to more effective safety interventions.

All mentioned studies that looked at estimating occupational accidents from diverse perspectives used conventional methods to incorporate machine learning models. The majority of ML-based models used pre-built or generic algorithms such as Genetic Algorithm (GA), Case-based Reasoning (CBR), Artificial Neural Networks (ANN), Linear Discriminant Analysis (LDA), Gaussian Mixture, KNN, and regression models by following traditional approaches (Ayhan and Tokdemir, 2019; Baker et al., 2020; Choi et al., 2020; Kang and Ryu, 2019; Kang et al., 2022; Khairuddin et al., 2022; Koc et al., 2023, 2022, 2021; Koc and Gurgun, 2022; Lee et al., 2020; Mohammadfam et al., 2015; Tixier et al., 2016; Xia et al., 2018). In addition, these studies have mainly focused on predicting accidents for general construction processes using traditional optimization techniques (Baker et al., 2020; Kang et al., 2022; Koc et al., 2023, 2021). None of the studies took into account accidents at work that occurred during maintenance construction projects with different feature selection processes. In other words, the unusual working environment in maintenance projects which is highly related to project characteristics is not considered in previous studies. One study by Hon et al. (2010) only considered the causes of accidents in maintenance projects by following a statistical method rather than introducing a predictive model. While these studies add much to the body of knowledge, new methodological approaches are needed to develop effective ML models and create more reliable proactive systems based on data-driven methods. Therefore, the main objective of this study is to introduce a new feature selection model based on the topological approach to forecast occupational accidents in building maintenance projects. By following this method, it is not necessary to use any optimization methods to achieve higher accuracy rates. With this ML-based model, the topological approach can also provide improved accuracy results without the need for parameter tuning or trial-and-error techniques. By following a different feature selection process, it is possible to reduce the number of attributes and improve the accuracy rates, which will provide solutions for data collection issues in construction workplaces. This method can also provide opportunities to interpret the importance of project characteristics-based attributes whether has a great impact on the likelihood of occupational accidents or not. It is expected that by harnessing the power of ML models with this new technique, construction stakeholders and government decision-makers will be able to proactively identify occupational risk factors related to project characteristics, improve decision-making, and ultimately increase the overall quality and safety of maintenance projects based on the project characteristics. Furthermore, the novel feature selection technique combined with ML-based algorithms will make it possible to create a proactive system that is reliable and efficient in preventing occupational accidents in maintenance projects.

3. Research methodology

3.1. Maintenance project dataset and preparation

The current study attempts to predict the likelihood of workplace accidents, whether they occur or not (yes or no) in school maintenance projects, using a novel ML method based on the topological approach. Although it is difficult and complex to understand the main causes of occupational accidents in construction workplaces, we have tried to focus on the project characteristics of maintenance projects that could have an impact on the likelihood of occupational accidents. To do so, we collected a total of 18 different attributes of 2018 public school maintenance projects implemented by the Turkish Ministry of Education between 2016 and 2022 (Table 1). First, the school maintenance projects were categorized into five main parts: general (G), building characteristics (B), project characteristics (P), financial (F), and time (T). The general dataset included the following characteristics: type of school (e.g. primary, secondary, and tertiary); location information (e.g. urban or rural); the age of the school; total square meters; the number of classrooms and students. Building characteristics relate to window (e.g., double hang, single hang, cavement, awning, and glas block etc.), roof (e.g., rubber, metal, clay, slate, floow, composite), electrical (e.g., joninig box and looping system) and heating characteristics (e.g., boiler, furnaces, heat pumps, hybrid, radiant and basebords) of public schools. Project characteristics included the size of the contractor (e.g. small, medium, or large) and the type of maintenance. The financial characteristics of maintenance projects included the cost breakdown ratio, approximate cost, allocated budget, and total cost paid recorded for each maintenance project. For these financial characteristics, we used the raw numbers for each project recorded. The ML model also included temporal information about maintenance projects, such as the total duration of the maintenance project and the season of the site handover (winter, summer, spring, and autumn) by the public owner. The target parameter was the likelihood of occupational accidents during these maintenance projects. If an occupational accident occurs during the school maintenance project, it will be coded as 1; otherwise, it will be coded as 0. The outcome of the occupational accident (fatal, lost workday, injury) was not considered in the ML models. The main focus of this study is to predict the likelihood of occupational accidents based on the different project characteristics using a new mathematical method, rather than predicting the outcome of occupational accidents. We were able to collect project data for a total of 2018 maintenance projects in Turkey, however, 211 projects were dropped due to missing values in some maintenance projects. The elimination of 211 projects due to missing data is unlikely to have a significant impact on the overall validity of the results, as they represent only about 10 % of the original sample of over 2,000 projects. Assuming that the excluded projects were missing at random and did not have systematic characteristics that would introduce bias, their removal does not affect the representativeness of the remaining data. With over 1,800 projects remaining, the sample size is still large enough to maintain statistical power and ensure the robustness and generalizability of the findings. As long as the excluded projects are not systematically different in terms of variables, the integrity of the conclusions remains intact. Finally, a 19x1807 matrix was generated and then we used our topologically based machine learning methods for the feature selection process.

Table 1
Attribute list for school maintenance projects.

CharacteristicCategory	Attributes	ID	DatasetType
General	Age of school	G1	Numerical
	Location	G2	Categorical
	Total square metres	G3	Numerical
	Number of classrooms	G4	Numerical
	Number of students	G5	Numerical
	School type	G6	Categorical
Building	Window properties	B1	Categorical
	Roof properties	B2	Categorical
	Properties of electrical installation	B3	Categorical
	Heating properties	B4	Categorical
Project	Size of contractor	P1	Categorical
	Type of maintenance	P2	Categorical
Financial	Cost breakdown ratio	F1	Numerical
	Approximate cost	F2	Numerical
	Allocated budget	F3	Numerical
	Total paid cost	F4	Numerical
Time	Total duration	T1	Numerical
	Season of site delivery	T2	Categorical

3.2. Topological method for data analysis

Rough set theory was first introduced by Pawlak (1982) and has since been used in many fields such as feature selection (Yang and Yang, 2012), decision systems (Liou, 2009; Shen and Loh, 2004), missing feature value problems (Kryszkiewicz, 1998; Salama, 2010), and medical analysis (Raza and Qamar, 2019; Thivagar et al., 2012; Zhang et al., 2009). By Pawlak's theory (Pawlak, 2002), objects are classified using an equivalence relation called the indiscernibility relation, which is explained below. Thivagar et al. (2012) used the rough topology to research real-world problems like medical data analysis. However, this method was only applied to nominal-type data. A generalization of the rough topology and the core to any numerical data is provided by Yiğit (2022). By transforming any numerical data to an ordinal form, which is viable for analyzing the data with this method, Yiğit (2022) gives a new topological method for dimension reduction of large-scale data processing.

Equivalence Relations. Let U be a non-empty set of objects. A relation R on U is a subset of the cartesian product $U \times U$. Typically, an element $(a, b) \in R$ is read as a is related to b. If $(a, a) \in R$ for all $a \in U$, then a relation R on U is said to be reflexive. It is said to be symmetric if $(a, b) \in R$ requires $(b, a) \in R$ for all $a, b \in U$. It is said to be transitive if $(a, b) \in R$ and $(b, c) \in R$ then $(a, c) \in R$ for all $a, b, c \in U$. A relation R on U is said to be an equivalence relation if and only if R is reflexive, symmetric, and transitive. Let R be an equivalence relation defined on a set U and $a \in U$. Let $[a]_R = \{x \in U | (a,x) \in R\}$ called the equivalence class of a under R, which is the set of all elements in U that are related to a. The set of all equivalence classes U/R of R in U gives a partition of U, which means that every equivalence class is disjoint, and U is the union of all equivalence classes.

Approximations. (Pawlak, (2002)) Assume that U is a non-empty finite set and R is an equivalence relation on U. A pair (U,R) is called an approximation space. Let X be a subset of U.

i. $\underline{R}(X) = \cup_{x \in U}\{x | [x]_R \subset X\}$ is called the lower approximation of X with respect to R.
ii. $\overline{R}(X) = \cup_{x \in U}\{x | [x]_R \cap X \neq \varnothing\}$ is called the upper approximation of X with respect to R.
iii. $B_R(X) = \overline{R}(X) - \underline{R}(X)$ is called the boundary region of X with respect to R.

If $B_R(X) \neq \varnothing$, then the set X is called a rough set with respect to R.

Topology. (Munkres, 2000) A topology on a set U is a collection τ of subsets of U satisfying the following properties:

(T1) $\varnothing$, $U \in \tau$

(T2) The union of the elements of τ is in τ.

(T3) The intersection of the finite number of elements of τ is in τ.

The pair (U,τ) is called a topological space. A basis for (U,τ) is a collection $\beta \subset \tau$ such that for each $A \in \tau$ and each $x \in A$, there exists $B \in \beta$ such that $x \in B \subset A$.

Rough Topology. (Thivagar et al., 2012) Let U be a non-empty finite set and R be an equivalence relation on U. A rough topology on U with respect to X is defined by $\tau_R = \{U, \varnothing, \underline{R}(X), \overline{R}(X), B_R(X)\}$ for $X \subset U$.

A basis for rough topology. (Thivagar et al., 2012) A basis for the rough topology τ_R on U with respect to X is $\beta_R = \{U, \underline{R}(X), B_R(X)\}$.

Example. Let $U = \{1,2,3,4,5,6,7,8\}$ be the set of objects, and $R = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6),(7,7),(8,8),(1,3),(3,1),(1,8),(8,1),(8,3),(3,8),$

$(4,6),(6,4)\}$ be an equivalence relation on U. Then the set of equivalence classes of U by the equivalence relation R is $U/R = \{\{1,3,8\},\{4,6\},\{2\},\{5\},\{7\}\}$. For $X = \{1,4,6\}$, $\underline{R}(X) = \{4,6\}$, $\overline{R}(X) = \{1,3,4,6,8\}$ and $B_R(X) = \{1,3,8\}$. Therefore, the rough topology $\tau_R = \{U,\varnothing,\{4,6\},\{1,3,4,6,8\},\{1,3,8\}\}$. The basis for τ_R is $\beta_R = \{U,\{4,6\},\{1,3,8\}\}$.

Core. (Thivagar et al., 2012) A subset M of the set of attributes is called the Core of R if $\beta_M \neq \beta_{(R-\{r\})}$ for every r in M. That is, the elements of the core cannot be removed without affecting the classification power of attributes.

Let (U,R) be a pair, where U is a set of objects (rows of the data) and R is an equivalence relation. R is defined as two rows that are related to each other if and only if all of their attribute values are the same. Nonetheless, most of the attribute values are not the same as one another. It is necessary to provide a method for converting numerical data into an ordinal format so that the core and rough topology may be used for analysis. We apply Algorithm 1, which utilizes the standard deviations of attribute values, to accomplish this.

Algorithm 1. ((*Yigit, 2022*)) **Step 1**: Given a data table, columns of which are attributes, rows of which are objects, and entries of the table are attribute values, pick one of the attributes.

Step 2: Take the maximum (*Max*) and the standard deviation (*St*) of the chosen attribute.

Step 3: Assign 1 to the values between (*Max*) and (*Max* −*St*) and discard these rows. Find the next maximum after discarding rows assigned as 1. Assign 2 to the values between the new (*Max*) and new (*Max* −*St*) and discard them. Repeat this process until every value is assigned to a new value.

Step 4: Repeat Step 2 and Step 3 for every attribute.

Step 5: Generate the new data table.

Example. Consider the table which shows 8 objects, 3 attributes, and a result (Table 2).

By applying the algorithm 1 to Table 2, we have Table 3:

Let $U = \{1,2,3,4,5,6,7,8\}$ be the set of objects, and $A = \{6,7,8\}$ be the set of objects which have a positive result. Let R be an equivalence relation, in which two objects are related to each other if and only if their values of all attributes are the same (Table 3 and Table 4).

Result: If we compare topologies (Table 4), we have only $\tau_{(R-Y)} \neq \tau_R$. So, we get $Core(A) = \{Attribute Y\}$. By taking $A = \{1,2,3,4,5\}$ as the set of objects having a negative result, similarly, we have $Core(A) = \{Attribute Y\}$. As a result, "Attribute Y" is the most crucial attribute in determining whether an object has a positive result or negative result.

Since the rough topology may only be changed by a single row, it is not suitable for large numerical data sets. As stated differently, topologies vary even when the border and the lower approximation are separated by a single object (row). However, the change can be quantified below. $\nu_r(X)$ is one tool that counts the number of rows that change the topology. Put otherwise, it's a method for determining which attributes in this classification are more important than others. It also lists these attributes in order of significance.

Accuracy. (Yigit, 2022) Let (U,R) be an approximation space and M be the set of attributes. For $r \in M$ and $X \subset U$, the accuracy of an attribute and the accuracy of the core with respect to X are defined respectively by

$$\mu_r(X) = \frac{|B_R(X)|}{|B_{(R-r)}(X)|}$$

$$\mu_{RT}(X) = Sup_{r \in M}\{\frac{|B_R(X)|}{|B_{(R-r)}(X)|}\}$$

Measurement for the core. (Yigit, 2022) Let (U,R) be an approximation space and M be the set of attributes. For $r \in M$ and $X \subset U$, the accuracy of the boundary for r with respect to X is defined by

$$\nu_r(X) = |B_R(X) - B_{(R-r)}(X)|$$

Remark. $\nu_r(X) = \nu_r(U - X)$ for any $r \in M$.

$\nu_r(X)$ is used to measure and arrange the data's tolerated unimportant object (rows). In other words, the attribute r has no bearing on this method's classification of the data if we delete rows in the $\nu_r(X)$ from the data. More examples can be found in (Yigit, 2022) (Fig. 1).

4. Results

4.1. Data core

Let X be the set of objects (rows) of the dataset resulting in an accident. The following are the outcomes and the $Core(X)$ that we acquire when we use the aforementioned topological technique for our data. The number of rows that alter the dataset's classification is measured by $\nu_r(X)$. Thus, taking into account at most $\nu_r(X) < 24$, which indicates a margin of maximum error of 1.3 percent (24 rows) across 1807 rows in the dataset, we extract at most 24 rows and establish our $Core(X) = \{1,7,10,11,12,13,14,15,16,17\}$. Put another way, these 10 attributes are the most crucial for the dataset's classification. Furthermore, as shown in

Table 2
Sample data.

Objects	X	Y	Z	Results
1	0,90	153,4	17,33	No
2	0,73	74,08	23,41	No
3	0,78	94,03	25,53	No
4	1,15	27,23	26,5	No
5	0,78	94,44	16,67	No
6	0,78	23,56	19,26	Yes
7	0,74	14,67	20,49	Yes
8	0,97	15,7	15,66	Yes
STD DEV	0,14	50,211	4,129	

Table 3
Converted sample data.

Objects	X	Y	Z	Results
1	2	1	2	No
2	3	2	1	No
3	3	2	1	No
4	1	3	1	No
5	3	2	2	No
6	3	3	2	Yes
7	3	3	2	Yes
8	2	3	3	Yes

Table 5, the attributes' respective orders of importance are 14, 12, 15, 16, 7, 1, 13, 10, 11, and 17, which are T2, F1, T1, P1, B2, G1, F2, G6, P2, F4, respectively. You can find the code link to calculate the values in the table below at Acknowledgment.

4.2. Machine learning algorithm results

In this section, we apply seven different machine learning algorithms that are commonly used in construction management issues to our data: Support Vector Classifier (SVC), Random Forest Classifier (RFC), Linear regression (LR), Stochastic Gradient Boosting (SGB), Extreme Gradient Boosting (EGB), Linear Discriminant Analysis (LDA), Gaussian Naive Bayes (GNB), and Hybrid (HYB). In this case, the hybrid algorithm is the

Table 4
Results for sample data.

	Equivalence Classes	Approximations	Topology
All Attributes	$U/R = \{\{1\},\{2,3\},\{4\},\{5\},\{6,7\},\{8\}\}$	$\overline{R}(A) = \{6,7,8\}\underline{R}(A) = \{6,7,8\}B_R(A) = \varnothing$	$\tau_R = \{U, \varnothing, \{6, 7, 8\}\}$
Remove X	$U/(R-X) = \{\{1\},\{2,3\},\{4\},\{5\},\{6,7\},\{8\}\}$	$\overline{(R-X)}(A) = \{6,7,8\}\underline{(R-X)}(A) = \{6,7,8\}B_{(R-X)}(A) = \varnothing$	$\tau_{(R-X)} = \{U, \varnothing, \{6,7,8\}\}$
Remove Y	$U/(R-Y) = \{\{1\},\{2,3\},\{4\},\{5,6,7\},\{8\}\}$	$\overline{(R-Y)}(A) = \{5,6,7,8\}\underline{(R-Y)}(A) = \{8\}B_{(R-Y)}(A) = \{5,6,7\}$	$\tau_{(R-Y)} = \{U, \varnothing, \{5, 6, 7, 8\}, \{8\}, \{5,6,7\}\}$
Remove Z	$U/(R-Z) = \{\{1\},\{2,3,5\},\{4\},\{6,7\},\{8\}\}$	$\overline{(R-Z)}(A) = \{6,7,8\}\underline{(R-Z)}(A) = \{6,7,8\}B_{(R-Z)}(A) = \varnothing$	$\tau_{(R-Z)} = \{U, \varnothing, \{6, 7, 8\}\}$

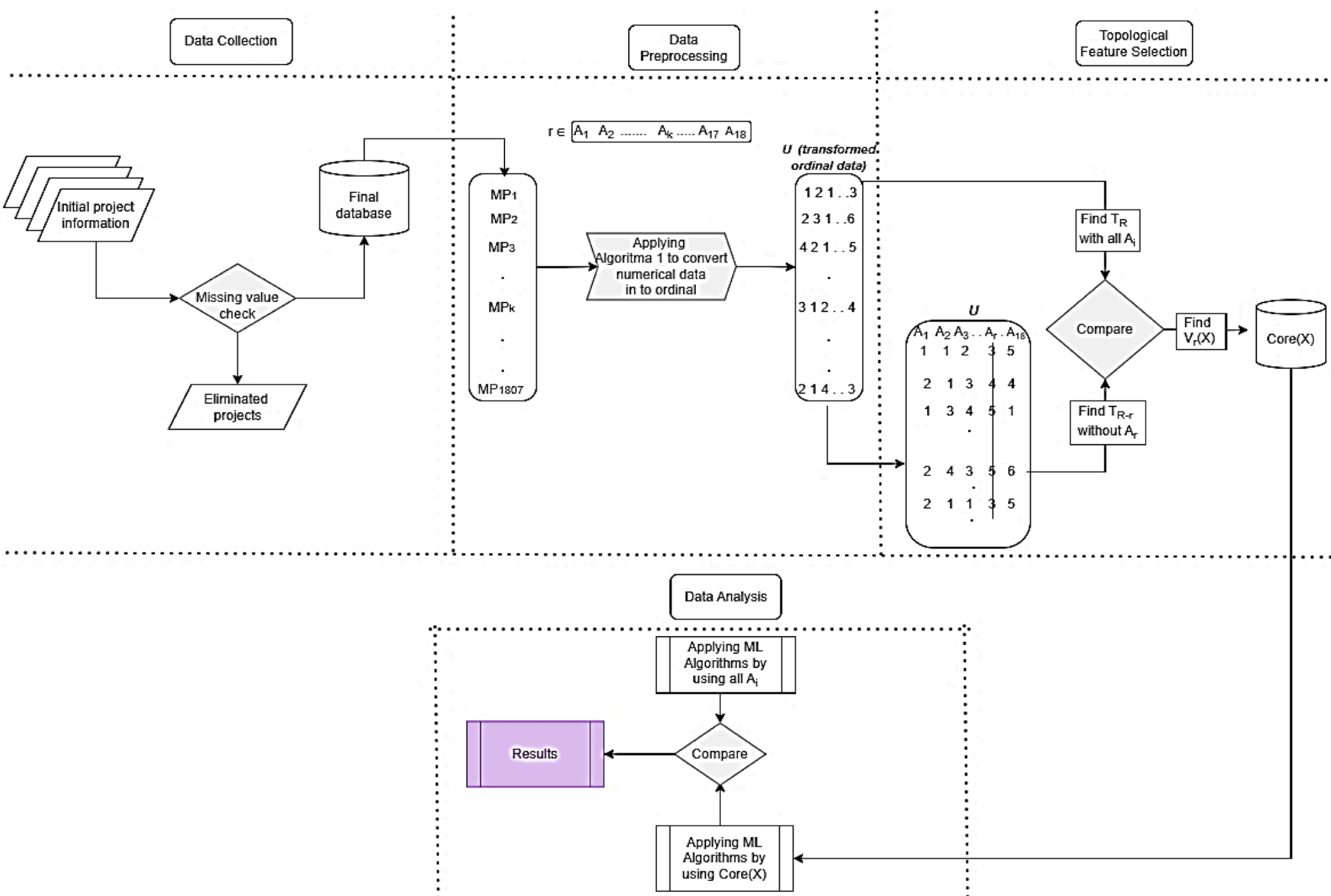

Fig. 1. Research methodology.

Table 5
Core dataset.

Column	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Attribute(r)	**G1**	G2	G3	G5	G4	B1	**B2**	B3	B4	**G6**	**P2**	**F1**	**F2**	**T2**	**T1**	**P1**	**F4**	F3
$\boldsymbol{v_r(X)}$	**52**	7	4	8	3	18	**63**	11	22	**42**	**42**	**105**	**50**	**109**	**103**	**81**	**41**	24

one that is developed according to whether or not four out of seven algorithms make accurate predictions. First, we use all attributes; after that, we use the attributes in the $Core(X)$ and then compare them.

Machine learning algorithms are carried out without any optimizations or parameters using a random selection procedure. Without using any optimization approaches, we employed ML algorithms in the default setup to demonstrate the effectiveness of the topological dimension reduction method. Since our data is homogeneous and balanced for binary classification, thus we employ fixed random state train-test splits that appropriately depict the distribution of classes as a whole. We divided the dataset in half, using 80 % for model training and 20 % for fixed-random selection testing (Koc et al., 2022; Koc and Gurgun, 2022). We consider executing the machine learning algorithms several times, comparing the average result, and then providing metrics.

Subsequently, the outcomes are then compared for every class and set of techniques. When evaluating the mean classification accuracy and other outcomes, using the $Core(X)$ yields superior results. It follows that, in contrast to all attributes as predicted, the approach employed here performed better for the 10 attributes in the $Core(X)$ (Fig. 2). More importantly, this classification includes the most significant attributes. Additionally, the order of the significance of the attributes is also provided in this prediction.

Table 6 and Table 7 demonstrate the classification outcomes of the seven ML algorithms and the hybrid algorithm. After applying topological dimension reduction by using the rough topological model for defining the Core and employing it, we get better prediction performance in five algorithms used (SGB, EGB, LDA, GNB, and Hybrid), the same results for two algorithms (SVC, RFC), but LR, as shown in tables. SGB, RFC, and Hybrid algorithms are more efficient than others, and they have accuracy rates between 0,754 and 0,771. GNB and SVC have the lowest prediction results according to others. In addition, one of our other aims in this study is to give which of the attributes are more important and the order of importance of these attributes. Looking at the results obtained in all algorithms by using the Core, it is concluded that this model works very well in this dataset.

Recall that the class-wise distribution of a classification model's predicted performance is called a confusion matrix. A 2×2 confusion matrix is a table used in classification tasks to evaluate the performance of a model. It has four cells: True Positives (TP), where the model correctly predicts the positive class; True Negatives (TN), where it correctly predicts the negative class; False Positives (FP), where it incorrectly predicts a positive class for a negative instance; and False Negatives (FN), where it incorrectly predicts a negative class for a positive instance. This matrix helps calculate key metrics like accuracy, precision, recall, and F1-score.

Below in Figs. 4 and 5, we give the confusion matrices that we created for the hybrid algorithm, using all the features and the features in the core, respectively. The confusion matrices of all other algorithms are given in Table 6 and Table 7.

Topological dimension reduction (TDR) is fundamentally a dimensionality reduction algorithm. In contrast to existing techniques in this field, it preserves the attributes in the original data set without directly causing any data loss. Furthermore, the user can choose how many dimensions to remove because it also provides the importance level of the attributes. We also shared with you the cases where the entire data set was used during the comparisons, but our rival in this area is Principal Component Analysis (PCA), the most well-known dimensionality reduction algorithm. As you can see in Fig. 3, we typically outperformed PCA in most cases, but LDA. Despite being entirely written in Python, it operates incredibly quickly.

When machine learning algorithms trained on Core Attributes are compared to those trained on PCA transformed data, Core Attributes typically produce higher accuracy. For algorithms such as SVC, RFC, LR, SGB, EGB, GNB and the Hybrid model, Core Attributes consistently outperforms PCA, with differences in accuracy ranging from 0.056 to 0.136. The most significant differences are observed in SGB and the Hybrid model, where Core Attributes achieves a significantly higher accuracy of 0.136 and 0.119 respectively. RFC and EGB also perform significantly better with core attributes, with a difference of 0.133 and 0.122 respectively. However, LDA is the only exception where PCA slightly outperforms Core Attributes by 0.011. Overall, Core Attributes shows superior performance in the majority of models, suggesting that the PCA transformation reduces accuracy for this dataset (Fig. 3).

5. Discussion

By following the topological approach to feature selection, we introduced a new ML-based model to predict the likelihood of occupational accidents in the maintenance of public school buildings. The results show that project characteristics such as season of site delivery (T2), cost breakdown ratio (F1), total duration (T1), and size of contractor (P1) are the most important characteristics that have a negative impact on the likelihood of occupational accidents in maintenance projects.

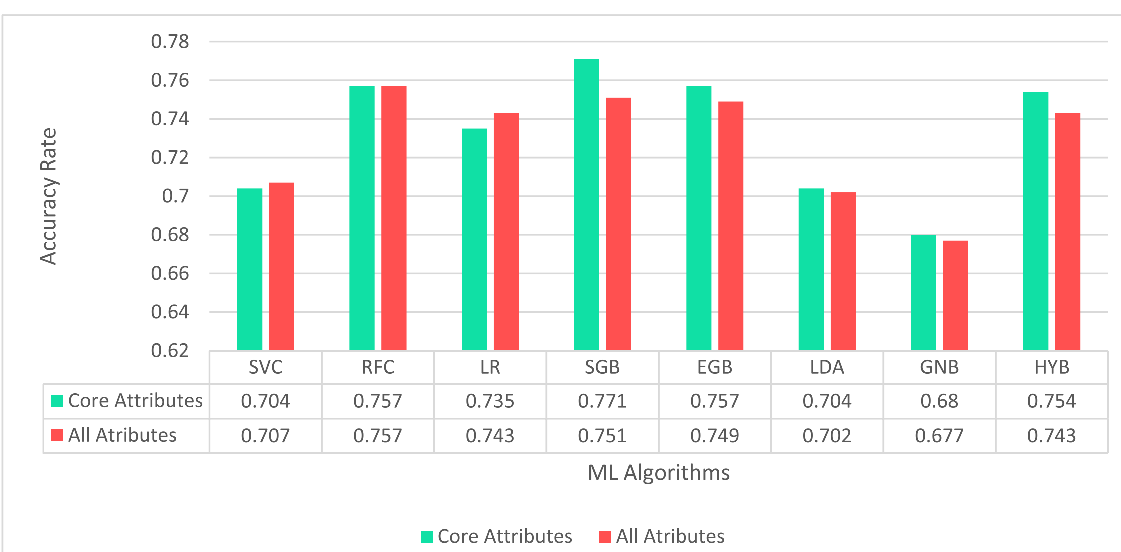

Fig. 2. Machine learning algorithm accuracy results.

Table 6
Use all attributes.

Parameters/Methods	SVC	RFC	LR	SGB	EGB	LDA	GNB	HYB
Accuracy	0.707	0.757	0.743	0.751	0.749	0.702	0.677	0.743
Precision	0.690	0.754	0.733	0.748	0.776	0.694	0.675	0.724
Recall	0.965	0.907	0.925	0.907	0.841	0.934	0.929	0.951
F1-score	0.805	0.823	0.818	0.820	0.807	0.796	0.782	0.822
True Negatives	38	69	60	67	81	43	35	54
True Positives	218	205	209	205	190	211	210	215
False Positives	98	67	76	69	55	93	101	82
False Negatives	8	21	17	21	36	15	16	11

Table 7
Use the core.

Parameters/Methods	SVC	RFC	LR	SGB	EGB	LDA	GNB	HYB
Accuracy	0.704	0.757	0.735	0.771	0.757	0.704	0.680	0.754
Precision	0.689	0.757	0.727	0.768	0.772	0.694	0.677	0.731
Recall	0.960	0.898	0.920	0.907	0.867	0.942	0.929	0.960
F1-score	0.802	0.821	0.812	0.832	0.817	0.799	0.783	0.830
True Negatives	38	71	58	74	78	42	36	56
True Positives	217	203	208	205	196	213	210	217
False Positives	98	65	78	62	58	94	100	80
False Negatives	9	23	18	21	30	13	16	9

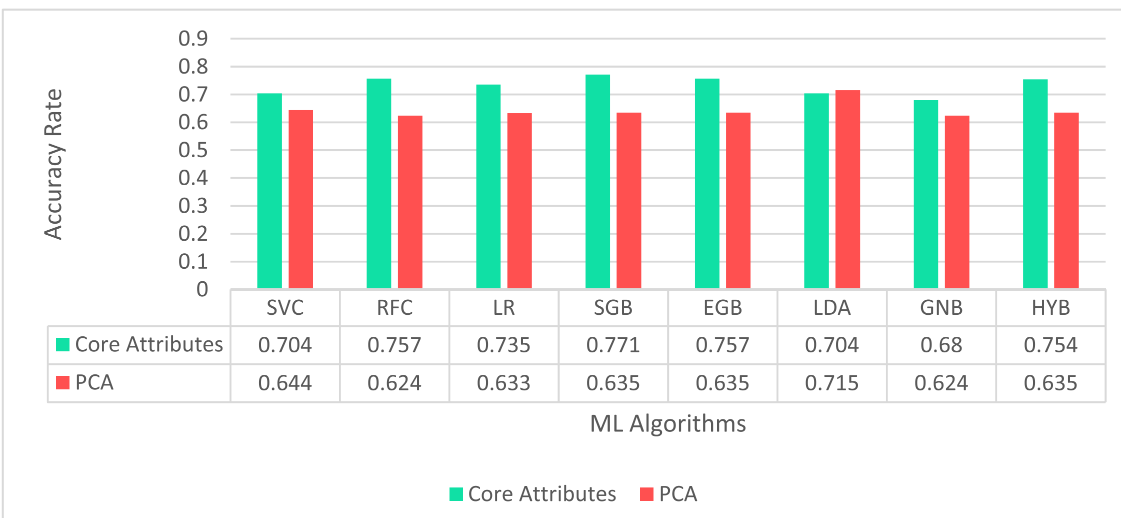

Fig. 3. PCA and core attributes results.

The season of site delivery is a crucial factor in stimulating occupational accidents in maintenance projects (Bendak et al., 2022; Kazar and Comu, 2022). Most maintenance projects are carried out during the summer season when schools are closed. High temperatures on summer days can cause heat stress, which can lead to fatigue, dehydration, and heat-related illnesses such as heat exhaustion or heat stroke (Yi and Chan, 2015). Fatigue reduces workers' alertness, reaction time, and decision-making ability, increasing the likelihood of accidents. Extreme heat can affect cognitive functions, reducing workers' ability to concentrate and increasing the likelihood of errors or lapses in safety protocols (Bendak et al., 2022). Wearing protective gear becomes more uncomfortable in hot weather, leading some workers to improperly use or remove essential safety equipment, exposing themselves to hazards. The physical strain of working in hot conditions can aggravate pre-existing health conditions, leading to accidents. Heat also increases the risk of muscle cramps, which can lead to sudden loss of control or stability (Szer et al., 2021).

Another critical parameter is the total duration of maintenance projects. This parameter is also strongly linked to the season of the maintenance projects. As most maintenance projects need to be carried out in the summer and completed before school starts, there is a time constraint on these projects. Under tight deadlines, workers and supervisors may feel pressured to complete tasks quickly, leading to rushed work. This can lead to cutting corners, skipping safety protocols, or neglecting proper procedures to save time, increasing the risk of accidents (XU et al., 2021). Time pressure often leads to longer working hours and fewer breaks. The pressure to meet tight deadlines can increase workers' stress levels. High levels of stress can distract workers, reduce their ability to concentrate, and lead to impulsive decisions that compromise safety (Han et al., 2014). Time constraints may also force supervisors and managers to prioritize speed over quality, resulting in poor inspections, missed safety checks, and undetected hazards (Man et al., 2021). This oversight can lead to unsafe working conditions.

Another important predictor of accidents on maintenance projects is the cost breakdown ratio. Most contractors cut construction costs to win the project from public owners. If a large proportion of the budget is allocated to other areas of the maintenance project, there may be insufficient funds left for critical safety measures. This can result in

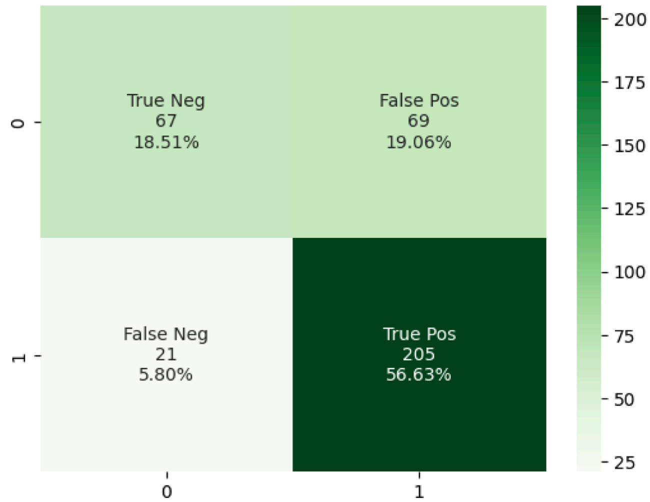

Fig. 4. Confusion matrix for hybrid algorithm (all attributes).

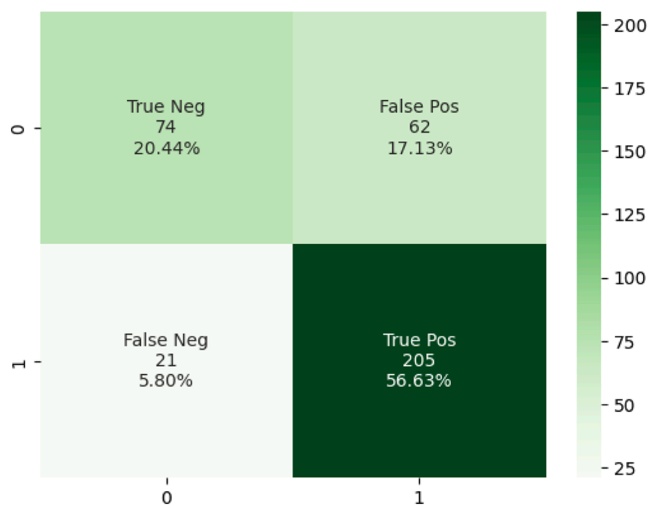

Fig. 5. Confusion matrix for hybrid algorithm (the core).

inadequate safety equipment, fewer safety training programs, or the use of cheaper, less reliable materials and tools, all of which can increase the likelihood of accidents (Wong et al., 2020). To save costs, projects may hire fewer workers, resulting in increased workloads for existing staff. Overworking workers can lead to fatigue, stress, and burnout, all of which increase the risk of errors and accidents on the job (Nnaji and Karakhan, 2020). Cost constraints can also lead to inadequate on-site supervision. Proper supervision is essential to ensure that safety procedures are followed, hazards are identified and mitigated, and workers perform their tasks safely. Without adequate supervision, the risk of accidents increases (Bhagwat et al., 2022).

Another important parameter is the size of the contractor carrying out the maintenance projects. Most contractors for maintenance projects are small and medium-sized enterprises (SMEs), which could lead to various occupational issues in these projects. Workplace accidents are particularly prevalent among SMEs in the construction industry due to a combination of factors that undermine safety protocols and management practices (Hon et al., 2010). One of the main reasons for the higher incidence of accidents in SMEs is inadequate enforcement of occupational health and safety (OHS) regulations (Al Mawli et al., 2021). Research shows that small businesses are inspected less frequently by regulatory bodies than their larger counterparts, leading to a lack of compliance with safety standards and inadequate monitoring of workplace conditions (Ahn, 2022; Singh and Misra, 2021). SMEs often operate on tight budgets, which can limit their ability to invest in safety equipment, training, and personnel. This can lead to inadequate safety measures, outdated or inadequate protective equipment, and less frequent safety inspections. SMEs may struggle to keep up with changing regulations and standards due to limited administrative support (Unnikrishnan et al., 2015). Failure to comply with safety regulations can increase the risk of accidents, as workers may not be working in a safe environment. Workers in SMEs often have to perform multiple roles due to limited staffing. This can increase the risk of accidents, as workers may perform tasks for which they are not fully trained, or work in high-risk areas without adequate support or supervision (Al Mawli et al., 2021).

The use of the topological technique to select new features from this data is one of the major theoretical contributions of the study. This approach eliminates the need for additional trial-and-error data training and parameter tuning processess. A further theoretical contribution of this method is that, although the number of features is reduced, we can still achieve improved performance in diverse ML-based algorithms without using optimization algorithms. Using the topological approach, it is also possible to derive the ranking of the most significant predictors. As a result, we can identify the key features that influence cost overruns in maintenance projects. We can use this method to evaluate each case and determine how cases affect cost overruns. Using this innovative approach, we have developed a novel ML-based prediction system for workplace accidents in the maintenance of public school buildings.

The practical contribution of this study is that the technique can be used to create new proactive systems that track occupational problems and their causes during maintenance projects, providing important insights for addressing construction safety issues. Contractors and construction professionals can focus on the key factors that indicate occupational injuries in maintenance projects. AI algorithms can analyze historical accident data alongside project-specific details (e.g. site location, project type, materials used, seasonal conditions) to identify patterns that correlate with higher risk. These insights can help anticipate potential risks early in the project design phase. Machine learning models, such as decision trees or neural networks, can predict the likelihood of accidents by analyzing the complex relationships between site characteristics, worker demographics, equipment used and project phases (Khairuddin et al., 2022). With more training, these models become more accurate at predicting risk factors. For example, using project-specific data, an algorithm could identify when scaffolding collapses are most likely to occur based on specific weather, schedule or material conditions. AI can analyse data from Internet of Things (IoT) sensors embedded in equipment and infrastructure to detect abnormal patterns, such as vibrations or temperature increases, which can signal potential failures or unsafe conditions (Koc and Gurgun, 2022). VR combined with AI can simulate dangerous construction scenarios, allowing workers to train in a controlled environment. AI can tailor simulations based on individual learning patterns, helping workers practice responses to potentially dangerous situations. For example, workers could virtually experience a situation where a structural failure occurs and learn the correct response, building muscle memory and speeding up reaction time in real life. Machine learning models can also be developed to predict the outcomes of various emergency scenarios, improving preparedness and reducing the risk of fatal accidents during construction (Tixier et al., 2016).

Data-driven approaches are critical to establishing effective proactive systems because they enable organizations to make informed decisions based on evidence rather than intuition or guesswork (Gondia et al., 2022; Sarkar et al., 2020). Data-driven approaches enable the accurate identification of potential risks and hazards before they become serious problems. By analyzing historical data, trends, and patterns, organizations can predict where problems are likely to occur and take preventative action to mitigate these risks (Koc et al., 2023, 2022). With access to accurate and comprehensive data, decision-makers can assess situations more objectively and make more informed decisions. This leads to more effective strategies for preventing accidents, optimizing processes, and improving overall safety and efficiency. Data-driven systems often incorporate real-time data collection and analysis, allowing organizations to monitor conditions as they develop and respond immediately to emerging threats or changes (Koc et al., 2021). This rapid response capability is essential to minimize the impact of potential problems.

The present research also has some limitations. The proposed novel feature selection method based on the topological approach was limited to the likelihood of work accidents in public school maintenance projects. Also, to evaluate the effectiveness of our method, we randomly selected only seven supervised machine learning algorithms. Different data sets and ML-based algorithms, such as unsupervised ones, can be added to these settings. The model could also include a variety of features to identify key predictors of workplace accidents. This model can be used to predict not only the likelihood of workplace accidents, but also schedule delays, cost overruns, or the outcome type of safety injuries. The data was collected in one region for a specific type of project, so new data from different regions can be included to generalize the results of the study to whole maintenance projects. Due to data accessibility limitations, we were unable to obtain extensive accident data from public maintenance projects in schools during the course of this study. It is well-known fact that public authorities have restriction to provide more detailed data due to laws and regulations. However, we do recognize that a comparison would provide a deeper validation of the identified risk factors and introduced ML algorithm by using a new feature selection method. The relation between the construction quality or type and its age can influence accidents in various projects. One could concerns about whether poor technical conditions of construction element were identified as contributing factors to accidents. The technical condition of the elements, such as material fatigue, and outdated design standards, are indeed significant risk factors. As buildings or infrastructure age, the likelihood of accidents can increase if the construction elements have deteriorated or if they were poorly designed. It is important to thoroughly assess both newer and older projects to ensure any potential risks due to poor construction or aging are identified and addressed. In this context, a more comprehensive study can be carried out as a future project by expanding the data with attributes such as corrosion or rusting of metal parts, outdated materials or construction techniques, deterioration of load-bearing elements over time. Other critical factors such as the risk management system, worker characteristics, worker-related factors such as human behaviour, demographic characteristics, mental health, working conditions, site characteristics, health and safety management system can have a significant impact on the likelihood of accidents. For example, safety training type of frequency can be coded for each project under new attributes. Overall, occupational safety in construction is a complex issue because of the number of factors that need to be considered. We have not been able to collect this information for all public schools as it requires more detailed research and official permissions from public authorities for all school buildings. Besides, the dataset should be expanded to include a wider range of projects such as infrastructure and infrastructural projects rather than only focusing maintenance projects to improve the representativeness and reliability of the findings. All these limitations can be addressed in future studies.

6. Conclusion

The current research aimed to predict occupational accidents in school building maintenance projects using a novel topological approach. A total of 1,807 public school buildings were renovated for various purposes between 2016 and 2022, providing a substantial dataset for testing our model. Using this methodology, we evaluated the effectiveness of our model with seven different ML-based algorithms, as well as a hybrid model, and successfully identified key factors that predict the likelihood of workplace accidents. The results show that the SGB, EGB, LDA, GNB and Hybrid models performed significantly better after implementing the topological technique, demonstrating the value of this approach in improving prediction accuracy. A particularly valuable aspect of this topological approach is its ability to improve the predictive performance of ML models by streamlining the feature set, reducing the number of required attributes while maintaining or even improving prediction rates. This feature is particularly beneficial in construction and maintenance contexts where data collection can be limited or difficult. By isolating only the most critical factors, this method allows us to build accurate, proactive safety systems without the burden of collecting extensive, potentially redundant data. In addition, the efficiency of this approach extends to the model development process, significantly reducing the time required for data pre-processing and model training. This advantage allows decision makers to implement safety models more quickly and with greater agility, a critical factor in responding to the dynamic, high-risk environments typical of building maintenance projects. As a result, this approach provides a practical solution for predicting and mitigating accidents, even in environments with tight timelines or limited data access. Beyond immediate project safety, the use of this topological approach has wider implications for the construction industry. By setting a precedent for efficient, data-driven risk management, this method can inspire future research and development of safety systems that are both resource-efficient and highly effective. In this way, the study not only contributes to the immediate goal of accident prediction, but also supports a shift towards sustainable, data-optimized approaches to construction safety management, fostering a proactive culture focused on preventing incidents before they occur.

All in all, decision-makers and construction professionals can effectively control the problems associated with these projects by understanding the key elements that significantly influence the likelihood of construction accidents in maintenance projects and developing predictive models using innovative feature selection techniques. Construction project management can be enhanced by using proactive systems and advanced ML techniques.

Data availability statement

Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request. See the link "https://github.com/EvReN-jr/TDR-Topological-Dimensional-Reduction" for Python and C++ codes to find the converted data, the approximations, topology, and the Core.

CRediT authorship contribution statement

Uğur Yiğit: Writing – review & editing, Visualization, Methodology, Formal analysis, Data curation. **Gökhan Kazar:** Writing – original draft, Validation, Investigation, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to thank the executives who anonymously provided the data and construction experts who participated in the feedback process for their efforts and contributions.

Data availability

Data will be made available on request.

References

Ahn, S., 2022. Derivation of risk factors to quantify the risk of safety accidents for small and medium-sized enterprises in construction industry. Sustain. 14. https://doi.org/10.3390/su14127306.

Al Mawli, B., Al Alawi, M., Elazouni, A., Al-Mamun, A., 2021. Construction SMEs safety challenges in water sector in Oman. Saf. Sci. https://doi.org/10.1016/j.ssci.2020.105156.

Ayhan, B.U., Tokdemir, O.B., 2019. Predicting the outcome of construction incidents. Saf. Sci. 113, 91–104. https://doi.org/10.1016/j.ssci.2018.11.001.

Baker, H., Hallowell, M.R., Tixier, A.J.P., 2020. AI-based prediction of independent construction safety outcomes from universal attributes. Autom. Constr. https://doi.org/10.1016/j.autcon.2020.103146.

Bendak, S., Jouaret, R., Rashid, H., 2022. Effects of high ambient temperature on construction workers performance: A longitudinal empirical study. J. Safety Res. 81, 197–202. https://doi.org/10.1016/j.jsr.2022.02.011.

Berhanu, F., Gebrehiwot, M., Gizaw, Z., 2019. Workplace injury and associated factors among construction workers in Gondar town, Northwest Ethiopia. BMC Musculoskelet. Disord. 20. https://doi.org/10.1186/s12891-019-2917-1.

Bhagwat, K., Delhi, V.S.K., Nanthagopalan, P., 2022. Construction safety performance measurement using a leading indicator-based jobsite safety inspection method: case study of a building construction project. Int. J. Occup. Saf. Ergon. 28, 2645–2656. https://doi.org/10.1080/10803548.2021.2012350.

Carrillo-Castrillo, J.A., Rubio-Romero, J.C., Guadix, J., Onieva, L., 2015. Risk assessment of maintenance operations: the analysis of performing task and accident mechanism. Int. J. Inj. Contr. Saf. Promot. https://doi.org/10.1080/17457300.2014.939196.

Chandra Dahal, R., Raj Dahal, K., 2020. A review on problems of the public building maintenance works with special reference to Nepal. Am. J. Constr. Build. Mater. 4, 39. https://doi.org/10.11648/j.ajcbm.20200402.12.

Cheng, C.W., Yao, H.Q., Wu, T.C., 2013. Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry. J. Loss Prev. Process Ind. https://doi.org/10.1016/j.jlp.2013.07.002.

Choi, J., Gu, B., Chin, S., Lee, J.S., 2020. Machine learning predictive model based on national data for fatal accidents of construction workers. Autom. Constr. 110, 102974. https://doi.org/10.1016/j.autcon.2019.102974.

Comu, S., Kazar, G., Marwa, Z., 2021. Evaluating the attitudes of different trainee groups towards eye tracking enhanced safety training methods. Adv. Eng. Informatics 49. https://doi.org/10.1016/j.aei.2021.101353.

Damayanti, F., Djakfar, L., Wisnumurti, Nugroho, A.M., 2023. Analysis of the role of big five personality throuch worker's safety culture and personal value as intervening variable on construction workers' safety behavior using sem-pls. Eastern-European J. Enterp. Technol. 3, 15–22. https://doi.org/10.15587/1729-4061.2023.280888.

De Marco, A., Ruffa, S., Mangano, G., 2010. Strategic factors affecting warehouse maintenance costs. J. Facil. Manag. https://doi.org/10.1108/14725961011041152.

Dragone, I.S., Biotto, C.N., Serra, S.M.B., 2021. Evaluation of Lean Principles in Building Maintenance Management, in: IGLC 2021 - 29th Annual Conference of the International Group for Lean Construction - Lean Construction in Crisis Times: Responding to the Post-Pandemic AEC Industry Challenges. pp. 13–22. doi: 10.24928/2021/0151.

Elkhoukhi, H., Bakhouya, M., El Ouadghiri, D., Hanifi, M., 2022. Using stream data processing for real-time occupancy detection in smart buildings. Sensors. https://doi.org/10.3390/s22062371.

Gao, Y., Gonzalez, V.A., Yiu, T.W., 2019. The effectiveness of traditional tools and computer-aided technologies for health and safety training in the construction sector: A systematic review. Comput. Educ. 138, 101–115. https://doi.org/10.1016/j.compedu.2019.05.003.

Gondia, A., Ezzeldin, M., El-Dakhakhni, W., 2022. Machine Learning–Based Decision Support Framework for Construction Injury Severity Prediction and Risk Mitigation. ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng. https://doi.org/10.1061/ajrua6.0001239.

Han, S., Saba, F., Lee, S., Mohamed, Y., Peña-Mora, F., 2014. Toward an understanding of the impact of production pressure on safety performance in construction operations. Accid. Anal. Prev. https://doi.org/10.1016/j.aap.2013.10.007.

Hauashdh, A., Jailani, J., Abdul Rahman, I., AL-fadhali, N., 2020. Building maintenance practices in Malaysia: a systematic review of issues, effects and the way forward. Int. J. Build. Pathol. Adapt. https://doi.org/10.1108/IJBPA-10-2019-0093.

Hauashdh, A., Jailani, J., Rahman, I.A., AL-fadhali, N., 2021. Structural equation model for assessing factors affecting building maintenance success. J. Build. Eng. 44. https://doi.org/10.1016/j.jobe.2021.102680.

Hauashdh, A., Jailani, J., Rahman, I.A., AL-fadhali, N., 2022. Strategic approaches towards achieving sustainable and effective building maintenance practices in maintenance-managed buildings: A combination of expert interviews and a literature review. J. Build. Eng. https://doi.org/10.1016/j.jobe.2021.103490.

Hire, S., Sandbhor, S., Ruikar, K., Amarnath, C.B., 2021. BIM usage benefits and challenges for site safety application in Indian construction sector. Asian J. Civ. Eng. https://doi.org/10.1007/s42107-021-00379-8.

Hon, C.K.H., Chan, A.P.C., Wong, F.K.W., 2010. An analysis for the causes of accidents of repair, maintenance, alteration and addition works in Hong Kong. Saf. Sci. 48, 894–901. https://doi.org/10.1016/j.ssci.2010.03.013.

Izudi, J., Ninsiima, V., Alege, J.B., 2017. Use of Personal Protective Equipment among Building Construction Workers in Kampala, Uganda. J. Environ. Public Health. https://doi.org/10.1155/2017/7930589.

Kadhim, E.M., Altaie, M.R., 2023. Factors affecting building maintenance practices: review. J. Eng. 29, 153–172. https://doi.org/10.31026/j.eng.2023.12.10.

Kamalinia, M., Mohammadi, H., Mohrz, M.H., Arassi, M., Faghih, M.A., Mirzajani, M., 2012. Investigation of occupational accidents induce sever injury in Fars, Iran from 2005 to 2007. Open J. Saf. Sci. Technol. https://doi.org/10.4236/ojsst.2012.23015.

Kang, K.S., Koo, C., Ryu, H.G., 2022. An interpretable machine learning approach for evaluating the feature importance affecting lost workdays at construction sites. J. Build. Eng. 53, 104534. https://doi.org/10.1016/j.jobe.2022.104534.

Kang, K., Ryu, H., 2019. Predicting types of occupational accidents at construction sites in Korea using random forest model. Saf. Sci. 120, 226–236. https://doi.org/10.1016/j.ssci.2019.06.034.

Karimi, H., Taghaddos, H., 2019. The influence of craft workers' educational attainment and experience level in fatal injuries prevention in construction projects. Saf. Sci. https://doi.org/10.1016/j.ssci.2019.04.022.

Kazar, G., Comu, S., 2021. Effectiveness of serious games for safety training: a mixed method study. J. Constr. Eng. Manag. https://doi.org/10.1061/(asce)co.1943-7862.0002119.

Kazar, G., Comu, S., 2022. Exploring the relations between the physiological factors and the likelihood of accidents on construction sites. Eng. Constr. Archit. Manag. https://doi.org/10.1108/ECAM-11-2020-0958.

Khairuddin, M.Z.F., Hasikin, K., Abd Razak, N.A., Lai, K.W., Osman, M.Z., Aslan, M.F., Sabanci, K., Azizan, M.M., Satapathy, S.C., Wu, X., 2022. Predicting occupational injury causal factors using text-based analytics: A systematic review. Front. Public Heal. https://doi.org/10.3389/fpubh.2022.984099.

Kim, J.M., Kim, T., Bae, J., Son, K., Ahn, S., 2019. Analysis of plant construction accidents and loss estimation using insurance loss records. J. Asian Archit. Build. Eng. 18, 507–516. https://doi.org/10.1080/13467581.2019.1687089.

Koc, K., 2023. Determining the Short-term Susceptibility of Construction Workers to Occupational Accidents Using Stochastic Gradient Boosting. J. Constr. Eng. Manag. Innov. https://doi.org/10.31462/jcemi.2023.01001015.

Koc, 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. Autom. Constr. 131. https://doi.org/10.1016/j.autcon.2021.103896.

Koc, K., Ekmekcioğlu, Ö., Gurgun, A.P., 2022. Accident prediction in construction using hybrid wavelet-machine learning. Autom. Constr. 133. https://doi.org/10.1016/j.autcon.2021.103987.

Koc, K., Gurgun, A.P., 2022. Scenario-based automated data preprocessing to predict severity of construction accidents. Autom. Constr. 140, 104351. https://doi.org/10.1016/j.autcon.2022.104351.

Koc, K., Ekmekcioğlu, Ö., Gurgun, A.P., 2023. Prediction of construction accident outcomes based on an imbalanced dataset through integrated resampling techniques and machine learning methods. Eng., Constr. Arch. Manage. https://doi.org/10.1108/ECAM-04-2022-0305.

Kryszkiewicz, M., 1998. Rough set approach to incomplete information systems. Inf. Sci. (ny). https://doi.org/10.1016/S0020-0255(98)10019-1.

Lee, J.Y., Yoon, Y.G., Oh, T.K., Park, S., Ryu, S.I., 2020. A study on data pre-processing and accident prediction modelling for occupational accident analysis in the construction industry. Appl. Sci. 10, 1–23. https://doi.org/10.3390/app10217949.

Liou, J.J.H., 2009. A novel decision rules approach for customer relationship management of the airline market. Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2008.05.002.

Liu, S.S., Ramadhan, G., 2017. Condition-based building maintenance strategy selection using constraint programming, in. In: ISARC 2017 - Proceedings of the 34th International Symposium on Automation and Robotics in Construction. https://doi.org/10.22260/isarc2017/0047.

Man, S.S., Chan, A.H.S., Alabdulkarim, S., Zhang, T., 2021. The effect of personal and organizational factors on the risk-taking behavior of Hong Kong construction workers. Saf. Sci. https://doi.org/10.1016/j.ssci.2020.105155.

Mohammadfam, I., Soltanzadeh, A., Moghimbeigi, A., Alizadeh Savareh, B., 2015. Use of Artificial Neural Networks (ANNs) for the analysis and modeling of factors that affect occupational injuries in large construction industries. Electron. Physician. https://doi.org/10.19082/1515.

Munkres, J., 2000. Topology, Second Edi. ed. Pearson.

Nnaji, C., Karakhan, A.A., 2020. Technologies for safety and health management in construction: Current use, implementation benefits and limitations, and adoption barriers. J. Build. Eng. 29. https://doi.org/10.1016/j.jobe.2020.101212.

Ogunbayo, B.F., Aigbavboa, C.O., Thwala, W., Akinradewo, O., Ikuabe, M., Adekunle, S. A., 2022. Review of culture in maintenance management of public buildings in developing countries. Buildings 12. https://doi.org/10.3390/buildings12050677.

Pandit, B., Albert, A., Patil, Y., Al-Bayati, A.J., 2019. Fostering safety communication among construction workers: Role of safety climate and crew-level cohesion. Int. J. Environ. Res. Public Health. https://doi.org/10.3390/ijerph16010071.

Pawlak, Z., 1982. Rough sets. Int. J. Comput. Inf. Sci. 11. https://doi.org/10.1007/BF01001956.

Pawlak, Z., 2002. Rough set theory and its applications. J. Telecommun. Inf. Technol. https://doi.org/10.26636/jtit.2002.140.

Rajaprasad, S.V.S., Chalapathi, P.V., 2015. Factors influencing implementation of OHSAS 18001 in indian construction organizations: Interpretive structural modeling approach. Saf. Health Work. https://doi.org/10.1016/j.shaw.2015.04.001.

Raza, M.S., Qamar, U., 2019. Understanding and Using Rough Set Based Feature Selection: Concepts, Techniques and Applications, Second Edition, Understanding and Using Rough Set Based Feature Selection: Concepts, Techniques and Applications, Second Edition. doi: 10.1007/978-981-32-9166-9.

Reiman, A., 2020. Human factors and maintenance in delivery transportation: drivers' work outside the cab in focus. J. Qual. Maint. Eng. https://doi.org/10.1108/JQME-05-2020-0035.

Salama, A.S., 2010. Topological solution of missing attribute values problem in incomplete information tables. Inf. Sci. (Ny). https://doi.org/10.1016/j.ins.2009.11.010.
Santiago, K., Yang, X., Ruano-Herreria, E.C., Chalmers, J., Cavicchia, P., Caban-Martinez, A.J., 2020. Characterising near misses and injuries in the temporary agency construction workforce: Qualitative study approach. Occup. Environ. Med. https://doi.org/10.1136/oemed-2019-106215.
Sarkar, S., Pramanik, A., Maiti, J., Reniers, G., 2020. Predicting and analyzing injury severity: A machine learning-based approach using class-imbalanced proactive and reactive data. Saf. Sci. 125, 104616. https://doi.org/10.1016/j.ssci.2020.104616.
Shen, L., Loh, H.T., 2004. Applying rough sets to market timing decisions. Support Syst Decis. https://doi.org/10.1016/S0167-9236(03)00089-7.
Shohet, I.M., 2003. Key performance indicators for maintenance of health-care facilities. Facilities 21, 5–12. https://doi.org/10.1108/02632770310460496.
Singh, A., Misra, S.C., 2021. Safety performance & evaluation framework in Indian construction industry. Saf. Sci. https://doi.org/10.1016/j.ssci.2020.105023.
Szer, I., Szer, J., Kaszubska, M., Miszczak, J., Hoła, B., Błazik-Borowa, E., Jabłoński, M., 2021. Influence of the seasons on construction site accidents. Arch. Civ. Eng. https://doi.org/10.24425/ace.2021.138067.
Thivagar, M., Richard, C., Rebecca Paul, N., 2012. Mathematical innovations of a modern topology in medical events. J. Inf. Sci Int. https://doi.org/10.5923/j.ijis.20120204.01.
Tixier, A.J.P., Hallowell, M.R., Rajagopalan, B., Bowman, D., 2016. Application of machine learning to construction injury prediction. Autom. Constr. 69, 102–114. https://doi.org/10.1016/j.autcon.2016.05.016.
Toptanci, Ş., Erginel, N., Acar, I.P., 2023. Predicting the severity of occupational accidents in the construction industry using standard and regularized logistic regression models. Ömer Halisdemir Üniversitesi Mühendislik Bilim. Derg. doi: 10.28948/ngumuh.1212385.
Unnikrishnan, S., Iqbal, R., Singh, A., Nimkar, I.M., 2015. Safety management practices in small and medium enterprises in India. Saf. Health Work. https://doi.org/10.1016/j.shaw.2014.10.006.
Wang, B., Xia, X., 2015. Optimal maintenance planning for building energy efficiency retrofitting from optimization and control system perspectives. Energy Build. 96, 299–308. https://doi.org/10.1016/j.enbuild.2015.03.032.
Wong, T.K.M., Man, S.S., Chan, A.H.S., 2020. Critical factors for the use or non-use of personal protective equipment amongst construction workers. Saf. Sci. 126. https://doi.org/10.1016/j.ssci.2020.104663.
Xia, N., Zou, P.X.W., Liu, X., Wang, X., Zhu, R., 2018. A hybrid BN-HFACS model for predicting safety performance in construction projects. Saf. Sci. 101, 332–343. https://doi.org/10.1016/j.ssci.2017.09.025.
Xu, N., Ma, L., Liu, Q., Wang, L., Deng, Y., 2021. An improved text mining approach to extract safety risk factors from construction accident reports. Saf. Sci. 138. https://doi.org/10.1016/j.ssci.2021.105216.
Yang, X., Yang, J., 2012. Incomplete information system and rough set theory: Models and attribute reductions, Incomplete Information System and Rough Set Theory: Models and Attribute Reductions. doi: 10.1007/978-3-642-25935-7.
Yi, W., Chan, A.P.C., 2015. Optimal work pattern for construction workers in hot weather: a case study in Hong Kong. J. Comput. Civ. Eng. https://doi.org/10.1061/(asce)cp.1943-5487.0000419.
Yigit, U., 2022. The rough topology for numerical data. Arxiv. doi: doi:10.48550/ARXIV.2206.05776.
Zákányiné Mészáros, R., Kárpi, M., Norbert, T., Zákányi, B., 2022. A human safety approach to safe maintenance. Multidiszcip. Tudományok 12, 93–99. https://doi.org/10.35925/j.multi.2022.4.9.
Zhang, Z., Shi, Y., Gao, G., 2009. A rough set-based multiple criteria linear programming approach for the medical diagnosis and prognosis. Expert Syst. Appl. 36, 8932–8937. https://doi.org/10.1016/j.eswa.2008.11.007.
Zolkafli, U.K., Zakaria, N., Mohammad Mazlan, A., Ali, A.S., 2019. Maintenance work for heritage buildings in Malaysia: owners' perspectives. Int. J. Build. Pathol. Adapt. https://doi.org/10.1108/IJBPA-07-2018-0062.